Yayın Geliş Tarihi (Submitted): **19/07/2024**

Yayın Kabul Tarihi (Accepted): **12/12/2024**

Makele Türü (Paper Type): **Araştırma Makalesi – Research Paper**

Please Cite As/Atıf için:
Öztürk, Z. (2024), The degree of challenge of achieving sustainable development goals on Turkey firms: Item response model approach, *Nicel Bilimler Dergisi*, 6(2), 158-180. *doi: 10.51541/nicel.1518755*

THE DEGREE OF CHALLENGE OF ACHIEVING SUSTAINABLE DEVELOPMENT GOALS ON TURKEY FIRMS: ITEM RESPONSE MODEL APPROACH

Zeynep ÖZTÜRK[1]

ABSTRACT

Sustainability can be determined as a way to protect the natural environment and biodiversity while meeting human needs. In 2015, the United Nations set 17 Sustainable Development Goals (SDGs) to create a better world in which to live. The challenges associated with achieving these goals have been documented in the literature, often with descriptive statistics and interpretations. The purpose of this study is to utilize the Item Response Theory (IRT) to assess the challenges of firms in Turkey to achieve sustainable development goals. In 2022, binary data was constructed from the sustainability and annual reports of 151 companies in Turkey. The Rasch model was identified as the most fitting model based on the results of the information criteria. Following the testing of the various assumptions, it was determined that the model is fully compatible with the infit, outfit, unidimensionality, and point serial correlation statistics. The difficulty levels of the items were identified using difficulty parameters and a Wright Map. The item difficulty levels were distributed as follows: It can be reasonably deduced that SDGs 5, 8, 12, and 13 can be readily achieved; SDGs 4, 7, 9, and 17 can also be attained with relative ease; SDGs 3 and 10 can be achieved with a moderate level of effort; SDGs 6, 11, 15 and 16 are more challenging to achieve; and SDGs 1, 2, 14 and 17 are extremely challenging to attain. The results show that companies in Turkey have different

[1]Corresponding Author, Asst. Prof. Dr, Department of Business Administration, Faculty of Economics, Administrative and Social Sciences, Artvin Çoruh University, Artvin, Türkiye. ORCID ID: https://orcid.org/0000-0003-0475-1357

challenges in achieving sustainable development goals. Increasing support and resources for difficult goals could make it easier for companies to achieve them.

Key Words: Sustainable Development Goals, Item Response Model, Rasch Model, Item Difficulty

TÜRKİYEDEKİ FİRMALARIN SÜRDÜRÜLEBİLİR KALKINMA HEDEFLERİNİN GERÇEKLEŞTİRİLMESİNDEKİ ZORLUK DERECELERİ: MADDE TEPKİ MODELİ YAKLAŞIMI

ÖZ

Sürdürülebilirlik, insan ihtiyaçlarını karşılarken doğal çevreyi ve biyoçeşitliliği korumanın bir yolu olarak belirlenebilir. 2015 yılında Birleşmiş Milletler, içinde yaşanacak daha iyi bir dünya yaratmak için 17 Sürdürülebilir Kalkınma Hedefi (SKH) belirlemiştir. Bu hedeflere ulaşılmasıyla ilgili zorluklar, literatürde genellikle tanımlayıcı istatistikler ve yorumlarla belgelenmiştir. Bu çalışmanın amacı, Türkiye'deki firmaların sürdürülebilir kalkınma hedeflerine ulaşma konusundaki zorluklarını değerlendirmek için Madde Tepki Teorisi'ni (MTK) kullanmaktır. 2022 yılında, Türkiye'deki 151 şirketin sürdürülebilirlik ve faaliyet raporlarından ikili veriler oluşturulmuştur. Rasch modeli, bilgi kriterlerinin sonuçlarına göre en uygun model olarak belirlenmiştir. Çeşitli varsayımların test edilmesinin ardından, modelin infit, outfit, tek boyutluluk ve nokta seri korelasyon istatistikleriyle tam uyumlu olduğu belirlenmiştir. Maddelerin güçlük düzeyleri, güçlük parametreleri ve Wright Haritası kullanılarak belirlenmiştir. Madde güçlük düzeyleri aşağıdaki gibi dağılmıştır: SKH 5, 8, 12 ve 13'e kolayca ulaşılabileceği; SKH 4, 7, 9 ve 17'ye görece kolaylıkla ulaşılabileceği; SKH 3 ve 10'a orta düzeyde bir çabayla ulaşılabileceği; SKH 6, 11, 15 ve 16'ya ulaşmanın daha zor olduğu; SKH 1, 2, 14 ve 17'ye ulaşmanın ise son derece zor olduğu sonucuna varılabilir. Sonuçlar, Türkiye'deki şirketlerin sürdürülebilir kalkınma hedeflerine ulaşmada farklı zorluklar yaşadığını göstermektedir. Zor hedefler için destek ve kaynakların artırılması, şirketlerin bu hedeflere ulaşmasını kolaylaştırabilir.

Anahtar Kelimeler: Sürdürülebilir Kalkınma Hedefleri, Madde Tepki Modeli, Rasch Modeli, Madde Zorluğu

1. INTRODUCTION

Sustainable development was identified as "*development that meets the needs of the present without compromising the ability of future generations to meet their own needs*" in 1987 by the Brundtland Commission Report (Imperatives,1987). Then, On September 25, 2015, the United Nations proposed 17 sustainable development goals (See Figure 1) to achieve aims such as safeguarding the planet, ending poverty, and promoting prosperity for everyone. The 17 sustainable development goals and 169 targets are focused on the dimensions of sustainable development (social, environmental, and economic), and the goals are indivisible and integrated. That is, the 2030 Agenda for Sustainable Development emphasizes the importance of the integrated and interlinkages nature of sustainable development goals.

Figure 1. Sustainable development goals (UN-SDGs,2015)

The notion of sustainable development is an inclusive and ideal concept, however, its achievement is extremely tough and a herculean task. This achievement required political will and social nods (Jain and Jain, 2020), but the effect of all countries and stakeholders is incontrovertible in achieving the goals and targets. There are some main challenges relating to

the implementation of action plans and strategies to achieve the goals. Assessing and evaluating progress in a particular sector, coordinating global, national, and local responses, and accessing resources and information to understand the goals are some of the key challenges (Fleming et al, 2017). Despite these difficulties, many firms strive to achieve these goals. Different companies will achieve different results in their efforts to meet the Sustainable Development Goals.

The primary objectives of these goals are to elevate living standards worldwide, reduce poverty, eliminate inequalities, ensure environmental sustainability, and create a fair, healthy, and secure future for all. The SDGs emphasize a global effort and partnership that involves not only governments but also firms, civil society, and individuals. Each of these sustainable development goals is difficult to achieve and countries and firms have different priorities and approaches towards sustainability. As such, there is no globally accepted sustainability policy, there are just only specified sustainable development goals. Countries and firms try to achieve these goals by investing in sustainability. Considering that achieving these goals has different degrees of difficulty for each country and business, the study investigates the difficulty levels of the applicability of sustainability goals. The approach adopted in this research includes the collection of information from different companies to study the correlations between sustainable development goals and their applicability. In this context, Item response theory (IRT) is preferred to use in analysis. This theory is designed to model the characteristics of each item, such as difficulty, discrimination, and predictability, and the responses of individuals (firms) to these items. Also known as Item Response Theory, which is another name for the Latent Variable Theory, it emerged as a response to Classical Test Analysis. While it was initially proposed in the 1930s, the main studies on the topic began in the 1950s. Despite being considered a more advanced theory than Classical Test Analysis, IRT has challenges in its application and calculation. Tucker (1946) introduced the term "Item Characteristic Curve," one of the key concepts of the Latent Variable Theory. Lord (1950) focused on this theory and developed the normal ogive model. Towards the end of the 1950s, Birnbaum replaced the normal ogive model with the logistic model. In the 1960s, Rash made a significant contribution to this field by developing a model named after himself. Subsequently, the Rash model became one of the most important models of this theory. More complex models, such as the three-parameter (3PL) and two-parameter (2PL) models, emerged in the 1980s. With the advancement of computer technology, the wider application and expansion of the field of IRT have become possible.

As far as we know, although the literature stated that there are challenges in implementing the strategies, there is no study on the level of difficulty of these goals for firms. Therefore, this study aims to determine the level of difficulty of these goals for firms. Which of these 17 goals is easier or more difficult for firms to achieve?

In brief, sustainable development goals aim for a combination of environmental sustainability, economic development, and social inclusion (Sachs, 2012; Fonseca et al., 2020). The primary objective of SDGs is to make a significant contribution to the sustainable development achievement for all societies and also address the needs of current and future stakeholders. Additionally, promoting globally the integration and operationalization of sustainability into organizations is important (Fonseca et al., 2020).

The model enables the determination of the difficulty levels of the items to be implemented and the targeted ratios of the items. Item characteristic curves, in conjunction with the Wright Map graph, provide a visual representation of the general characteristics, comparisons, and implementation difficulty levels of the sustainable development goal items. This model serves as a tool for comprehending and evaluating the obstacles encountered in the implementation of sustainable development goals. Consequently, the obstacles encountered by firms in Turkey in the application of SDG items are evaluated and analyzed. Notwithstanding the difficulties encountered in implementing the SDGs, they offer crucial guidance for a sustainable global future. Efforts to achieve these goals persist in numerous countries, businesses, and individuals. It is therefore anticipated that this study will provide valuable insights for the business world. After the introduction, a literature review is presented. The study ends with a methodology section with results and findings, and then a discussion and conclusion.

2. THE SUSTAINABLE DEVELOPMENT GOALS

The world is in a new geological time, in which humankind has come to play an effective and threatening role in earth dynamics. Human beings face many human-caused crises such as climate change and ocean acidification (based upon human-produced greenhouse gases), extensive environmental pollution, the loss of biodiversity, fossil resources depletion, and the conversion of wilderness and forests into pastures and farms. At this point, SDGs are a significant idea and help ultimately put the earth on a sustainable path (Sachs, 2012).

As mentioned, the SDGs are significant ideas for the entire world, and the effort to achieve SDGs has accelerated because of the urgency of sustainability. Generally, The SDGs focus on targets (known as the five Ps) for “people, planet, prosperity, partnership, and peace” (Sachs et al., 2019). However, specific targets may change globally, between and within societies (Sachs, 2012: 2206). SDGs' nature and scope greatly differ and the targets have diverse functions. Some targets are connected to multiple goals, while other goals have weak relations with other targets. For example, eliminating poverty (SDG1), promoting economic growth and employment (SGD 8), reducing inequality (SDG 10), and sustainable consumption and production (SDG12) are directly or indirectly related to at least ten other targets. Life on land (SDG 15) is also related to six other targets (Katila et al., 2019). Additionally, Fonseca et al, (2020) stated that achieving one SDG may either harm or reinforce another target. For example, Industrial growth and economic expansion contributed to hunger or poverty reduction and providing clean water/sanitation. Anyway, this industrial and economic development also had no positive effects on some social and environmental targets.

To successfully achieve sustainable goals, some considerations must be taken into account. If the sustainable development goals are to succeed, an inclusive approach to growth must be promoted eliminating investments in unsustainable acts in all countries and mobilizing innovative financing sources (Stafford et al, 2017). Again, "accountability" is important to have effective goals. The question that who is supposed to do what to get the proposed goal accomplished must be answered (Pogge and Sengupta, 2015). According to Filho et al (2019), Inclusiveness and reducing inequalities are significant. Sustainable development goals universally apply to all developed and developing countries. The private sector, public sector, and civil society should be in strong alliances and improved cooperation The implementation of the Sustainable Development Goals will demand financial support and increased accountability.

In addition to this, social protection, decent employment, well-managed policies, more resources for essential services, strong national ownership, and supported coherently by partners can increase the success of the SDGs (Singh, 2016). Sustainable Development Goals can move communities away from the model of economic prosperity toward a more sustainable and holistic prosperity. Hence, the success of these goals rests on its implementation (Mair et al, 2017). Also, the agenda should cover strong human rights language to enlarge its impact and ambition (Pogge and Sengupta, 2015).

Another outstanding point is the regional disparity. For example, Africa is the largest region to not achieve the sustainable development goals, followed by Asia (Moyer and Hedden, 2020). In brief, the planet, countries, and communities need to determine and express shared goals, and also create ways to follow progress in meeting them. Sustainable well-being should be measured for a desirable and sustainable future (Costanza et al., 2014) Robert et al (2005) also stated that another way to describe sustainable development is in how it is measured. When we consider all these success factors, the efforts of some firms in Turkey will be evaluated in the study. In this context, 151 companies are reviewed.

3. METHODOLOGY

3.1. Item Response Model (IRT)

Item Response Models (IRT) describe the relationship between test items and individuals, to determine the probability that a person has a particular trait. The relation between an individual's item performance and the set of traits related to that performance is represented by a monotonically increasing function known as the item characteristic function, which is also referred to as the item characteristic curve (ICC). This function is determined by increases in the probability of a correct response to an item, which are in turn determined by increases in the level of the trait. In this way, IRT is used to assess the reliability and validity of a test, to revise items, or to help design tests more effectively according to people's abilities. In item response theory, people's responses to an item on a test can be measured as true or false, or as belonging to one of several categories. IRT deals with the concept of "ability". The underlying theory posits that the performance of an individual on a test is indicative of their ability, which is the construct that the test is designed to measure. The unidimensionality, local independence, invariance, model fit, and item characteristic function assumptions are employed to evaluate the reliability, accuracy, and validity of tests and to gain insight into the characteristics being measured. IRT is classified into one-parameter, two-parameter, and three-parameter item response models for binary responses.

3.1.1. Binary Item Response Models

Rasch Model

The Rasch model (Rasch, 1960) is a one-parameter logistic model and is one of the most popularly used item response models. A one-parameter response model is expressed by the equation

$$P(\mathrm{Y_{ik}} = 1|\theta_i\,, \mathrm{b_k}\,) = \frac{e^{\theta_i - b_k}}{1+e^{\theta_i - b_k}} = \frac{1}{1+e^{b_k - \theta_i}} \qquad i = 1,2\ldots, n \tag{1}$$

where the ability level is given by θ_i, the number of items is given by n, and the difficulty parameter b_k is the probability of 0.5 correct response for the i-th individual.

Two-parameter Model

In the two-parameter logistic model, a discrimination parameter is added to the model and is defined by the equation

$$P(\mathrm{Y_{ik}} = 1|\theta_i\,, a_k, b_k\,) = \frac{e^{a_k\theta_i - b_k}}{1+e^{a_k\theta_i - b_k}} = \frac{1}{1+e^{b_k - a_k\theta_i}} \tag{2}$$

where the discrimination parameter is given by a_k.

Three-parameter Model

In the three-parameter logistic model, a chance parameter is added to the model and is defined by the equation

$$P(\mathrm{Y_{ik}} = 1|\theta_i\,, a_k, b_k, c_k) = c_k + \frac{(1-c_k)}{1+e^{b_k - a_k\theta_i}} \tag{3}$$

$$= \frac{1}{1+e^{b_k - a_k\theta_i}} + \frac{c_k}{1+e^{a_k\theta_i - b_k}}$$

where the chance parameter is c_k.

3.2. Applicability of the Item Response Model to Sustainable Development Goals (SDGs)

IRT addresses issues encountered in psychometrics and education, such as test development, creating question banks, developing tailored tests for individuals, determining item bias, assessing item difficulty and ease, weighing options, and test equating. Although primarily used in education and psychometrics, it is also applied in various fields such as health sciences, recruitment processes, social sciences, marketing, and others. Its flexibility

and analytical power allow adaptation to different measurement situations. In the literature, IRT on specific marketing and business research topics (i.e., non-directly observable concepts such as consumer satisfaction, satisfaction when purchasing a product, brand loyalty, or work motivation) has been available for 30 years (Ewing et al. 2005; De Jong et al. 2008; Kamakura and Balarubramian, 1989; Schultz et al. 2013; Raykov and Calantone, 2014). Busco et al. (2018), "Achieving SDGs will be a challenging battle in the coming years, and many commercial organizations worldwide have begun to determine and implement sustainable strategies as the fundamental driving forces of their goals, visions, and business models". Commercial organizations in Turkey have rapidly joined this effort. To achieve sustainable development goals, firms are developing new business models, exploring market opportunities, and adapting to new needs. Jones et al.(2016) that the financial services industry has a pivotal role in facilitating sustainable consumption, yet it faces substantial hurdles in engaging effectively with the SDGs. This is compounded by the fact that many Turkish firms tend to focus on specific SDGs, such as gender equality, without a comprehensive evaluation of their overall sustainability impact, as discussed by Alkan and Kamaşak (Alkan and Kamaşak, 2023). The selective approach to SDG implementation can lead to significant gaps in achieving broader sustainability goals.

One of the aims of the Item Response Model is to determine item difficulty. Studies on item difficulty usually aim to assess the quality of a test and to identify items that require revision or reorganization of the test. A binary measure was created by coding 1 for SDG items targeted by firms and 0 for SDG items not targeted by firms. The Rasch model, which can be used for binary data, can provide the use of the measurement model for SDG items and improve the processes of developing, changing, or adding targets. For example, the Rasch model approach has been used in the literature to assess the difficulty of healthy eating and which dietary guidelines consumers have more or less difficulty following (Henson et al., 2010). Darmana et al. (2002) evaluated the difficulty level of the items examined using the Rasch model for the validation of Chemistry National Examination Tools. By comparing students' abilities with the difficulty levels of the questions, the probability of answering correctly was analyzed. In this study, the Rasch model was used to measure the difficulty level of the applicability of sustainable development goal items.

4. RESULTS

4.1. Data Set

The SDGs comprise 17 goals, all of which are framed within the context of economic, social, and environmental considerations. Concerning the Rasch model, it has been suggested that 151 samples can be used to accurately estimate the difficulty parameter in tests with 10, 20, or 30 items (Şahin & Anıl, 2016). Data from Turkish businesses were sampled, and based on the responses to sustainable development goals from the sustainability reports of firms in 2022, a binary format was created (present=1/absent=0). The Binary Rasch Model was analyzed using the Jamovi 2.3.28 and R programs. The fact that the data was collected from sustainability reports provides a limited perspective only with the information reported by the companies. Therefore, the accuracy of the reports can be known to the extent that the companies declare in writing. Alkan and Kamaşak (2023) conducted in-depth interviews with human resources personnel of 13 multinational companies operating in Turkey and stated that global companies adopted the Sustainable Development Goals to different degrees and that the companies used the Sustainable Development Goals as initiatives and projects. It should also be taken into account that companies may generally tend to emphasize positive results when reporting their own sustainability performance. This may lead to a phenomenon known as "greenwashing". There is evidence in the literature that some companies may provide misleading information in their reports in order to make their sustainability performance seem better than it is (Delmas & Burbano, 2011). Being aware of this risk, the extent to which the reported data is actually accurate and objective is open to debate. In our study, we take into account that this possible bias may have an impact on the results and we clearly state this limitation in this article.

4.2. Analysis

In this study, Rasch analysis was conducted in Jamovi, an open-source statistical software that provides a user-friendly interface using the statistical power of the R program. The SnowRMM and snowIRT modules were employed. The snowIRT module utilizes the 'erm' package from the R software. Furthermore, the software enables the generation of item fit statistics, Item Characteristic Curves (ICCs), Wald test, Wright's map, Martin-Löf Likelihood Ratio test for unidimensionality, and Anderson's Likelihood test for invariance

(Mair et al., 2023). Additionally, the optimal model can be selected and some graphical representations can be created with the 'mirt' package from R software.

4.2.1. Fitting Model

Before undertaking the analyses, it is essential to determine the optimal model. To this end, comparisons have been made between various fit criteria, including the Akaike Information Criterion (AIC), the Schwarz Bayesian Information Criterion (SABIC), the Hannan-Quinn Information Criterion (HQ), the Bayesian Information Criterion (BIC), and the log-likelihood (loglik) values of the 1PL, 2PL and 3PL models. These are presented in Table 1.

Table 1. Optimal model selection

	AIC	SABIC	HQ	BIC	logLik
1PL(Rasch)	2713.185	2710.528	2735.249	2767.496	-1338.592
2PL	2715.499	2710.480	2757.176	2818.087	-1323.750
3PL	2715.454	2707.926	2777.969	2869.335	-1306.727

From Table 1, given that the lowest fit statistic values align with the Rasch model, it can be concluded that the Rasch model is the most fitting in this context.

4.2.2. Item characteristic curves (ICC)

Item characteristic curves are graphs used to assess item performance during the analysis of the Rasch model. These curves show, for each item, the probability that individuals with different ability levels will answer the item correctly. A good model fit is indicated when the expected and observed ICCs fall within the 95% confidence interval. Figure 2 illustrates the ICCs for SDG 1 and SDG 13. Figure 2 reveals that SDG 13 has a strong fit, with the expected probabilities (represented by a solid curve) aligning closely with the observed probabilities (shown as dots). The dots are evenly distributed along and near the curve. Conversely, SDG 1 exhibits a poor fit; the expected probabilities (solid line) and observed probabilities (dots) diverge significantly above logit 0, with the dots positioned far from the curve. This suggests that SDG 13 achieves a good model fit, whereas SDG 1 does not.

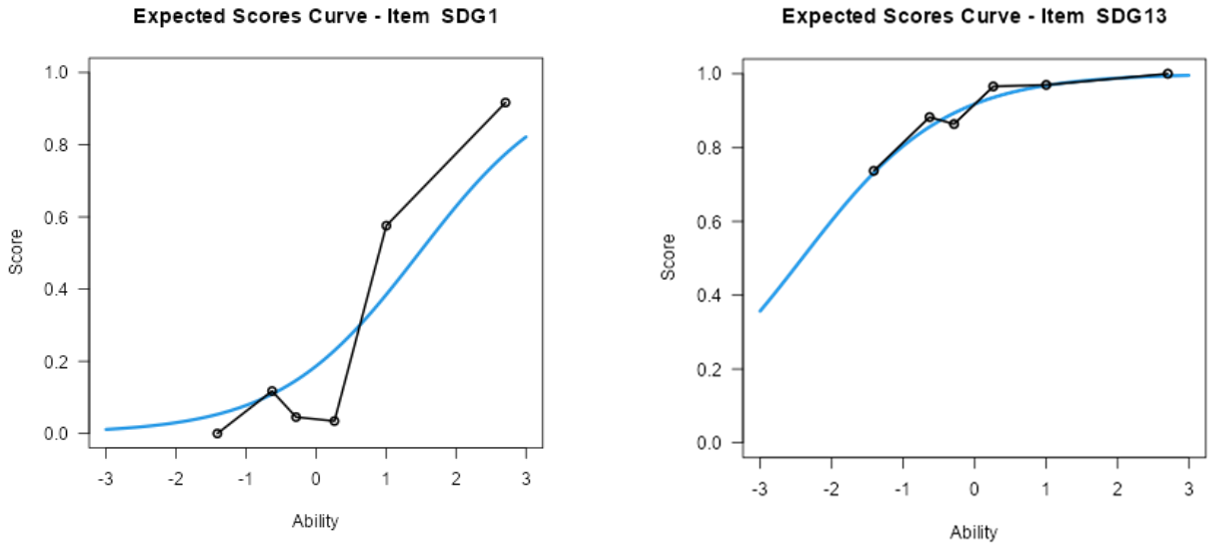

Figure 2. SDG 1 and SDG 13 item characteristic curves

The item characteristic functions of the items are shown in Figure 3.

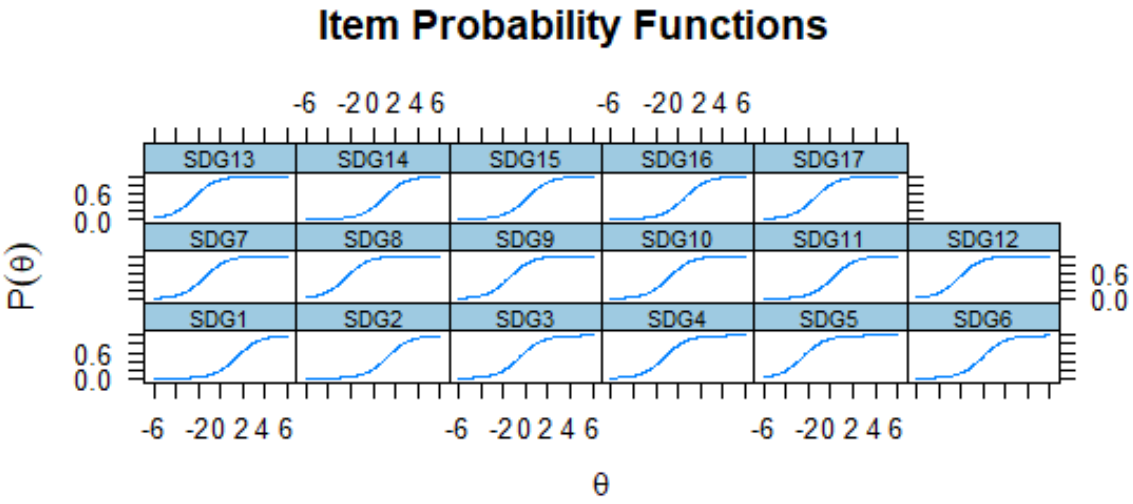

Figure 3. Item probability functions

4.2.3. Item Fit Statistics

Item fit statistics are comprised of two fit statistics: infit MNSQ and outfit MNSQ. MNSQ is a metric that evaluates the degree of correspondence between a test item or a person's performance and the expected value according to the Rasch model. It is defined as the mean squared error. The MNSQ represents the square of the deviation of each observation

from the expected model values, averaged over all observations. This metric assesses the variance of the deviations and measures the overall fit of the test. Infit is employed to assess the impact of discrepancies between the anticipated and observed performance, with a particular focus on items of moderate difficulty and participants of moderate ability. Outfit is a metric that assesses the effect of the deviation between the model's expected performance and observed performance, particularly the effect of extreme or unexpected responses. While the fit is an information-weighted fit statistic sensitive to responses to test items, the outfit is a fit statistic sensitive to outliers (Bond & Fox, 2007). Both values are considered to be within the acceptable range when they fall between 0.5 and 1.5. Values exceeding 1 may indicate a discrepancy in the fit with the model, whereas values below 1 may indicate overfitting (Linacre, 2002). Point series correlation refers to the correlation between the item scores of the observations in the data and the total item scores. A high correlation value indicates that an item can discriminate between respondents in terms of their ability (Linacre, 2021). Consequently, it is recommended that the point series correlation values exceed 0.30 (Bond et al., 2015). In contrast, Othman et al. (2014) also categorized the desired measurement values as weak and small correlation (less than 0.35), moderate and reasonable correlation (0.36-0.67), and strong and high correlation (0.68-1.00). Table 2 presents item fit statistics and point serial correlation.

Table 2. Item fit statistics and point serial correlation

Goals	Infit	Outfit	Point Serial Correlation
SDG1	0.800	0.651	0.622
SDG2	1.031	1.397	0.449
SDG3	1.080	1.041	0.403
SDG4	0.988	0.971	0.404
SDG5	0.976	1.057	0.278
SDG6	1.107	1.097	0.405
SDG7	1.022	0.931	0.370
SDG8	0.865	0.561	0.391
SDG9	0.953	0.993	0.395
SDG10	0.927	0.864	0.481
SDG11	1.119	1.222	0.394
SDG12	0.923	1.705	0.348
SDG13	1.009	0.909	0.285
SDG14	1.104	1.031	0.439
SDG15	0.895	0.825	0.553
SDG16	0.958	0.954	0.514
SDG17	1.134	1.309	0.280

Table 2 indicates that the infit values are within the range of 0.80 to 1.12, while the outfit values are within the range of 0.65 to 1.70. The only item, SDG 12, which exhibited an outfit value of 1.70, was found to be outside the acceptable limits of the agreement. However, all other outfits and infit values were found to be within acceptable limits. Consequently, the results obtained indicate that the items are in agreement about the measurement. Furthermore, an analysis of the point serial correlation data in Table 2 revealed that all values were within the range of 0.28-0.62. The only point serial correlation value below the desired limit is that of SDG17, which is 0.28. In other words, the results indicated that the remaining items, except SDG17, exhibited moderate and reasonable correlation. Consequently, the capabilities of firms with sustainable development goals are distinctive.

4.2.4. Unidimensionality

Another component evaluated to determine construct validity is unidimensionality analysis. The Rasch model states that a measurement tool should have a unidimensional structure, as it is expected that it will contribute to the measurement of a single trait (Bond et al., 2015; Linacre, 2023). Unidimensionality can be analyzed using both principal component analysis and the Martin-Löf (MLoef) likelihood ratio (LR) test.

4.2.5. Martin-Löf (MLoef) Likelihood Ratio (LR) Test

The Martin-Löf (MLoef) likelihood ratio (LR) test was employed to ascertain unidimensionality. The MLoef test classifies the items into two distinct groups based on the median of the raw scores for each item and then tests whether the two groups are homogeneous. The results are as follows: The LR value was found to be 53.9, with a degree of freedom (df) of 71 and a p-value of 0.934 ($p > 0.05$). This indicates that the item difficulty values based on the two subgroups of students are statistically identical, thereby confirming the unidimensionality of the item measures. Table 3 shows the unidimensionality.

Table 3. Martin-Loef Test-Median

	Value	df	p
Likelihood ratio	53.9	71	0.934

As a result of testing the assumptions, the model is fully fit according to the infit, outfit, unidimensionality, and point serial correlation statistics.

4.2.6. Item difficulty

The purpose of obtaining the item difficulty parameter is to determine the chance of answering a question correctly at a given ability level. The item difficulty parameter is measured in logit units. The range of difficulty of good questions is between -2.0 and +2.0 logits (Hambleton and Swaminathan, 1985). A question item is considered to be very difficult if it has a difficulty index above +2.00 logit; a question item is accounted for to be very easy if it has a difficulty index below -2.0 logit. A question item is considered "very difficult if its value b (measurement item) > 1; difficult if $0.5 \leq b < 1$; moderate $-0.5 \leq b < 0.5$; easy $-0.5 \leq b < 1$; and very easy $b \leq -1$" (Adedoyin & Mokobi, 2013).

Item parameter estimates (difficulty and ability) were obtained using the Bootstrap Joint Maximum Likelihood (JML) estimation method. Table 4 shows the estimates of the ability and difficulty parameters and their standard errors. Higher scores represent more difficult or less accessible items, while lower scores represent more achieved items.

From Table 4, the targets consisted of samples ranging from 2,338 to -1,916, with the hardest achievable target and the easiest achievable target items. Considering the reference values of Adedoyin and Mokobi (2013), SDG5(-1.916), SDG8(-1.839), SDG12(-1.564), SDG13(-1.839) are very easily achievable, SDG4(-0.568), SDG7(-0.827), SDG9(-0.827), SDG17(-0. 782) are easily achievable, SDG3(0.185), SDG10(-0.140) are moderately achievable, SDG6(0.672), SDG11(0.952), SDG15(0.881) are difficult achievable and SDG1(2.338), SDG2(2.140), SDG14(1.784), SDG16(1.350) are very difficult achievable target items.

Table 4. Item ability and difficulty parameter estimates

Goals	Ability	Difficulty	S.E. Difficulty
SDG1	0.225	2.338	0.227
SDG2	0.252	2.140	0.218
SDG3	0.596	0.185	0.188
SDG4	0.728	-0.568	0.203
SDG5	0.894	-1.916	0.280
SDG6	0.503	0.672	0.186
SDG7	0.768	-0.827	0.212
SDG8	0.887	-1.839	0.273
SDG9	0.768	-0.827	0.212
SDG10	0.656	-0.140	0.193
SDG11	0.450	0.952	0.188
SDG12	0.861	-1.564	0.252
SDG13	0.887	-1.839	0.273
SDG14	0.305	1.784	0.205
SDG15	0.464	0.881	0.187
SDG16	0.377	1.350	0.194
SDG17	0.762	-0.782	0.211

4.2.7. Wright's map

The Wright map is a visual representation of the distribution of item difficulties and firm abilities on the same graph. It shows how well the items discriminate between different levels of the basic trait being measured. The Wright map is presented as two vertical histograms. The left side shows the firms and the right side shows the items. The capabilities of the firms are measured from most to least capable. The items on the right are measured from most difficult to least difficult. The gap indicates that the items at that level are missing and the items in that gap cannot be correctly distinguished (Liu, 2020). The distribution of firms' ability levels has a range where the lowest is above -2 and the highest is above +2. Figure 4 shows the Wright map.

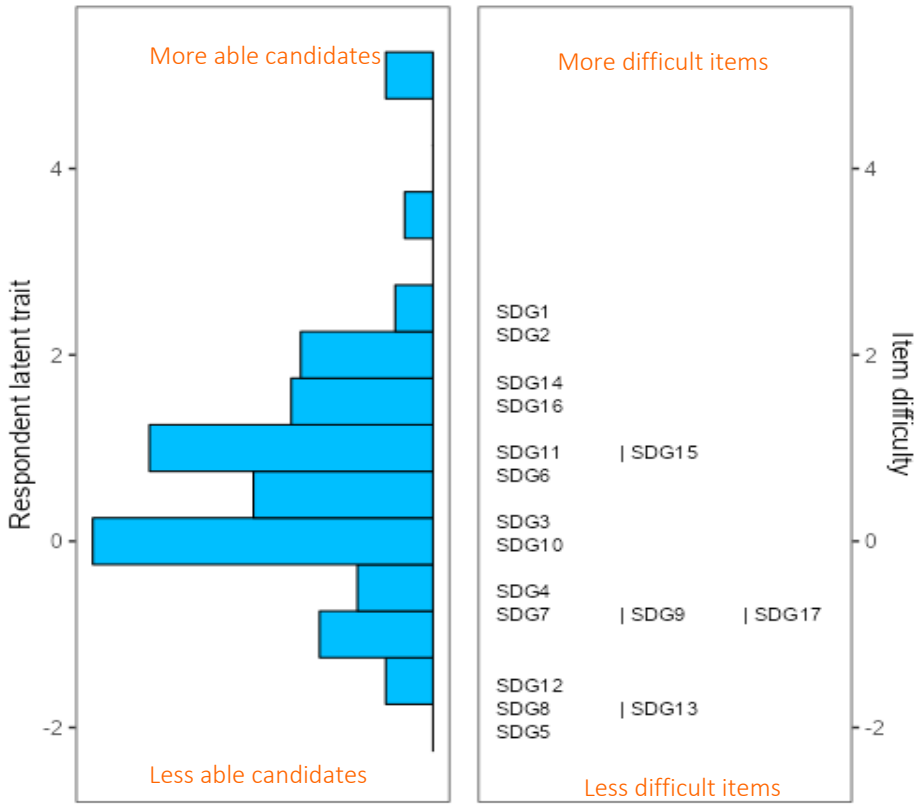

Figure 4. Wright Map

Figure 4 indicates that the ability levels of the firms to achieve the objectives are almost similar. However, some firms differ from other firms in terms of ability – a level above 4. The distribution of firms' ability levels is skewed to the left. The item difficulty levels are distributed as follows: SDG5, SDG8, SDG512, SDG13 (23.5% of the total) very easily achievable, SDG4, SDG7, SDG9, SDG17 (23.5% of the total) easily achievable, SDG3, SDG10 (12% of the total) medium achievable, SDG6, SDG11, SDG15 (17.6% of the total) difficult achievable, and SDG1, SDG2, SDG14, SDG16 (23.5% of the total) very difficult achievable. The 17 Global Goals (SDGs) have been designed to be challenging, with a range of difficulty levels.

4.2.8. Invariance

A reliable measurement tool must demonstrate invariance for both the sample and the set of items used to generate ability measures, as supported by Liu (2020). The Wald test splits the sample into two groups based on the median raw score and uses a Z-test to assess whether the item difficulty values between the two subsamples are statistically equivalent. Due to the multiple tests performed and the subsequent inflation of Type I error, a Bonferroni-

type adjustment of the alpha level is necessary. For the 17 SDG items, the adjusted p-value for significance is 0.05/17, which equals 0.003. Therefore, a p-value threshold of 0.003 is employed to determine if the observed differences are statistically significant. Table 5 presents the outcomes of the Wald test.

Table 5. Wald test - Median

Item	**Z statistic**	**p**
SDG1	-1.6258	0.104
SDG2	0.7048	0.481
SDG3	1.5488	0.121
SDG4	0.0215	0.983
SDG5	-0.0753	0.940
SDG6	1.6139	0.107
SDG7	0.2202	0.826
SDG8	0.2202	0.826
SDG9	-1.2110	0.226
SDG10	1.6855	0.092
SDG11	-1.2395	0.215
SDG12	-0.1910	0.848
SDG13	1.6164	0.106
SDG14	-0.1068	0.915
SDG15	-0.8652	0.387
SDG16	1.0935	0.274
SDG17		

According to Table 5, Wald test, the null hypothesis is accepted for each item ($p>0.003$), thus invariance is ensured and the model is valid.

4.2.9. Anderson's Probability Test

The fundamental principle of this statistical test is subgroup homogeneity in Rasch models. Andersen's Likelihood Ratio (LR) test (Andersen, 1973) is calculated by dividing data into two groups based on median raw scores.

Table 6. Andersen's LR test – median

	Value	**df**	**p**
Likelihood ratio	17.8	15	0.273

The results of Anderson's LR test are as follows: The LR value is chi-square = 17.8, df = 15, p = 0.273. The p-value for the Anderson-LR test shows the degree of fit between the observed data set and the distribution predicted by the Rasch model. A small p-value (typically less than the alpha level of 0.05) indicates that the observed data set does not align with the distribution predicted by the Rasch model. Conversely, a large p-value suggests that the data set is more consistent with the distribution predicted by the Rasch model.

5. CONCLUSIONS AND DISCUSSION

Item response models are not a widely known and used method in marketing and business research. This study focuses on application-oriented research for sustainable development goals to encourage researchers in marketing and related fields to use item response models more frequently and benefit from their strengths. The purpose of this study is to utilize the Item Response Theory (IRT) to assess the challenges faced by firms in Turkey in achieving sustainable development goals. By scanning the literature, no similar study was found where IRT was used in the context of sustainability in the business world. In the literature, IRT is widely used, especially in the fields of education and psychometrics, but there is no direct reference to its use in the business world and in measuring sustainability performance. This shows that our study fills the gap in the literature and makes an innovative contribution.

The findings of the study offer a comprehensive insight into the obstacles Turkish businesses encounter in their pursuit of sustainable development goals. Some of the Sustainable Development Goals (SDGs) are relatively straightforward to achieve, whereas others present significant challenges. The relatively straightforward SDGs (5, 8, 12, and 13) were found to be relatively easily achievable. This indicates that companies are successful and demonstrate a commitment to gender equality, decent work, economic growth, sustainable consumption and production, and climate action. A considerable number of companies have adopted policies aimed at promoting gender equality and have implemented a range of initiatives to increase the participation of women in the workforce. The relatively straightforward SDGs (4, 7, 9, and 17), including those related to quality education, affordable and clean energy, industry, innovation, and infrastructure, as well as the achievement of these goals through partnerships, are relatively straightforward for companies to achieve and are included as targets. SDGs of medium difficulty (3, 10) are capable of attaining a moderate level of success in the domains of health and well-being, as well as in the reduction of inequalities. This indicates that there are some challenges, but that they can be surmounted. The challenging SDGs (6, 11, 15, 16), namely those of clean water and sanitation, sustainable cities and communities, living on the land, and peaceful, equitable, and strong institutions were identified as being particularly challenging for firms to achieve. The inability of companies to achieve these goals is largely attributable to factors beyond their control, such as climate conditions and government policies. Additionally, the costs associated with achieving these goals are often prohibitively high. SDGs that were identified as extremely challenging (1, 2, 14, 17) are those related to ending poverty, and hunger, protecting marine life, and forming partnerships. It was found that achieving these goals is particularly difficult for companies. The attainment of these goals is rendered exceedingly challenging for companies by a multiplicity of factors, including the constraints of limited resources, the complexities of policy and regulatory compliance, the inherent difficulties of navigating social and environmental variables, and the necessity of forging collaborative relationships with stakeholders.

These findings indicate that firms in Turkey encounter varying levels of difficulty in their pursuit of sustainable development goals. While easily achievable targets indicate that firms are performing well in some areas, extremely difficult targets reveal that firms face significant challenges in these areas. This offers valuable insight for policymakers and business leaders alike. The provision of increased support and resources in areas where

challenges are present can facilitate the achievement of sustainable development goals by firms.

Furthermore, the findings of this study indicate that the attainment of sustainable development goals is not solely contingent on economic strength, but also on the capacity of firms to engage in strategic planning and implementation. It is therefore of great importance for firms to take the necessary steps to strengthen their sustainability strategies. As a result, the original contribution of this study is to reveal the applicability of IRT in measuring the difficulties of firms in achieving the SDGs in the field of sustainability. In this respect, the study makes an important contribution to the literature both methodologically and in terms of application area. Applying future studies to different countries with more firm and sector data based on this study will increase the generalizability of the findings.

ETHICAL DECLARATION

In the writing process of the study titled "The Degree of Challenge of Achieving Sustainable Development Goals on Turkey Firms: Item Response Model Approach", there were followed the scientific, ethical and the citation rules; was not made any falsification on the collected data and this study was not sent to any other academic media for evaluation.

REFERENCES

Alkan, D. P. and Kamaşak, R. (2023), The practice of sustainable development goals washing in developing countries, *Proceedings of the 6th International Academic Conference on Management and Economics*.

Assembly, G. (2015), *Resolution adopted by the General Assembly on 19 September 2016*, Online: 3.10.2016 *(The New York Declaration), Tech. Rep*.

Busco, C.; Granà, F and Izzo, M.F. (2018), *Sustainable development goals and integrated Reporting*, Routledge: London, UK.

Costanza, R., McGlade, J., Lovins, H. and Kubiszewski, I. (2014), An overarching goal for the UN sustainable development goals, *Solutions*, 5(4), 13-16.

De Jong, M. G., Steenkamp, J.-B. E. M., Fox, J.-P. and Baumgartner, H. (2008), Using item response theory to measure extreme response style in marketing research: A global investigation, *Journal of Marketing Research*, 45, 104–115.

Delmas, M. A. and Burbano, V. C. (2011), The Drivers of Greenwashing, *California Management Review*, 54(1), 64-87.

El Zein A. S., Consolacion-Segura C., and Huertas-Garcia R. (2020), The role of sustainability in brand equity value in the financial sector, *Sustainability*, 12(1), 254.

Ewing, M. T., Salzberger, T. and Sunkovic, R. R. (2005), An alternate approach to assessing cross-cultural measurement equivalence in advertising research, *Journal of Advertising*, 34, 17–36.

Filho L., W., Tripathi, S. K., Andrade Guerra, J. B. S. O. D., Giné-Garriga, R., Orlovic Lovren, V. and Willats, J. (2019), Using the sustainable development goals towards a better understanding of sustainability challenges, *International Journal of Sustainable Development & World Ecology*, 26(2), 179-190.

Fleming, A., Wise, R. M., Hansen, H. and Sams, L. (2017), The sustainable development goals: A case study, *Marine Policy*, 86, 94-103.

Fonseca, L. M., Domingues, J. P. and Dima, A. M. (2020), Mapping the sustainable development goals relationships, *Sustainability*, 12(8), 3359.

Imperatives, S. (1987), Report of the World commission on environment and development: Our common future, *http://www.ask-force.org/web/Sustainability/Brundtland-Our-Common-Future1987-2008.pdf*

Jain, P. and Jain, P. (2020), Are the sustainable development goals sustainable? A policy perspective, *Sustainable Development*, 28(6), 1642-1651.

Jones, P., Hillier, D. and Comfort, D. (2016), The sustainable development goals and the financial services industry. *Athens Journal of Business &Amp; Economics*, 3(1), 37-50.

Kamakura, W. A. and Balasubramanian, S. (1989), Tailored interviewing: an application of item response theory for personality measurement, *Journal of Personality Assessment*, 53, 502–519.

Katila, P., Colfer, C. J. P., De Jong, W., Galloway, G., Pacheco, P. and Winkel, G. (Eds.). (2019), *Sustainable development goals*, Cambridge University Press.

Mair, S., Jones, A., Ward, J., Christie, I., Druckman, A. and Lyon, F. (2018), A critical review of the role of indicators in implementing the sustainable development goals, *Handbook of sustainability science and research*, 41-56.

Moyer, J. D. and Hedden, S. (2020), Are we on the right path to achieve the sustainable development goals?. *World Development*, 127, 104749.

Pogge, T. and Sengupta, M. (2015), The sustainable development goals: A plan for building a better world, *Journal of Global Ethics*, 11(1), 56-64.

Raykov, T., Calantone, R.J. (2014), The utility of item response modeling in marketing research. *J. of the Acad. Mark. Sci.* 42, 337–360.

Robert, K. W., Parris, T. M. and Leiserowitz, A. A. (2005), What is sustainable development? Goals, indicators, values, and practice, *Environment: Science and policy for sustainable development*, 47 (3), 8-21.

Sachs, J. D. (2012), From millennium development goals to sustainable development goals, *The Lancet*, 379(9832), 2206-2211.

Sachs, J. D., Schmidt-Traub, G., Mazzucato, M., Messner, D., Nakicenovic, N. and Rockström, J. (2019), Six transformations to achieve the sustainable development goals, *Nature Sustainability*, 2(9), 805-814.

Schultz, C., Salomo, S. and Talke, K. (2013), Measuring new product portfolio innovativeness: how differences in scale width and evaluator perspectives affect its relationship with performance, *Journal of Product Innovation Management,* 30, 93–109.

Singh, Z. (2016), Sustainable development goals: Challenges and opportunities, *Indian Journal of Public Health*, 60(4), 247-250.

Stafford-Smith, M., Griggs, D., Gaffney, O., Ullah, F., Reyers, B., Kanie, N. and O'Connell, D. (2017). Integration: The key to implementing the sustainable development goals, *Sustainability Science*, 12, 911-919.