ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Research Article; 2023; 14(4); 508-523

Modelling the Differences in Social and Emotional Skills with Polytomous Explanatory IRT: The Example of Assertiveness Skill*

Fatma Nur AYDIN **   Kübra ATALAY KABASAKAL ***

Abstract

Explanatory item response theory models can simultaneously decompose the covariance between persons and items, as well as analyze items by adding item-related predictors for differences between item difficulties and/or person-related predictors for differences between individuals. In the current study, we calculated the parameter estimations regarding the skill of assertiveness according to the rating scale model and partial credit model, which are descriptive (traditional) item response theory models as well as latent regression partial credit model including only person-level predictors, and then examined the results comparatively. We used the raw score belonging to the skill of assertiveness of Türkiye belonging to the OECD Social and Emotional Skills Study, and we included gender, socioeconomic level, perceived relationships with teachers, bullying at school, sense of belonging at school, global mindedness, and test anxiety as person-level predictors. Current study findings suggest that; (1) the latent regression partial credit model produces a better data fit when compared to the rating scale model and partial credit model, and (2) sense of belonging at school, global mindedness, and socioeconomic level are significant predictors to explain the differences between persons. We discussed the current study findings in terms of the rich body of knowledge provided by explanatory item response theory and presented some suggestions.

Keywords: polytomous IRT models, polytomous explanatory IRT models, social and emotional skills, assertiveness

Introduction

Item response theory (IRT) is a mathematical theory of measurement that indicates that it is possible to establish a relationship between one's performance in a test, and latent traits or abilities assumed to underlie this performance (Hambleton & Swaminathan, 1985). IRT aims to make inferences about the features measured by a test (Baker, 2016). There are various IRT models such as those including items in two categories (Rasch model; one, two, three parameter models) (Embretson & Reise, 2000) and polytomous models (partial credit model (Masters, 1982), rating scale model (Andrich, 1978a; 1978b), graded response model (Samejima, 1969). When IRT models are considered within the framework of generalized linear mixed models and non-linear mixed models, it is possible to get descriptive and explanatory models (De Boeck & Wilson, 2004). If a model explains item qualities with parameters such as item difficulty and item discrimination, while it explains individuals' performances in terms of ability scores, it is called a "descriptive" measurement model (De Boeck & Wilson, 2004). The aforementioned traditional IRT models are considered to be descriptive models. On the other hand, generalized linear mixed models allow IRT models to be addressed with a multi-level approach. According to that, responses to items are dependent on the emerging hierarchical structure. In other words, responses to items are addressed as repetitive measurements nested in individuals (De Boeck & Wilson, 2004). It is possible to estimate the impact of predictor variables with the help of such a

* A part of this study was presented as an oral presentation at the 8. Congress of Measurement and Evaluation in Education and Psychology (CMEEP)

**Teacher., Republic of Türkiye Ministry of National Education, Ankara-Türkiye, fatmanuraydin.2012@gmail.com, ORCID ID: 0000-0003-0887-395X
*** Assoc. Prof., Hacettepe University, Faculty of Education, Ankara-Türkiye, kkatalay@gmail.com, ORCID ID: 0000-0002-3580-5568

To cite this article:
Aydın, F.N., Kabasakal Atalay, K. (2023). Modelling the differences in social and emotional skills with polytomous explanatory IRT: The example of assertiveness skill. *Journal of Measurement and Evaluation in Education and Psychology, 14*(4), 508-523. https://doi.org/10.21031/epod.1299468

Received: 19.05.2023
Accepted: 21.11.2023

modelling approach (Stanke & Bulut, 2019). Explanatory item response theory models (De Boeck & Wilson, 2004) refer to models that analyse items by focusing on differences between item difficulties through adding item-related predictors and/or focusing on differences between individuals through adding person-related predictors as well as dissociating the common variance between person and items at the same time (Briggs, 2008; De Boeck & Wilson, 2004). Although studies in the literature often use descriptive measurement models that make it possible to find an answer to many problems, these models do not provide information on systematic effects that can explain observations, and so they cannot explain the common variability among persons and items (De Boeck & Wilson, 2004; Stanke & Bulut, 2019). However, explanatory IRT models can meet this demand, and these models can be divided into four according to the predictors they include. According to that, if the model does not include a predictor at the level of person or item, it is called "doubly descriptive (i.e. traditional IRT models)"; if it includes a predictor at the person level, it is called "person explanatory (i.e. the latent regression Rasch model)"; if it includes a predictor only at the level of item, it is called "item explanatory (i.e. linear logistic test model)"; if it includes a predictor both at the level of person and item, it is called "doubly explanatory (i.e. the latent regression linear logistic test model)"(De Boeck & Wilson, 2004). Various studies are using these models in the literature. Atar and Çobanoğlu Aktan (2013) added gender, positive attitude towards science, the importance given to science, self-confidence towards science and parents' education level to the model as individual-level predictors to explain the differences between student achievement. Demirkol and Ayvallı Karagöz (2023) compared various explanatory IRT models in which item format and cognitive domain level of items were added as predictors to explain the differences in item difficulty parameters. Demirkol and Kelecioğlu (2022) examined the item position effect and its interaction with various student characteristics (gender, SES, test anxiety, achievement motivation) in a test in reading. Stanke and Bulut (2019) examined individuals' reactions to items by adding various predictor variables at the item level for the verbal aggression (Vansteelandt, 2000) dataset. In the study, type of behaviour (curse, scold, or shout), type of blame (others or self), and blame mode (want or do) were used as item-level explanatory variables. Atar (2011) established explanatory and descriptive IRT models in her study. Accordingly, the variables of gender, positive attitude towards mathematics, giving importance to mathematics and self-confidence in learning mathematics were used as individual-level characteristics in the study. In the same study, two different predictors were used as item-level predictors: cognitive domain (knowledge, application, reasoning) and subject area (numbers, algebra, data analysis and probability, geometry).

Studies on IRT mostly focus on data with items scored in dichotomous form. However, it is more common to use polytomous measurement tools when it comes to ability tests or scales. Polytomous data give more information on response patterns and more detailed insight into the construct to be measured as well as measurement tools (De Boeck & Wilson, 2004; Stanke & Bulut, 2019). It is necessary to use appropriate models to analyze the data that can reveal this extra information. For example, data loss will be inevitable when polytomous data is turned into one having two categories and then analysed (De Boeck & Wilson, 2004; Stanke & Bulut, 2019). The analysis given in the book written by De Boeck and Wilson (2004) about explanatory IRT can be an example of this. In related examples, they compared the fit of various descriptive and explanatory IRT models with the verbal aggression (Vansteelandt, 2000) dataset converted into a dichotomous response format (Wilson & De Boeck, 2004) and its polytomous version (Tuerlinckx & Wang, 2004). Firstly, results obtained from the partial credit model were compared to the latent regression Rasch model, and the coefficients were found to be almost equal, while standard errors regarding the coefficients were found to be 60% more in the dichotomous data set. According to another finding, in the dichotomous data set, gender was not a significant variable, whereas it was found to be significant in the polytomous data set (Tuerlinckx & Wang, 2004). On the other hand, Kim and Wilson (2019) conducted a study to extend the linear logistic test model approach, and they developed two different item explanatory models to find out that polytomous item explanatory IRT models could contribute to test development processes more than descriptive models due to the information they provided on the content of the items.

Literature review shows that there is a limited number of studies on polytomous explanatory IRT (i.e. Kahraman, 2014; Kim & Wilson, 2019, Stanke & Bulut, 2019; Tuerlinckx & Wang, 2004). Tuerlinckx and Wang (2004) conducted the study that founded the basis to formalize and interpret explanatory IRT models via a polytomous data set. In that study, the researchers analysed the data set on verbal aggression via partial credit model, person explanatory partial credit model, rating scale model, person explanatory rating scale model, and explanatory partial credit model including person and item-level predictors. Study findings show that person-explanatory models display a better fit when compared to their traditional counterparts. On the other hand, Stanke and Bulut (2019) conducted a study to make a new parameterization that made it possible to explain the distances between the thresholds by flexing the formulas regarding polytomous item response theory models. In that study, they analysed data sets on verbal aggression via rating scale model, partial credit model, explanatory partial credit model and cross-classified explanatory partial credit model. The study findings show that the partial credit model resulted in a better fit according to the AIC value, while the cross-classified explanatory partial credit model resulted in a better fit according to the BIC value. Another study on polytomous explanatory IRT models was carried out by Kim and Wilson (2019), who developed two different item explanatory IRT models. In that study, the researchers analyzed two different data sets (carbon cycle and verbal aggression). Results of the analysis conducted with the data set on the carbon cycle show that the explanatory many-facet Rasch model resulted in a better fit, while the researchers could not reach a definite result at the end of the analysis conducted with verbal aggression according to AIC and BIC values. Another study which made use of the explanatory IRT approach was conducted by Kahraman (2014). That study which used the data obtained from a performance test in the field of medicine took the advantage of partial credit model to compare explanatory IRT models to which various predictor variables (gender, time to respond, number of the items, test score) were added individually. The study results show that the explanatory partial credit model to which the test score was added as a predictor displayed a better fit to the data.

According to Stanke and Bulut (2019), polytomous explanatory IRT models have mostly focused on the first threshold between item response categories (e.g., Tuerlinckx & Wang, 2004). However, possible variances between the thresholds can be ignored in such a case (Stanke & Bulut, 2019). Therefore, Stanke and Bulut (2019) added a new parameter and flexed the explanatory IRT models. In this context, the log-odd of response in category j instead of $j-1$ given by the individual n for the item i is written as below according to the explanatory partial credit model employed in the current study:

$$\log\left(\frac{P_n(j)}{P_{n(j-1)}}\right)= Z_n\theta_n - X_n'\delta_i + W_n'\tau_{ii}$$

Here, Z_n represents a matrix used to estimate fixed and random effects related to personal traits θ_n refers to the level of latent qualities of a person, and it has a normal distribution ($N(\mu_n, \sigma_n^2)$). X_n' is a matrix of item-related information that describes the characteristics of individual items. δ_i is the position of the first threshold between the first and second response categories for the item i. W_n' is a matrix that is used to estimate the fixed and random effects regarding the distances between the thresholds. τ_{ii} refers to the distance between the threshold of $(j-2)/(j-1)$ and $(j-1)/j$ for the item i (Stanke & Bulut, 2019). In the current study, in addition to the explanatory partial credit model, predictions were also made according to the rating scale model (RSM) and the partial credit model (PCM). Accordingly, the model equation for RSM and PCM is given below (Stanke and Bulut, 2019):

$$\log\left(\frac{P_n(j)}{P_{n(j-1)}}\right) = \theta_n - (\delta_i + \tau_{ii})$$

In this equation, θ_n and δ_i have the same meaning as the explanatory partial credit model. The only difference between RSM and PCM is that τ_{ii} is the same for all items in RSM (Stanke and Bulut, 2019).

Social and Emotional Skills

Today, continuously changing social, economic and environmental conditions lead to changes in individuals' lives and the flow of social activities. As globalization and digitalization connect people, the world has become a more complicated place full of uncertainties. The content of these skills that are necessary to be successful in such a world and adapt to these changes are also changing in time (Kankaraš & Suarez-Alvarez, 2019; OECD, 2021a). Cognitive skills that are commonly associated with academic achievement are thought to be of prime importance. These skills are very significant as they increase the likelihood of people getting positive results in later life by making them competent in many social and emotional skills such as perseverance, sociability, and self-respect. In today's world, for individuals to become competent in transforming skills, it is important to measure social and emotional skills that interact with cognitive skills in education systems and to take initiatives to support the development of skills accordingly (OECD, 2015). Social and emotional skills are known to be effective in many fields such as academic achievement, productivity in work life or subjective well-being. A high level of social and emotional skills increases trust and tolerance in society, and they lead to a decrease in criminal and anti-social behaviours (Kankaraš & Suarez-Alvarez, 2019; OECD, 2021a). Accepted to be one of the most comprehensive evaluations of these skills in the international arena, the OECD Social and Emotional Skills Study (2021a) was conducted to identify the factors that support or hinder students' social and emotional skills. The study findings were intended to provide the shareholders of education with reliable information (OECD, 2021b). The social and emotional skills specified in that study are described as "individual capacities that can be (a) manifested in consistent patterns of thoughts, feelings and behaviours, (b) developed through formal and informal learning experiences, and (c) important drivers of socioeconomic outcomes throughout the individual's life" (OECD, 2015, p.35). In that study, which was conducted at an international level, the theoretical framework of social and emotional skills relied on the basic components of "big five" personality traits to develop "big five social and emotional skills model." In this context, the basic skills were listed as engaging with others (sub-domains; assertiveness/dominance, sociability, energy/enthusiasm), task performance (sub-domains; persistence, self-control/self-discipline, responsibility/trustworthiness), emotional regulation (sub-domains; optimism/positive emotion, stress resistance/resilience vs. anxiety, emotional control), collaboration (sub-domains; empathy/compassion, trust, co-operation/relationship harmony), open-mindedness (sub-domains; creativity/imagination, tolerance/cultural flexibility, intellectual curiosity) (Kankaraš & Suarez-Alvarez, 2019). The assertiveness skill examined in the current study is associated with expressing one's ideas, feelings and needs responsibly and liking leadership. Individuals having this skill can express their thoughts directly when they disagree with others, and they do not need the guidance of others (Kankaraš & Suarez-Alvarez, 2019). Because of that, being assertive plays an important role in increasing one's level of well-being, while emphasizing individual rights at the same time (Eskin, 2003). As individuals with this skill can more clearly reveal their will (Kankaraš & Suarez-Alvarez, 2019), they can find a solution for their problems at work, school or home more rationally and appropriately. Those who have a high level of assertiveness will have a higher level of self-confidence as well as better decision-making skills, and they will be able to deal with negative feelings such as anger more healthily. This will also have a positive impact on school performance (Sitota, 2018).

It is important to try to explain individual differences as to the social and emotional skills that have an important role in people's family, work and school life and that are also related to cognitive skills. Literature review shows that studies that employed an explanatory IRT mostly focused on cognitive skills (see. Atar, 2011, Atar & Çobanoğlu Aktan, 2013; Briggs, 2008; Büyükkıdık & Bulut, 2022; Demirkol & Ayvallı Karagöz, 2023; Demirkol & Kelecioğlu, 2022; Kahraman, 2014; Kim & Wilson, 2020). However, we think that identifying the variability of social and emotional skills among individuals will make it possible to understand the construct better and accordingly to give more effective feedback to individuals at schools. On the other hand, when the studies that employed an explanatory IRT were evaluated in terms of the qualities of the data set, it was clear that studies mostly employed dichotomous data (see. Atar, 2011; Atar & Çobanoğlu Aktan, 2013; Briggs, 2008; Büyükkıdık & Bulut, 2022). However, most measurement tools that measure latent features have a polytomous data format. In that sense, it is important to make use of explanatory IRT models appropriate for polytomous data sets. Also, the studies on polytomous explanatory IRT generally ignored the distances between the thresholds and focused only on the estimation of the first threshold parameter. This can cause data loss due to ignoring the possible variability between the thresholds (Stanke & Bulut, 2019). In this context, the current study aims at identifying the individual differences in an affective skill via a polytomous explanatory IRT model that makes it possible to estimate the distances between the thresholds. In addition, the explanatory IRT model used is analyzed in comparison with the predictions obtained from RSM and PCM. In line with the study purpose, the research questions are as below:

How are the parameter estimations obtained from the rating scale model regarding the skill of assertiveness?

How are the parameter estimations obtained from the partial credit model regarding the skill of assertiveness?

How are the parameter estimations obtained from the latent regression partial credit model regarding the skill of assertiveness?

What are the significant predictors at the personal level regarding the skill of assertiveness?

How are the results of model-data fit regarding the rating scale model, partial credit model and latent regression partial credit model?

Method

In the current study, we comparatively investigated individual differences regarding the skill of assertiveness within the scope of polytomous descriptive and explanatory IRT models. Therefore, we employed a descriptive study model (Büyüköztürk et al., 2014).

Participants

This study uses data collected in Türkiye as part of the OECD Social and Emotional Skills Study conducted in 2019. Only Istanbul was included in this survey conducted by the OECD. A total of 5869 individuals in the age group of 10 and 15 years participated in the study. We conducted the analysis using the responses of individuals included in the age group of 15 (n=3168), who responded to the items in the Assertiveness Scale. We used the responses given by 2968 individuals after examining the data set in terms of missing data and outliers. The study group included 1764 (59.43%) female and 1204 (40.57%) male students.

Measurement Tools

In the current study, we used the raw data obtained from the sub-scale of assertiveness measured within the scope of the OECD Social and Emotional Skills Study as well as the indices calculated using the raw data. Social and emotional skills were measured using questionnaires administered to the student, the teacher, and the parents. With the raw data from the questionnaires, indices representing different characteristics were calculated (Kankaraš & Suarez-Alvarez, 2019, OECD, 2021b).

In the current study, we used a 5-point (strongly disagree, disagree, neutral, agree, strongly agree) Likert-type scale including 8 items (Item 5 is reverse coded) for the skills of assertiveness, while we did not include item number 7 as it was found to be statistically insignificant in estimations conducted via IRT models. This item was also excluded in the study conducted by OECD as it produced a high tau (slope) value, and it displayed a duplication with item number 6 (OECD, 2021b). The indices we used as a predictor at the person level were perceived relationships with teachers, bullying at school, sense of belonging at school, global mindedness, and test anxiety. Furthermore, the other two predictor variables of the study were gender and socio-economic level. Table 1 below gives the items of the Assertiveness Scale used in the study.

Table 1.
Scale Items

Item Number	Item Content
1	A leader
2	Want to be in charge
3	Know how to convince others to do what I want
4	Enjoy leading others
5	Dislike leading a team
6	Like to be a leader in my class
7	-
8	Dominant, and act as a leader

Table 1 above presents information about the content of scale items as explained in the technical report published by OECD. As item number 7 was excluded from the study at the end of the analysis, the report did not include content about this item. Table 2 below shows the content of the variables that were used as predictors in the study (see. OECD, 2021b).

Table 2.
Content about the Person-Level Predictors

Name of the Index	Question	Alternatives
Perceived relationships with teachers	During the past 12 months, how often did you have the following experiences at school?	Most of my teachers treated me fairly. I got along well with most of my teachers. Most of my teachers were interested in my well-being.
Bullying at school	During the past 12 months, how often have you had the following experiences in school?	Other students made fun of me. I was threatened by other students. Other students took away or destroyed things that belonged to me. I got hit or pushed around by other students.
Sense of belonging at school	Thinking about your school: To what extent do you agree with the following statements?	I feel like an outsider (or left out of things) at school. I make friends easily at school. I feel like I belong at school. I feel awkward and out of place in my school. Other students seem to like me. I feel lonely at school.
Global mindedness	How informed are you about the following topics?	Climate change and global warming Global health (e.g. epidemics) International conflicts Causes of poverty Equality between men and women in different parts of the world
Test anxiety	To what extent do you agree or disagree with the following statements about yourself?	I often worry that it will be difficult for me to take a test. Even if I am well prepared for a test I feel very anxious. I get very tense when I study for a test.

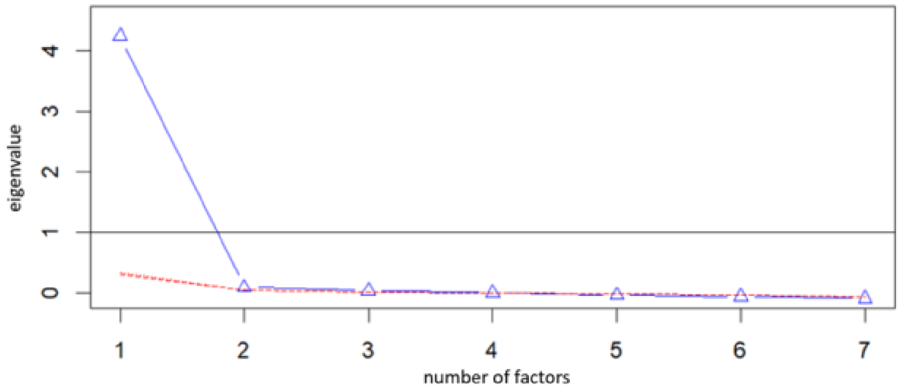
Data Analysis

In the current study, we conducted data cleaning and checked the assumptions on R software (R Core Team, 2022) via the packages of haven (Wickham et al., 2022), stringr (Wickham, 2022), olsrr (Hebbali, 2020), dplyr (Wickham et al., 2022), ltm (Rizopoulos, 2006), psych (Revelle, 2022), MVTests (Bulut, 2019), ggplot2 (Wickham, 2016), mirt (Chalmers, 2012). For estimations relating to the partial credit model, rating scale model, and latent regression partial credit model, the eirm package (Bulut, 2021) that is also included in R software was used. We examined the data and checked the assumptions as stated below:

We excluded 37 individuals based on missing value analysis and 163 individuals based on univariate and multivariate outlier analysis. We then conducted a multicollinearity analysis. Simple pairwise correlation values between independent variables (ranging from -0.261 to 0.291), tolerance values (ranging from 0.826 to 0.961) and variance influence factor values (ranging from 1.039 to 1.209) showed that there was no multicollinearity problem. In the final case, we continued the analyses with data from 2968 individuals. On the other hand, Cronbach's alpha value calculated for reliability was found to be 0.603[1]. For Cronbach alpha, a value equal to or higher than 0.70 is acceptable, while a value higher than 0.80 implies a high level of reliability (Nunnally & Bernstein, 1994). Hence, it is possible to state that the reliability value calculated in the current study was low. This might result from the length of the test which is a factor that affects reliability. It is known that a short test affects reliability in a negative way (Crocker & Algina, 1986). In the current study, the reliability value for the data set belonging to the 7 items was low, which we think might be due to the shortness of the test[2].

Finally, we tested the assumptions of unidimensionality and local independence, which are necessary assumptions for item response theory analyses. First, we conducted an exploratory factor analysis and parallel analysis to test the unidimensionality assumption. To simplify the narrative, only parallel analysis results are given[3]. For this analysis, we used the function of fa.parallel in the package psych of R software (Revelle, 2022). Figure 1 below shows the related results.

Figure 1.
Results of parallel analysis



In Figure 1, the blue line refers to values regarding the real data, and the red line refers to values regarding the data produced randomly. One of the factors obtained from the real data set has an eigenvalue noticeably higher than the eigenvalues of data produced randomly. In this case, it is possible to state that the scale has only one factor. When all the results are taken into consideration, the factor loads of items of the scale which was decided to be unidimensional are 0.811; 0.549; 0.428; 0.904; -0.877; 0.877; 0.862 respectively. Secondly, to test the local independence assumption, we performed parameter estimations for the rating scale model and the partial credit model. We examined correlation

[1] The Cronbach's alpha value calculated without removing the 7th item from the data set was 0.725.
[2] In the technical report published by the OECD, both alpha and omega coefficients for the assertiveness subscale for the 15-year-old age group are reported as 0.88 (OECD, 2021b).
[3] Exploratory factor analysis (EFA) was also conducted to examine the unidimensionality assumption. According to the EFA results, the assertiveness subscale was found to be unidimensional (eigenvalue of the first factor=4.550, variance explained=61%).

values between the residuals. According to the results obtained from the rating scale model, these values range between -0.789 and 0.421, and the results obtained from the partial credit model range between -0.670 and 0.408. As stated by Christensen et al. (2016), the studies in the literature mostly use the critical value of 0.2 suggested by Chen and Thissen (1997) for local independence. The values higher than this value are said to violate local independence. However, other critical values are also used (Christensen et al., 2016). In this context, Christensen et al. (2016) mentioned some studies in which critical values of 0.1; 0.3; 0.5 and even 0.7 were used. On the other hand, it is stated that local independence can be violated as personality assessments include very similar items (Steinberg & Thissen, 1996, cited by Embretson & Reise, 2000, p.232). According to that, considering all the studies conducted previously, we concluded that the assumption of local independence in the scale was not violated.

Preparing the Data Set for Analysis

In this part, the procedures for making the data set suitable for analysis are explained. Estimations regarding dichotomous and polytomous explanatory IRT models can be done via package eirm (Bulut, 2021) which conducts transactions through function glmer in package lme4 (Bates et al., 2015). In explanatory IRT models, as items are nested in persons, it is necessary to turn data into a long format in which there are answers about one item in each line and each person has more than one line. On the other hand, to use the functions in the package, the responses should display binominal distribution, and so it should be dichotomous. Therefore, the responses should first be transformed into multiple dichotomous formats to analyse polytomous data. These two processes can be done with the function polyreformat (Bulut et al., 2021). Table 3 gives information on multiple dichotomous coding conducted to analyse the 5-category data set.

Table 3.
Transforming Polytomous Responses into Multiple Dichotomous Responses

Original response	"I disagree"	"Neutral"	"I agree"	"I strongly agree"
Strongly disagree	0	NA	NA	NA
Disagree	1	0	NA	NA
Neutral	NA	1	0	NA
Agree	NA	NA	1	0
Strongly agree	NA	NA	NA	1

According to the recording given in Table 3, for example, for an individual who responded to an item in the strongly disagree category, the responses transformed into multiple dichotomous responses were recoded as 0 in the "disagree" category and NA in the other categories. Table 4 below shows the R codes used in the study.

Table 4.
R codes used in the Study

Model	R codes
Rating Scale Model	rsm <- eirm(formula = "polyresponse ~ -1 + item + polycategory + (1|person)", data = long_data)
Partial Credit Model	pcm <- eirm(formula = "polyresponse ~ -1 + item + item:polycategory + (1|person)", data = long_data)
Latent Regression Partial Credit Model	lrm_poly <- eirm(formula = "polyresponse ~ -1 + item + item:polycategory + gender+relteach+ bully+belong+ global+ anxtest+ SES+ (1|person)", data = long_data)

Results

In this part, we explained the results obtained from the rating scale, partial credit, and latent regression item response theory model according to the research questions. The tables show the estimations regarding the easiness parameters.

Table 5.

Estimations Regarding Rating Scale Model

Item Number	**Location for Strongly Disagree/Disagree**			**Distance for Disagree/Neutral****			**Distance for Neutral/Agree**			**Distance for I agree/I strongly agree**		
	Estimation	SE	Z value	Estimation	SE	Z value	Estimation	SE	Z value	Estimation	SE	Z value
1	0.580	0.036	16.072*									
2	0.804	0.037	21.512*									
3	1.130	0.040	28.392*									
4	0.732	0.037	19.781*	-0.561	0.033	-17.196*	-0.689	0.033	-21.149*	-1.195	0.034	-34.729*
5	0.325	0.035	9.315*									
6	0.681	0.037	18.592*									
7	0.614	0.036	16.953*									

*p<0.05

**In empty cells, the values written in the relevant columns are repeated. For clarity, we have not repeated the same values.

Table 5 gives the results obtained from the rating scale model for the sub-scale of assertiveness. According to that, item number 3 (Know how to convince others to do what I want) with logit 1.130 is the item which has the easiest likelihood to respond in a higher category according to location parameter estimation for the threshold of “I strongly disagree/I disagree”. For this item, it seems exp(1.130)= 3.095 times easier to respond in the category of “I disagree” instead of “I strongly disagree”. In the rating scale model, distance value is estimated the same in all items for each category. According to the distance value obtained for the threshold of “I disagree/I am neutral”, after checking the level of assertiveness, it is exp(-0.561)=0.570 times easier for items to respond in the category of “I am neutral” than the categories of “I strongly disagree” and “I disagree”. According to the distance value obtained for the threshold of “I am neutral/I agree”, after checking the level of assertiveness, it is exp(-0.689)=0.502 times easier for items to respond in the category of “I agree” than the categories of “I strongly disagree” and “I disagree”. Lastly, according to the distance value obtained for the threshold of “I agree/I strongly agree”, after checking the level of assertiveness, it is exp(-1.195)=0.302 times easier for items to respond in the category of “I strongly agree” than the categories of “I strongly disagree” and “I disagree”.

Table 6.

Estimations Regarding Partial Credit Model

Item Number	Location for Strongly Disagree/Disagree			Distance for Disagree/Neutral			Distance for Neutral/Agree			Distance for I agree/I strongly agree		
	Estimation	SE	Z value	Estimation	SE	Z value	Estimation	SE	Z value	Estimation	SE	Z value
1	0.563	0.061	9.289*	-0.562	0.080	-7.056*	-0.703	0.081	-8.694*	-1.086	0.088	-12.306*
2	1.525	0.098	15.600*	-1.092	0.111	-9.827*	-1.512	0.108	-13.946*	-2.288	0.114	-20.070*
3	1.485	0.145	10.215*	-0.434	0.162	-2.672*	-1.082	0.153	-7.073*	-1.765	0.152	-11.578*
4	0.490	0.064	7.644*	-0.636	0.087	-7.355*	-0.106	0.085	-1.254	-0.842	0.084	-10.001*
5	0.204	0.050	4.054*	-0.521	0.072	-7.193*	-0.577	0.080	-7.231*	-0.597	0.090	-6.605*
6	0.661	0.065	10.179*	-0.606	0.084	-7.242*	-0.735	0.084	-8.770*	-0.989	0.088	-11.259*
7	0.620	0.062	9.963*	-0.605	0.081	-7.471*	-0.734	0.082	-8.972*	-1.094	0.088	-12.400*

*p<0.05

Table 6 gives the results obtained according to the partial credit model for the sub-scale of assertiveness. There is one thing to be careful about while interpreting the distance values. According to that, while calculating threshold values except for the location parameter of the related item, it is necessary to add the distance value estimated in each category and the value estimated for the location parameter (1st threshold). For instance, the location value for the first item is 0.563. For the same item, the estimated distance value for the category of “I disagree/I am neutral” is -0.562. This value points to the distance between the first and second thresholds for the first item. When these two values are added (0.563+(-0.562)), the second threshold is obtained. In other words, the first threshold value is taken as a reference to make calculations to find each threshold. According to that, item number 2 (Want to be in charge) with logit 1.525 is the item which has the easiest likelihood of responding in a higher category. For this item, it seems exp(1.525)= 4.595 times easier to respond in the category of “I disagree” instead of “I strongly disagree”. Unlike the rating scale model, the distance value is estimated differently in all items for each category when it comes to the partial credit model. According to the distance value obtained for the threshold of “I disagree/I am neutral”, after checking the level of assertiveness, it is exp(0.563+(-0.562)) =exp(1.051)=2.860 times easier for the third item to respond in the category of “I am neutral” than the categories of “I strongly disagree” and “I disagree”. According to the distance value obtained for the threshold of “I am neutral/I agree”, after checking the level of assertiveness, it is exp (0.403) =1.496 times easier for the third item to respond in the category of “I agree” than the categories of “I strongly disagree” and “I disagree”. Lastly, according to the distance value obtained for the threshold of “I agree/I strongly agree”, after checking the level of assertiveness, it is exp (-0.281)=0.755 times easier for the third item to respond in the category of “I strongly agree” than the categories of “I strongly disagree” and “I disagree”.

Table 7.

Estimations Regarding Latent Regression Partial Credit Model

Item Number	Location for Strongly Disagree/Disagree			Distance for Disagree/Neutral			Distance for Neutral/Agree			Distance for I agree/I strongly agree		
	Estimation	SE	Z value	Estimation	SE	Z value	Estimation	SE	Z value	Estimation	SE	Z value
1	-0.354	0.126	-2.810*	-0.580	0.080	-7.264*	-0.751	0.081	-9.245*	-1.180	0.089	-13.289*
2	0.617	0.147	4.199*	-1.109	0.111	-9.971*	-1.561	0.108	-14.390*	-2.379	0.114	-20.827*
3	0.589	0.182	3.231*	-0.453	0.163	-2.784*	-1.126	0.153	-7.343*	-1.849	0.153	12.081*
4	-0.430	0.128	-3.368*	-0.650	0.087	-7.494*	-0.139	0.085	-1.633	-0.914	0.085	-10.792*
5	-0.781	0.124	-6.321*	-0.482	0.073	-6.634*	-0.520	0.080	-6.485*	-0.530	0.091	-5.838*
6	-0.253	0.128	-1.980*	-0.622	0.084	-7.415*	-0.770	0.084	-9.267*	-1.079	0.088	-12.214*
7	-0.298	0.127	-2.349*	-0.619	0.081	-7.615*	-0.778	0.082	-9.465*	-1.182	0.089	-13.326*

*p<0.05

Table 7 gives the results obtained from the latent regression partial credit model for the sub-scale of assertiveness. Distance values are calculated as in the partial credit model. According to that, considering the location parameters values, item number 2 (Want to be in charge) with logit 0.617 is the item which has the easiest likelihood to respond in a higher category. For this item, it seems exp (0.617)=1.53 times easier to respond in the category of “I disagree” instead of “I strongly disagree”. According to the distance value obtained for the threshold of “I disagree/I am neutral”, after checking the level of assertiveness, it is exp(0.135)=1.144 times easier for the third item (Know how to convince others to do what I want) to respond in the category of “I am neutral” instead of the categories of “I strongly disagree” and “I disagree”. According to the distance value obtained for the threshold of “I am neutral/I agree”, after checking the level of assertiveness, it is exp(-0.538)=0.583 times easier for the third item to respond in the category of “I agree” than the categories of “I strongly disagree” and “I disagree”. Lastly, according to the distance value obtained for the threshold of “I agree/I strongly agree”, after checking the level of assertiveness, it is exp(-1.260)=0.283 times easier for the third item to respond in the category of “I strongly agree” than the categories of “I strongly disagree” and “I disagree”.

Table 8 below shows the estimation regarding the predictors in the latent regression partial credit model in which analysis is conducted by including predictors at the person level.

Table 8.

Estimations Regarding Predictors

Predictors	b	SE	exp (b)	Z value
Gender	0.045	0.023	1.046	1.929
Perceived relationships with teachers	0.000	0.001	1.000	0.056
Bullying at school	0.002	0.001	1.002	1.660
Sense of belonging at school	0.009	0.001	1.009	8.194*
Global mindedness	0.009	0.001	1.009	8.109*
Test anxiety	-0.001	0.000	0.999	-1.544
Socio-economic level	0.028	0.012	1.028	2.306*

*p<0.05

According to the results, the predictors of a sense of belonging at school, global mindedness and socio-economic level were found to be significant, and the predictors of gender, perceived relationships with teachers, bullying at school and test anxiety were found to be insignificant. According to that, for a 1-unit change in sense of belonging at school, the likelihood of getting a higher score from assertiveness is 1.009 times more. For a 1-unit change in the level of global mindedness, the likelihood of getting a higher score from assertiveness is 1.009 times more. Lastly, for a 1-unit change in the level of socio-economic level, the likelihood of getting a higher score from assertiveness is 1.028 times more. In other words, it is possible to state that individuals with a more advantageous socio-economic level are more likely to have a higher level of assertiveness.

Table 9.

Results of Model-Data Fit

Model	AIC	BIC	Deviation
Rating Scale Model	47996.6	48089.9	47974.6
Partial Credit Model	47680.0	47926.0	47622.0
Latent Regression Partial Credit Model	47502.3	47807.7	47430.3

Table 9 gives the results of model-data fit regarding the three models developed in the current study. As the models were not nested, they were compared to relative fit indices. According to that, the results were examined according to the Akaike Information Criterion (AIC) (Akaike, 1974), Bayesian Information Criterion (BIC) (Schwarz, 1978) and deviation values. Having a small index refers to a better model-data fit. According to that, the latent regression explanatory partial credit model is the one with the lowest AIC, BIC, and deviation values. In other words, the model that best fits the data is the latent regression partial credit model.

Discussion and Conclusion

In the current study, individual differences in assertiveness skills of individuals were examined through the rating scale model, partial credit model and latent regression partial credit model. We interpreted the study findings within the framework of parameter estimations and model-data fit. According to the first study finding which we obtained from the rating scale model and partial credit model, neither of which included a predictor, the partial credit model showed a better fit to the data. According to the distance values obtained from the partial credit model, it was easier for the third item to respond in a category in the 2nd, 3rd and 4th thresholds, while it was easier for the second item to respond in a category in the 1st threshold. Therefore, there was a variance in the distance values. In this context, according to the rating scale model, which makes the same estimations as to all items for each category distance, it is possible to state that the partial credit model gives deeper information and reveals variances between items better. This study finding is supported by a previous study conducted by Tuerlinckx and Wang (2004). In that study, the researchers concluded that the partial credit model provides a better fit than the rating scale model. This study finding is partially in line with the results of the study conducted by Stanke and Bulut (2019). The researchers in that study reported that when the items in the data set of that study were estimated via a partial credit model, the resulting values varied between -0.10 and 1.13. According to that, the researchers compared the estimations of distance values obtained from the partial credit model and rating scale model and concluded that the partial credit model explained the variance between the items more. Also, in the same study, the partial credit model produced a better fit according to the AIC value, whereas the rating scale model produced a better fit according to the BIC value.

Secondly, the current study reveals that the items which are easier to respond to in a higher category are the same according to the partial credit model and latent regression partial credit model. Thus, we concluded in the current study that the results of these two models were overlapping in terms of location and distance value estimation.

According to the third finding of the current study, the best-fitting model for all indices is the latent regression partial credit model. The fact that explanatory IRT models are models that can simultaneously decompose the covariance between persons and items in addition to analyzing items (Briggs, 2008) by adding predictors related to items for differences in item difficulties and/or adding predictors related to persons for differences between persons (De Boeck & Wilson, 2004) was thought to be effective in the emergence of a tendency to better fit the data. Furthermore, it is stated in the literature that one of the reasons why explanatory IRT models have a better fit is the increase in the number of estimated parameters (De Boeck & Wilson, 2004). Stanke and Bulut (2019) came up with a similar finding in their study. In that study, they compared the explanatory partial credit model and cross-classified explanatory partial credit model and found out that the second model produced a better fit. This was thought to result from the latter model including more parameters. When all models developed in the study were compared, they concluded the model that produced the best fit according to AIC value was the partial credit model, while the model that produced the best fit according to BIC value was the cross-classified explanatory partial credit model. Tuerlinckx and Wang (2004) developed the rating scale model, partial credit model and person-explanatory partial credit model, and they compared the values obtained from these models to find out that person explanatory partial credit model produced the best fit of all. On the other hand, Kahraman (2014) found that the explanatory partial score model including the test score variable provided a better fit. Another important finding belongs to a study carried out by Briggs (2008) to explain the differences in achievement in science. In that study, after conducting ability estimation through one multidimensional Rasch model, the researcher conducted linear regression by taking ability scores obtained from these estimations as the dependent variable and race/ethnicity as the independent variable. Then the researcher compared the results of this analysis which was stated to be a two-step approach and the estimations were done via the latent regression Rasch model. It was stated that if the reliability of the test scores is high, the results of the two-step approach and explanatory IRT approach will overlap. On the other hand, when reliability is low-medium, the ability estimates in the two-stage approach will narrow towards the population mean and therefore the regression coefficients obtained in the second stage will be weakened due to measurement error. The researcher suggested using an explanatory IRT approach when the aim is to identify group differences. In addition, it was emphasized that the explanatory approach allows for more detailed interpretations of achievement differences at the group level depending on race/ethnicity. Büyükkıdık and Bulut (2022) conducted a study to model individuals' achievement in science through test, student, and school-related predictors, and they developed explanatory IRT models including the Rasch model and various predictors. The study results showed that the best fit was obtained from the explanatory model including the variables of gender and school type. Atar and Çobanoğlu Aktan (2013) also carried out a study in which they comparatively examined the differences in achievement in science using the latent regression two-parameter logistic model and traditional two-parameter logistic model, and they found that the latent regression two-parameter model produced a better fit.

Fourthly, according to the results of the latent regression partial credit model including person-level predictors, the predictors of sense of belonging at school, global mindedness and socio-economic level were found to be significant, whereas the values regarding the predictors of gender, bullying at school, perceived relationships with teachers and test anxiety were found to be insignificant. According to the result related to school belonging, it was observed that the higher the sense of belonging to the school, the higher the probability of having a higher level of assertiveness. According to the Social and Emotional Skills Turkey results published by OECD (2021a), a statistically significant positive relationship was found between a sense of belonging to school and assertiveness in the 15-year-old age group. On the other hand, for the variable of the level of awareness of global events, which was found to be significant in the study, it was found that the higher the level of awareness, the higher the probability of receiving higher assertiveness scores. However, the report published by OECD (2021a) did not include an explanation of this variable being significant. In the study, it was also found that there was a positive relationship between socioeconomic level and assertiveness scores, which means that those having a more advantageous socioeconomic level are more likely to have a higher score of assertiveness. A similar result was obtained from OECD (2021a) Türkiye results. There are various studies in the literature that focus on the relationship between assertiveness and socioeconomic level. Kılıç (2009), one of these studies, concluded that the level of assertiveness varied statistically significantly according to the perceived economic level. It was found in that study that individuals included in the middle income and over-middle/high-income families had a higher level of assertiveness than those included in under-middle/low-income families at a statistically significant level. These findings can be attributed to the fact that families from a more advantageous socioeconomic background could invest more in the development of their kids' social and emotional skills (OECD, 2021a).

Gender is one of the variables that was not found to be significant in the current study. In the study, the mean score of girls was found to be 21.535, whereas it was 22.379 for boys. Studies on the relationships between assertiveness and gender have various results. Some studies reported that male participants had a higher score of assertiveness than female participants (Karaaslan Arkan, 2016; OECD, 2021a; Sitota, 2018). On the other hand, some other studies concluded that there was no significant difference between assertiveness and gender (Eskin, 2003; Kılıç, 2009). In some studies that concluded that there was no significant difference, it was found that the assertiveness scores of males were slightly higher than females (Castedo et al., 2015; Kaya & Karaca, 2018). It is striking to see that males have a higher level of assertiveness when studies that examine the relationships between assertiveness and gender are in question. Furthermore, the level of significance regarding this relationship varies from study to study. In the current study, the fact that males had slightly higher assertiveness scores than females, but that there was no significant difference between them, overlaps with the findings in the literature to a certain extent. However, we think that the diversity of findings among the studies in the literature may be because each study has a different study group, uses different analysis techniques, etc. Another variable of the current study, perceived relationships with teachers wasn't found to be a significant predictor. However, the OECD (2021a) report indicated that there was a statistically significant positive relationship between the said variable and the skill of assertiveness. On the other hand, Stake et al. (1983) found a significant increase in the self-esteem levels of individuals who thought that they received more positive reactions from their teachers due to their assertive behaviours. In the current study, we found out that there was a negative but insignificant relationship between assertiveness and exposure to bullying, whereas it was stated in the OECD report (2021a) that there was a significant negative relationship between the two variables. Similarly, Keliat et al. (2015) found a low-level statistically significant negative relationship between stories of abuse and assertiveness. Lastly, the variable of test anxiety was not found to be a significant predictor in the current study. A similar result was reported in the OECD (2021a) study.

When the results of the current study are evaluated together, it is seen that the explanatory IRT model used in the study, in line with the literature, shows a better fit and explains the differences between individuals and the variability in the data better. In addition to that, although some of the estimation results about the predictors are in parallel with the literature, there are also some differences. There are some disagreements, especially with the results reported by OECD (2021a). We think that this might result from using different techniques of analysis.

In conclusion, in light of the findings of the current study, we believe that explanatory IRT models can contribute to improving the scope of studies on the latent traits targeted to be measured. For instance, for a polytomous data set, as is the case in the current study, revealing the potential variance between categories can contribute to developing measurement tools more felicitously. Also, identifying the variables that are related to the skill of assertiveness can help enrich educational content and preventive guidance activities that can be presented for this skill at schools. A clearer understanding of the construct and the contextual factors associated with it can support formal and non-formal education activities involving students, teachers, and parents.

Limitations and Suggestions

The study has several limitations. Only personal characteristics were included as predictors in the study. In other studies, both items and personal characteristics can be added as predictors. In addition, the rating scale model, partial credit model and latent regression partial credit models were compared in the study. Different models can be established in other studies. On the other hand, only the responses of 15-year-olds to the assertiveness scale of the OECD Social and Emotional Skills survey were used in the study. Similar studies can be conducted with other sub-skills in the OECD study and/or with the 10-year-old age group. At the same time, studies comparing the two age groups can be conducted. Finally, the eirm

(Bulut, 2021) package was used for the analysis in this study. This package only analyzes explanatory IRT models based on the Rasch model family. Therefore, different explanatory IRT models including discrimination parameters cannot be estimated with this package (Bulut et al., 2021).

Declarations

Conflict of Interest: No potential conflict of interest was reported by the authors.

Ethical Approval: This research study complies with research publishing ethics. Secondary data were used in this study. Therefore, ethical approval is not required.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723. https://doi.org/10.1109/TAC.1974.1100705.

Andrich, D. (1978a). Application of a psychometric model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2*(4), 581-594. https://doi.org/10.1177/014662167800200413.

Andrich, D. (1978b). A rating formulation for ordered response categories. *Psychometrika, 43* (4), 561-573. https://doi.org/10.1007/BF02293814.

Atar, B. (2011). Tanımlayıcı ve açıklayıcı madde tepki modellerinin TIMSS 2007 Türkiye matematik verisine uyarlanması. *Eğitim ve Bilim, 36* (159), 255-269.

Atar, B., & Çobanoğlu Aktan, D. (2013). Birey açıklayıcı madde tepki kuramı analizi: Örtük regresyon iki parametreli lojistik modeli. *Eğitim ve Bilim, 38* (168), 59-68.

Baker, F. B. (2016). *Madde tepki kuramının temelleri*. (N., Güler, Trans Ed.) & (M., İlhan, Trans.). Ankara. https://doi.org/10.14527/9786053185673.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01.

Briggs, D. C. (2008). Using explanatory item response models to analyze group differences in science achievement. *Applied Measurement in Education, 21*(2), 89 - 118. https://doi.org/10.1080/08957340801926086.

Bulut, H. (2019). An R Package for Multivariate Hypothesis Tests: MVTests. *Technological Applied Sciences, 14* (4), 132-138. https://doi.org/10.12739/NWSA.2019.14.4.2A0175.

Bulut, O. (2021). *eirm: Explanatory item response modeling for dichotomous and polytomous item responses*. https://CRAN.R-project.org/package=eirm.

Bulut, O., Görgün, G., & Yıldırım Erbaşlı, S. N. (2021). Estimating explanatory extensions of dichotomous and polytomous Rasch models: The eirm package in R. *Psych, 3*(3), 308-321. https://doi.org/10.3390/psych3030023.

Büyükkıdık, S., & Bulut, O. (2022). Analyzing the effects of test, student, and school predictors on science achievement: An explanatory IRT modelling approach. *Journal of Measurement and Evaluation in Education and Psychology, 13*(1), 40-53. https://doi.org/10.21031/epod.1013784.

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö., Karadeniz, Ş., & Demirel, F. (2014). *Bilimsel araştırma yöntemleri*. Pegem Akademi.

Castedo, A. L., Juste, M. P., & Alonso, J. D. (2015). Social competence: Evaluation of assertiveness in Spanish adolescents. *Psychological reports, 116* (1), 219-229. https://doi.org/10.2466/21.PR0.116k12w5.

Chalmers, R.P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software, 48* (6), 1-29. https://doi.org/10.18637/jss.v048.i06.

Christensen, K.B., Makransky, G., & Horton, M.C. (2017) Critical values for yen's q3: identification of local dependence in the rasch model using residual correlations. *Applied Psychological Measurement, 41* (3), 178-194. https://doi.org/10.1177/0146621616677520.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer-Verlag.

Demirkol, S., & Ayvallı Karagöz, M. (2023). PISA 2015 okuma becerisi maddelerinin güçlük indeksini etkileyen madde özelliklerinin incelenmesi. *Ana Dili Eğitimi Dergisi, 11*(3), 567-579. https://doi.org/10.16916/aded.1212049

Demirkol, S., & Kelecioğlu, H. (2022). Analyzing the interaction of item position effect and student characteristics within explanatory IRT models. *Journal of Measurement and Evaluation in Education and Psychology, 13*(4), 282-304. https://doi.org/10.21031/epod.1126368

Embretson, S. E., & Reise, S. P. (2000*). Item response theory for psychologists*. Lawrence Erlbaum Associates Inc.

Eskin, M. (2003). Self-reported assertiveness in Swedish and Turkish adolescents: A cross-cultural comparison. *Scandinavian Journal of Psychology, 44*(1), 7-12. https://doi.org/10.1111/1467-9450.t01-1-00315.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Springer Science & Business Media.

Hebbali A. (2020). *olsrr: Tools for building OLS regression models*. https://CRAN.R-project.org/package=olsrr.

Kahraman, N. (2014). An explanatory item response theory approach for a computer-based case simulation test. *Eurasian Journal of Educational Research, 14* (54), 117-134. https://doi.org/10.14689/ejer.2014.54.7.

Kankaraš, M., & Suarez-Alvarez, J. (2019). *Assessment framework of the OECD study on social and emotional skills*. OECD Publishing. https://doi.org/10.1787/5007adef-en.

Karaaslan Arkan, R. (2016). *Ergenlerin atılganlık, utangaçlık ve yalnızlık düzeyleri*. [Unpublished master thesis]. Beykent Üniversitesi.

Kaya, Z., & Karaca, R. (2018). Ergenlerin atılganlık ve sürekli kaygı düzeylerinin bazı değişkenlere göre incelenmesi. *Van Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi, 15*(1), 1490-1516. https://doi.org/10.23891/efdyyu.2018.113

Keliat, B. A., Tololiu, T. A., Daulima, N. H. C., & Erawati, E. (2015). Effectiveness assertive training of bullying prevention among adolescents in West Java Indonesia. *International Journal of Nursing, 2*(1), 128-134. https://doi.org/10.15640/ijn.v2n1a14.

Kılıç, G. (2009). *Lise öğrenimi görmekte olan ergenlerin atılganlık düzeylerinin ebeveynlerine bağlanma örüntülerine ve bazı demografik değişkenlere göre incelenmesi: Darıca ilçesi örneği*. [Unpublished master thesis]. Maltepe Üniversitesi.

Kim, J., & Wilson, M. (2020). Polytomous item explanatory item response theory models. *Educational and Psychological Measurement, 80*(4), 726-755. https://doi.org/10.1177/0013164419892667.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174. https://doi.org/10.1007/BF02296272.

Nunnally, J., & Bernstein, I. (1994), *Psychometric Theory*, McGraw-Hill.

OECD. (2015). Skills for social progress: The power of social and emotional skills. OECD Publishing. https://doi.org/10.1787/9789264226159-en.

OECD. (2021a). Beyond academic learning: First results from the survey of social and emotional skills. OECD Publishing. https://doi.org/10.1787/92a11084-en.

OECD. (2021b). OECD Survey on social and emotional skills technical report. OECD Publishing.

R Core Team (2022). *R: A language and environment for statistical computing*. https://www.R-project.org/.

Revelle, W. (2022) *Psych: procedures for personality and psychological research*. https://CRAN.R-project.org/package=psych.

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17* (5), 1-25. https://doi.org/10.18637/jss.v017.i05.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika, 34* (1), 1-97. https://doi.org/10.1007/BF03372160.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136.

Sitota, G. (2018). Assertiveness and academic achievement motivation of adolescent students in selected secondary schools of Harari peoples regional state, Ethiopia. *International Journal of Education and Literacy Studies, 6*(4), 40-46. https://doi.org/10.7575/aiac.ijels.v.6n.4p.40.

Stake, J. E., DeVille, C. J., & Pennell, C. L. (1983). The effects of assertive training on the performance self-esteem of adolescent girls. *Journal of Youth and Adolescence, 12*(5), 435-442. https://doi.org/10.1007/BF02088725.

Stanke, L., & Bulut, O. (2019). Explanatory item response models for polytomous item responses. *International Journal of Assessment Tools in Education, 6*(2), 259–278. https://doi.org/10.21449/ijate.515085.

Tuerlinckx, F., & Wang, W.C. (2004). Models for polytomous data. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 75–109). Springer-Verlag.

Vansteelandt, K. (2000). *Formal models for contextualized personality psychology* [Unpublished doctoral dissertation]. K.U.Leuven.

Wickham H. (2022). *stringr: Simple, consistent wrappers for common string operations*. https://CRAN.R-project.org/package=stringr.

Wickham H., François R., Henry L., & Müller K. (2022). *dplyr: A grammar of data manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham H., Miller E., & Smith D. (2022). *haven: Import and export 'SPSS', 'Stata' and 'SAS' files*. https://CRAN.R-project.org/package=haven.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.

Wilson, M., & De Boeck, P. (2004). Descriptive and explanatory item response models. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 43–74). Springer-Verlag