ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Research Article; 2022; 13(4); 282-304

Analyzing the Interaction of Item Position Effect and Student Characteristics within Explanatory IRT Models*

Sinem DEMİRKOL** Hülya KELECİOĞLU***

Abstract

The location of the same item in different positions among booklets leads to a biased estimation of item parameters. This undesirable effect on the probability of answering the items correctly is referred as the item position effect. The purpose of this study is to examine the items that are more sensitive to the item position effect and to investigate the student characteristics related to the item position effect. In the study, the items in the PISA 2015 reading domain are used. The study group consists of 2418 students who responded to the items in the reading domain from PISA 2015 Turkey Sample. Explanatory IRT models are used in the analysis of the research. According to the results, 42% of the items are affected by the item position. The most important characteristic related to item position is the SES level of students. In addition, male students are more affected by item position than female students.

Keywords: Item position effect, Explanatory IRT models, student characteristics reading domain

Introduction

In general, tests are tools that measure and understand the natural or learned abilities, knowledge, and characteristics of a person or community (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 2014). The general purpose of these tools is to reveal the properties of individuals to be measured through observable response behaviors to a series of items (Rose et al., 2019). Items in a booklet are always given in a context. As the test and item properties change, the context of the item in the test changes. These adjustments are possible sources of construct-irrelevant variability in test scores. Such undesirable variability in response behaviors, which cannot be explained by the ability level of students and stem from the context of the test, is expressed as context effect (Leary & Dorans, 1985; Yen, 1980). According to Brennan (1992), although there is no full definition of context effects, differences in one or more statistics should be taken as evidence for context effects.

As a type of context effects, the effect of the same item in different locations among the booklets on the probability of answering the item correctly is expressed as the item position effect (Leary & Dorans, 1985). According to Weirich et al. (2017), item position effects can be considered a special case of context effects, since the position in a test is part of the context. Therefore, although context and position effects are tried to separate from each other, it is not correct to assert the cause of the construct-irrelevant effect as the context or the position effect only (Brennan, 1992; Leary & Dorans, 1985).

The fact that the same item is in different positions among test forms affects the item parameters (Bulut, Quo & Gierl, 2017; Mollenkopf, 1950; Qian, 2014). Studies generally examined the effect of item position on item difficulty (Alexandrowicz & Matschinger, 2008; Guertin, 1954; Wise et al., 1989; Harting & Buchholz, 2012, Hahne, 2008; MacNicol, 1956). Items at the end of the test may become easier or more difficult than items at the beginning of the test. There are usually two possible

* The present study is a part of PhD Thesis conducted under the supervision of Prof. Dr. Hülya KELECİOĞLU and prepared by Sinem DEMİRKOL.

** PhD., Hacettepe University, Faculty of Education, Ankara-Türkiye, dmrklsinem@gmail.com, ORCID ID: 0000-0002-9526-6156
*** Prof. Dr., Hacettepe University, Faculty of Education, Ankara-Türkiye, hulyaebb@hacettepe.edu.tr, ORCID ID: 0000-0002-0741-9934

To cite this article:
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

Received: 5.06.2022
Accepted: 3.11.2022

explanations, depending on the direction of the effect on item difficulty. The increase in item difficulty can be interpreted as an effect of students' fatigue or low motivation to answer. On the other hand, the decrease in item difficulty can be interpreted as a practice or learning effect (Kingston & Dorans, 1984). The learning effect may be due to test takers becoming more familiar with the items during the test. (Hohensinn et al., 2011). Although both explanations seem plausible, in order to examine the position effects, item or individual properties should be included in the models (Debeer & Janssen, 2013). Investigating such variables will help to reduce the item position effects; thus, the reliability and validity of the measurement tool will improve.

The assumption that the item and person parameters are not affected by the booklet selection is violated by the item position effect (Albano, 2013). If item responses are not independent under one dimension, another dimension may cause dependency. The effect of the item position on the item parameters violates the local independence assumption of the item response theory (IRT) in particular (Hahne, 2008). This situation causes biased results in the estimation of the item parameters, and thus in the ability estimations of the individuals (Whiteley & Davis, 1976). One of the most important advantages of the models of IRT is that the item parameters are independent of the latent trait. (DeMars, 2016; Embretson & Reise, 2000). In cases where IRT assumptions are met, item parameter invariance is also ensured (Hambleton & Swaminathan, 1985). As mentioned above, item position effects violate the local independence assumption and item parameter invariance assumption, which is shown as one of the most important differences between classical test theory (CCT) and IRT. Violation of this assumption causes problems, especially in equating studies using common items (Angoff, 1971; Meyers, Miller & Way, 2009).

The item position effect differed among students, that is, the students were not exposed to the same level of item position effect (Christiansen & Janssen, 2020; Deeber & Janssen, 2013; Demirkol & Kelecioğlu, 2022). These results raised the question of what individual characteristics might be related to the item position effect, and the relationship between different student characteristics and the item position effect was investigated (Smouse & Munz, 1968; Nagy et al., 2018; Qian, 2014; Weirich et al., 2017; Wu et al., 2019). Therefore, in this study, the relationship between motivation, anxiety levels, gender, and SES, which are thought to have the most effect on students' response behaviors, and the item position effect are discussed.

When the literature is examined, it has been emphasized that the item position effect may be caused by the differences in the motivation level of the test takers (Kingston & Dorans, 1982; Albano, 2013; Debeer & Janssen, 2013). Wu et al. (2019) examined the relationship between students' motivation levels and item position effect to explain individual differences in item position effects. In the study, students' motivation levels were represented by the variables; enjoyment and interests, effort thermometer, and perseverance. According to the results of the research, in most countries, students who enjoyed reading had higher persistence levels in 2009 PISA (The Programme for International Student Assessment), but this relationship was not observed in 2006 PISA science and 2012 PISA mathematics. In addition, in some countries, it was stated that students' test-solving efforts had an effect on persistence in the 2006 and 2012 PISA. However, it was emphasized that a general motivational effect was not consistently associated with item position in all countries and PISA cycles. Weirich et al. (2017) stated that even when the initial motivation levels and variability of the motivation levels of the students are controlled, the position effects continue in an "ideal" group of highly motivated students during the test.

One of the most important variables associated with students' academic achievement is considered to be SES (Taylor, 2005). Students with higher SESs may have more chance to focus on their studies and maintain their attention (Sirin, 2005). This can make students more advantageous during the exam. Therefore, SES affects both academic skills and motivation of students (Duncan & Magnuson, 2005). Nagy et al. (2018), using PISA 2006 data, examined the relationship of the item position effect with student characteristics with structural equation models at school and student levels. In the reading field, it was found that the most important student characteristic related to item position was SES, and students

with high SES levels were less affected by the item position effect by 32%. Contrary to this study, Wu et al., (2019) stated that in most of the countries they examined, there was no significant relationship between the item position effect and SES.

Grandy (1987) stated that the gender factor is effective in students' response behaviors. When the literature is examined, female students have more effort and their attention levels are higher than male students in low-stake exams (Butler & Adams, 2007; Eklöf, 2007). It can be thought that the motivation, attention, and effort of male students' change/decrease more than female students during the test, and therefore they are exposed to more item position effect. When the relationship between item position effect and gender was examined, Qian (2014) found that male students were more affected by item position in the 2007 NAEP (National Assessment of Educational Progress) writing data. Nagy et al. (2018) stated that gender was related to the item position effect in the domains of mathematics, reading, and science, and that male students were more affected by the item position effect than female students. Again, Wu et al. (2019) stated that male students were more affected by the item position effect than female students.

The effects of item position effect and test anxiety on achievement test scores were first studied by Smouse and Munz (1968). A multiple-choice test consisting of 100 items prepared in the field of psychology was administered to 113 undergraduate students. Three different forms of the test were prepared as items from easy to difficult, difficult to easy, and at random, and students were randomly divided into two groups. The first group was given information to increase the anxiety level of the students, while the normal test atmosphere was maintained in the second group. As a result of the analysis of variance, the ordering of the items according to different difficulty levels did not have any interaction with anxiety. Later, the research was expanded and the study was repeated with 40 students with the lowest and highest anxiety levels using the Achievement Anxiety Test. According to the results of this study, a significant interaction was found between item position and anxiety. This result was interpreted as students with very low or high anxiety levels might be affected by the item position. On the other hand, Berger et al. (1969) and Towle & Merrill (1975) found that there was no significant relationship between the item position effect and anxiety.

Purpose of the Study

In many test programs, it is assumed that context or position effects have no or negligible effect on students' responses. Violation of this assumption may lead to biased estimates in item and person parameters. Rather than a general item position effect, it will be valuable for test developers and practitioners to investigate the problematic items that are significantly affected by the item position and to find solutions for the undesirable effects that occur in these items (Albano, 2013). Depending on the psychological events that may occur during the test (such as frustration, excitement, fatigue) or individual characteristics (such as gender, SED), the position of an item contributes to the probability of correct answers. This result gave rise to the question, "Which characteristics of individuals related to the item position effect?" (Nagy et al. 2018; Bulut et al., 2017) The purpose of this study is to examine items that are more sensitive to item position rather than a general item position effect and individual characteristics that are thought to be related to item position effect. For this purpose, the relationship between the motivation, anxiety, SES, and gender variables of students and the effect of item position are examined on an item basis. In this study, it is aimed to answer the following questions.

1. How is the item position interaction at item level?
2. How does the item position effect interact with the SES?
3. How does the item position effect interact with gender?
4. How does the item position effect interact with students' test anxiety?
5. How does the item position effect interact with students' achievement motivations?

Method

Working Group

5895 (50.2% female, 49.5% male) students participated in the PISA 2015 in Turkey. Since the research focused on the items in the reading domain, all students who answered the items in the reading were included in the research. Therefore, the study group of the research consists of 2418 (49.8% female, 50.2% male) students who were drawn from the PISA 2015 Turkey sample and answered the items in the reading domain.

Data Collection Methods

PISA, financed by the OECD (Organization for Economic Cooperation and Development), is an international research to measure and evaluate the purposes of education (OECD, 2017). PISA mainly assesses students' literacy in science, math, and reading fields. Knowledge about students' motivations, their opinions about themselves, their psychological characteristics about learning processes, and school environments are collected via student, teacher, and school questionnaires.

PISA 2015 was carried out on computer based in Turkey. For computer-based assessment, 66 different main booklets were used. There were 12 different clusters in science, 6 in mathematics, 6 in reading, and 4 in collaborative problem-solving. The booklets used in PISA 2015 were created to include four of these clusters prepared in the fields of science, mathematics, reading, and collaborative problem-solving. The positions of the clusters in the booklets differ, but the positions of the items in the clusters are fixed. Therefore, the item position effect was examined on cluster-based in this study. Since the test forms consisted of four different clusters, the first cluster was coded as 0, the second cluster 1, the third cluster 2, and the fourth cluster 3. In this way, the position variable is a variable that takes a value between 0 and 3.

In this study, the effect of item position on reading items was investigated. Thirty-six booklets containing reading items were used. There were 88 items in the reading area, and all items were included in the analysis. Since the analysis model used was in accordance with the dichotomous scored item format, the partially scored (7 items) items were converted into dichotomous scoring with 0 for "incorrect" and "partially correct" items and 1 for "fully correct" items. In addition, omitted items were considered incorrect and inaccessible items or missing items due to other reasons were considered missing.

Variables Used in the Research

Gender. The gender variable is a two-category variable, coded as 0 for females and 1 for males.

Socio-economic status (SES). SES index is built by the PISA study team via principal components analysis using parent education (PARED), highest parent occupation (HISEI), and home possessions (HOMEPOS). This variable is standardized to have a mean of 0 and a standard deviation of 1. In the sample of Turkey, the minimum value of the SES variable is -5.131, and the maximum value is 3.123. In the PISA technical report, factor loadings of the variables used in the SES are given for each country. For the Turkish sample, the factor load of HISEI is calculated as 0.82, the factor load of PARED as 0.79, and the factor load of HOMEPOS as 0.77. In addition, the scale reliability (Cronbach's alpha) of the SES variable for the Turkish sample is estimated as 0.68.

Test anxiety. In order to investigate the test anxiety of test takers, there are five items in the PISA 2015 student questionnaire. Students answer these items in the categories of "strongly agree", "agree", "disagree", and "strongly disagree". The scores obtained from these categories are included in the

student questionnaire with the ANXTEST code. This variable is standardized to have a mean of 0 and a standard deviation of 1. In the sample of Turkey, the minimum value of the anxiety variable is -2.505, and the maximum value is 2.549. For the Turkish sample, the reliability coefficient (Cronbach Alpha) of the "test anxiety" scale is estimated as 0.825.

Achievement motivation. There are five items in the PISA 2015 student questionnaire in order to investigate the achievement motivation of test takers. Students answered these items in the categories of "strongly agree", "agree", "disagree", and "strongly disagree". The scores obtained from these categories are included in the student questionnaire with the MOTIVAT code. This variable is standardized to have a mean of 0 and a standard deviation of 1. The minimum value of this variable is -3.087, and the maximum value is 1.854. For the Turkish sample, the reliability coefficient (Cronbach Alpha) of the scale is estimated as 0.825. For detailed information, the PISA 2015 technical report can be viewed (OECD, 2017).

Analysis of Data

A two-stage procedure can be used when the purpose of research is to explain the differences in person and item parameters with traditional IRT models. In the first stage, the abilities and item parameters are estimated, and in the second stage, the variables that are thought to cause differences in ability and item parameters can be modeled with various methods (De Boeck & Wilson, 2004; Atar & Çobanoğlu Aktan, 2013).

When IRT models are considered within the framework of generalized linear or non-linear mixed models, they can simultaneously estimate the variation among individual and item parameters by including the predictors of the individual, item, or interaction of both in the model. These models are called Explanatory IRT models (De Boeck & Wilson, 2004). The main advantage of explanatory IRT models is that they provide flexibility to analyse the covariance between these parameters simultaneously while estimating item and individual parameters (Briggs, 2008).

In explanatory IRT models, the responses of items are considered repeated measurements. So item answers are embedded within students. Examining the responses to the items in a multi-level framework also allows to consider the effect of the explanatory variables as a fixed or random effect across the levels (De Boeck & Wilson, 2004). Within the framework of Explanatory IRT models, the item position variable can be included in the model as a predictor variable to explain the difference between the difficulty of the items (Atar, 2011; De Boeck & Wilson, 2004; Debeer & Janssen, 2013).

There are four explanatory IRT models that are widely used. These are the Rasch model, the latent regression Rasch model (LRRM), the linear logistic test model (LLTM), and the latent regression LLTM. In the Rasch Model, while estimating the difficulties of the items and the abilities of the persons, there are no item or person properties to explain the differences in these parameters. Hence this model is referred to as "doubly descriptive model". The Latent Regression Rasch Model (LRRM) is a model in which person properties are included as explanatory variables in order to explain the differences among persons' ability levels, but no explanatory properties are included at the item level. Since this model only includes explanatory variables at the person level, it is also known as "person explanatory model". The Linear Logistic Test Model (LLTM), unlike the LRRM, includes item-level explanatory variables to explain the differences in the difficulty of items, but no explanatory properties are included at the person level. It is also known as "item explanatory model". Finally, the Latent Regression LLTM includes item and person properties simultaneously to explain the differences in item and person parameters. In addition, in this model, interactions of item and person properties can be added. This model is also called the "double explanatory model" as it includes explanatory variables at both the item and person level (De Boeck & Wilson, 2004).

The purpose of this study is to investigate items that are more sensitive to the item position effect rather than an overall effect. Therefore, the study started by examining the interaction between the item and item position on an item basis. Then, the relationship between these interactions (item-position) and

individual characteristics was examined. The interaction model between the item and the position is given in equation 1.

$$logit\left[P\left(Y_{pik}=1\right)\right]=\theta_p+\beta_i+\delta_p\left(k_{pi}-1\right)+\gamma_{1i}\left(k_{pi}-1\right) \quad (1)$$

$P\left(Y_{pik}=1\right)$, is the probability of the person p giving the correct answer to the item i in the k position, β_i is the ease of item i in the reference position, k_{pi} is the position in which p person takes item i, γ_{1i} is the item-position interaction effect of item i. In the models, the random effect of θ_p is person ability and it is assumed to have a normal distribution ($\theta_p \sim N(0,\sigma_\theta^2)$). β_i is fixed effect of items, δ_p is the random effect of item position among persons. In other words, it is the deviation of the person p from the general position effect and it is assumed to have a normal distribution (($\delta_p \sim N(0,\sigma_\delta^2)$).

When individual characteristics are added to the item-position interaction, equation 1 is extended as equation 2.

$$logit\left[P\left(Y_{pik}=1\right)\right]=\theta_p+\beta_i+\delta_p\left(k_{pi}-1\right)+\gamma_{1ip}\,Z_p\left(k_{pi}-1\right) \quad (2)$$

In Equation 2, Z_p is the value of the Z property of person p (person level covariate). γ_{1ip} is the interaction of item, position, and person properties of item i.

The item position interactions were examined on an item basis with the model (Model 1) given in equation 1. Then, with equation 2, the interaction of the item position and the variables of gender (Model 2), SES (Model 3), anxiety (Model 4), and motivation (Model 5) were examined. The analysis of the study was carried out in the R program, within the framework of GDKMs, with the glmer function of the lme4 (Bates et al., 2014) package and the eirm (Bulut, 2021) package suitable for the analysis of Explanatory IRT models. The R codes used in the models are given in Table 1. Maximum Likelihood (Laplace Approximation) method was used in the estimations of the models.

Table 1

R Codes used in Models

Models	R codes used in lme4
Model 1	responses ~ -1 + items + position:items + (1 + position | id)
Model 2	responses ~ -1 + items + position:items:gender + (1 + position | id)
Model 3	responses ~ -1 + items + position:items:SES + (1 + position | id)
Model 4	responses ~ -1 + items + position:items:anxiety + (1 + position | id)
Model 5	responses ~ -1 + items + position:items:motivation + (1 + position | id)

Results

Model fit indices are given in Table 2. When AIC, BIC, and log-likelihood indexes were examined, it was seen that M3 was the best model with model data fit. This result can be interpreted as that the interaction between item position and SES is more than other variables.

Table 2

Model Fit Indices

Models	AIC	BIC	logLik	deviance
M1	64540	66145	-32091	64182
M2	64546	66151	-32094	64188
M3	64535	66140	-32088	64177
M4	64664	66269	-32153	64304
M5	64653	66258	-32147	64295

First, the interaction of each item with the position was examined. According to the results, item-position interactions were significant in 37 of 88 items and these effects were in the range of -0.549 to -0.162 logit. When the direction of these effects was examined, the significant effect in all items was negative. Locating these items one cluster later reduces probability of answering the item correctly. Table 3 shows statistically significant item-position interactions. Item-position interaction for all items are given in Appendix A.

Table 3

Item-Position Interactions

Item No	Estimate	Item No	Estimate	Item No	Estimate
3	-0.307 (0.081)***	27	-0.228 (0.080)**	55	-0.225 (0.087)***
4	-0.162 (0.079)*	29	-0.242 (0.117)*	57	-0.264 (0.081)***
5	-0.254 (0.083)**	30	-0.404 (0.078)***	64	-0.176 (0.086)*
6	-0.184 (0.077)*	31	-0.368 (0.157)*	69	-0.234 (0.095)*
7	-0.310 (0.083)***	35	-0.177 (0.083)*	72	-0.271 (0.082)**
9	-0.298 (0.079)***	37	-0.276 (0.077)***	73	-0.211 (0.078)**
10	-0.196 (0.080)*	39	-0.229 (0.086)**	74	-0.274 (0.088)**
12	-0.336 (0.082)***	40	-0.240 (0.077)**	76	-0.169 (0.081)**
13	-0.444 (0.101)***	41	-0.250 (0.080)**	79	-0.181 (0.079)*
18	-0.549 (0.192)**	51	-0.189 (0.095)*	82	-0.224 (0.090)*
20	-0.238 (0.087)**	52	-0.172 (0.085)*	83	-0.212 (0.094)*
22	-0.249 (0.079)**	54	-0.312 (0.103)**	85	-0.224 (0.088)*
23	-0.172 (0.077)*				

* p< .05; ** p< .01; *** p<.001. Standard errors of estimates are shown in parentheses.

For example, responding to the 18th item one cluster later reduces the probability of answering the item correctly by 0.549 logits. When the exponential of 0.549 on the logit scale is taken (exp(0.549)), the obtained value of 1.731 gives the odd ratio. Answering item 18 after one cluster will reduce the odds ratio to approximately 1.731. If the probability of answering this item correctly is 0.50 at the reference position, answering it one cluster later reduces the probability of answering correctly to approximately 0.36.

Figure 1 shows the relationship between the probability of answering the items correctly and the item position. Each quadruple dot (Red-Blue-Green-Purple) in the graphs shows the variation in the probability of responding to items correctly according to the item position.

Figure 1

Item-position interactions

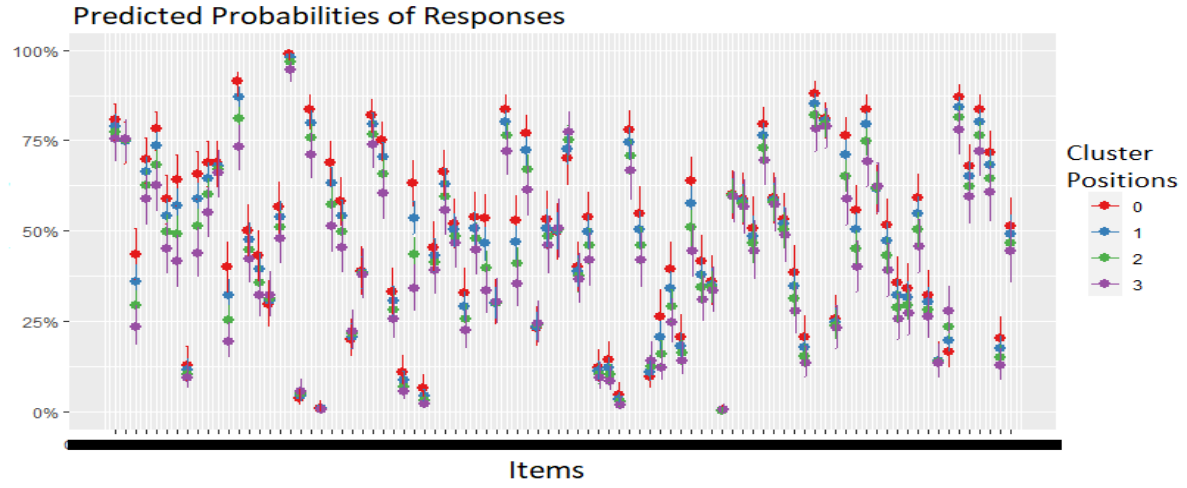


Figure 1 shows that, generally, the red dots (reference position) are at the top and the purple dots are at the bottom. In other words, the items in the reference position are more likely to be answered correctly, while the items in the last positions are less likely to be answered correctly. The points that overlap or are very close to each other can be interpreted as the probability of answering the item correctly is not affected by its position in the test. In addition, in appendix B, the graphs of the items with significant item position interactions are shown on an item basis.

The relationship between item-position interaction and gender was examined. The purpose here is to examine how the probability of answering the item changes according to the item position for different gender groups. Since females are coded as 0 and males as 1, the reference group is females. Table 4 shows the estimates for items that have significant interactions. Item-position-gender interactions for all items are given in Appendix A.

Table 4

Item-Position-Gender Interaction

ItemNo	Estimate	ItemNo	Estimate	Item No	Estimate
4	-0.167 (0.078)*	32	-0.174 (0.078)*	67	-0.386 (0.097)***
7	-0.225 (0.081)**	37	-0.198 (0.077)*	74	-0.209 (0.083)*
9	-0.153 (0.079)*	44	-0.203 (0.088)*	79	-0.253 (0.082)**
12	-0.182 (0.083)*	47	-0.210 (0.083)*	80	-0.188 (0.090)*
14	-0.155 (0.076)*	54	-0.428 (0.128)***	85	-0.199 (0.083)*
18	-0.352 (0.154)*	55	-0.183 (0.092)*	86	-0.197 (0.085)*
30	-0.169 (0.080)*	57	-0.173 (0.084)*		

* p< .05; ** p< .01; *** p<.001. Standard errors of estimates are shown in parentheses.

In 20 of the 88 items, the interaction of item, position, and gender is significant. The interaction values range from -0.428 to -0.153 logit. All significant interaction values are negative. The items that have statistically significant interaction values are answered one cluster later by male students reduces the probability of answering the items correctly. Figure 2 shows the effect of students' gender and item-position interaction on the probability of answering the item correctly.

Figure 2

Item-position-gender interaction

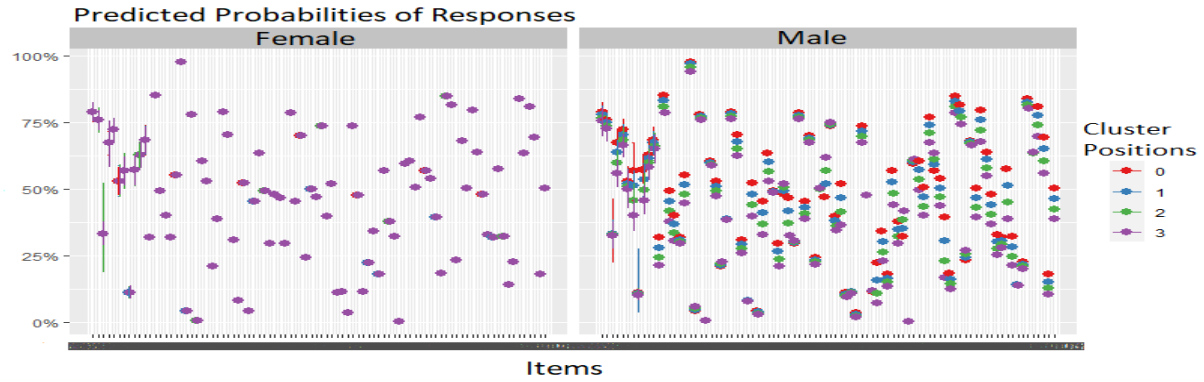


Figure 2 shows that if an item is located in the last position (purple dot), the probability of being answered correctly decreases compared to the first position (red dot). On the other hand, in items where the colors overlap, it can be interpreted that the item position effect does not differ according to gender. When female students are taken as the reference group, male students' answering the items in later positions decreases the probability of answering the items correctly. In other words, male students are more affected by item-position interaction than female students. In addition, in appendix C, the graphs of the items with significant item position and gender interactions are shown on an item basis.

The relationship between the item-position interaction and the SES levels of the students was examined. The purpose here is to examine how the probability of correct answers to the items varies according to the item position for students from different SES levels. Table 3 shows the estimates of the items that have significant interactions. Item-position-SES interactions for all items are given in Appendix A.

Table 5

Item-Position-SES Interactions

Item No	Estimate	Item No	Estimate	Item No	Estimate
3	0.144 (0.034)***	43	0.076 (0.038)*	67	0.117 (0.040)**
5	0.099 (0.031)**	47	0.087 (0.034)*	70	0.083 (0.036)*
10	0.089 (0.032)**	49	0.125 (0.052)*	72	0.092 (0.033)**
12	0.111 (0.035)***	52	0.120 (0.036)***	73	0.086 (0.033)**
13	0.092 (0.039)*	54	0.113 (0.048)*	74	0.099 (0.035)**
16	0.065 (0.033)*	55	0.084 (0.038)*	76	0.092 (0.035)**
29	0.111 (0.054)*	56	0.092 (0.045)*	77	0.095 (0.037)*
30	0.095 (0.032)**	57	0.083 (0.033)*	78	0.103 (0.038)**
36	0.083 (0.031)**	62	0.126 (0.036)***	84	0.063 (0.032)*
41	0.082 (0.032)**	66	0.087 (0.036)*	87	0.146 (0.046)**

* p< .05; ** p< .01; *** p<.001. Standard errors of estimates are shown in parentheses.

Table 5 shows that in 30 of the 88 items, the interaction of item, position, and SES is significant and the interaction estimates are in the range from 0.063 to 0.146 logit. All significant interactions are positive. Students with higher SES will be more likely to answer the item correctly as the item position increases (when the item is located in the later parts of the test).

Figure 3

Item-position-SES interaction at different SES levels

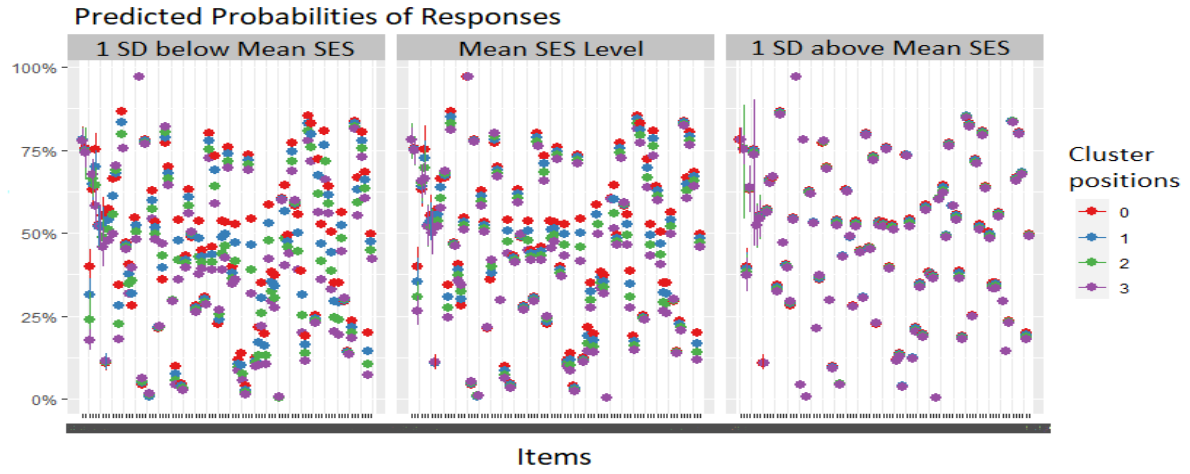


Figure 3 shows that in groups whose SES level is one standard deviation below the mean SES level, answering the item one cluster later reduces the probability of a correct answer. As the SES level of the individuals increases, the probability of answering the item correctly in the subsequent positions increases and there is almost no item position effect above one standard deviation of the mean SES level. In addition, in appendix D, the graphs of the items with significant item position and SES interactions are shown on an item basis.

How the probability of correct answers to the items changes according to the item position was examined for individuals who have different anxiety levels. Table 6 shows the estimates of the items with significant interactions. Item-position-anxiety interactions for all items are given in Appendix A.

Table 6

Item-Position-Anxiety Interaction

Item No	Estimate	Item No	Estimate	Item No	Estimate
3	-0.086 (0.043)*	13	-0.108 (0.050)*	41	-0.105 (0.043)*
5	-0.098 (0.044)*	18	-0.173 (0.086)*	53	0.141 (0.058)*
9	-0.088 (0.042)*	30	-0.103 (0.044)*	65	0.124 (0.049)*
12	-0.105 (0.049)*				

* p< .05; ** p< .01; *** p<.001. Standard errors of estimates are shown in parentheses.

Table 6 shows that in 10 of the 88 items, the interaction of item, position, and anxiety is significant and the interaction estimates ranged from -0.173 to 0.144. 8 of the 10 items that have significant interaction are negative, and 2 of them are positive. Student with higher anxiety level responding to the item one set later decreases the probability of answering correctly in 8 out of 10 items, and increases it in 2 of 10 items (items 53 and 65).

Figure 4

Item-position-anxiety interaction at different anxiety levels

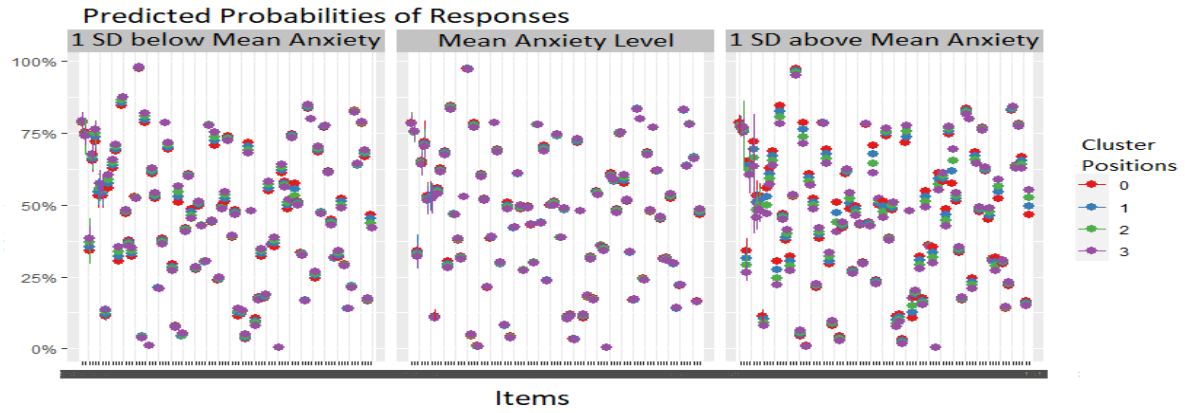


Figure 4 shows that the item position effect is less when the student's anxiety level is average, and the item position effect increases when the student's anxiety level is high. This result can be interpreted as that when the anxiety level is average, the student maintains their attention more during the test and is less affected by the effect of fatigue. In addition, in appendix E, the graphs of the items with significant item position and anxiety interactions are shown on an item basis.

How the probability of correct answers to the items changes according to the item position was examined for individuals with different motivation levels. Table 7 shows the estimates of the items that have significant interactions. Item-position-motivation interactions for all items are given in Appendix A.

Table 7

Item-Position-Motivation Interaction

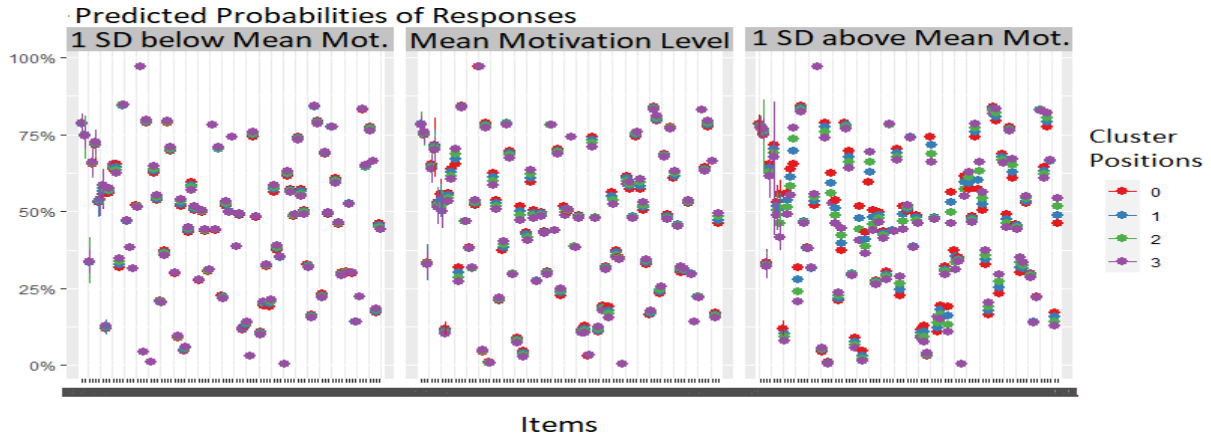
Item no	Estimate	Item no	Estimate	Item no	Estimate
7	-0.113 (0.043)**	30	-0.091 (0.041)*	33	0.089 (0.043)*
11	0.188 (0.046)*	31	-0.266 (0.078)***	56	-0.133 (0.059)*
12	-0.119 (0.046)**				

* p< .05; ** p< .01; *** p<.001, Standard errors of estimates are shown in parentheses.

In 7 of the 88 items, the interaction of item, position, and motivation was significant, and the interaction values ranged from -0.266 to 0.188. In 5 of the 7 items (items 7, 12, 30, 31, and 56), the direction of the interaction is negative. Answering these items in later positions reduces the probability of answering the item correctly. In 2 of the 7 items (items 11 and 33), the direction of the interaction is positive. Answering these items in later positions increases the probability of answering the items correctly.

Figure 5

Item-position-motivation interaction at different motivation levels



While there were 37 items with statistically significant item-position interactions, the number of items with significant interactions decreased to 7 when the motivation levels of individuals were included in this interaction. In addition, when the direction of these interactions was examined, it was seen that consistent results were not obtained. This result suggests that, among other variables used in the study, the motivation levels of individuals are less related to the item-position interaction. In addition, in appendix F, the graphs of the items with significant item position and motivation interactions are shown on an item basis.

Discussion and Conclusion

A general item position effect has been investigated in many studies in the literature (Christiansen & Janssen, 2020; Debeer & Janssen, 2013; Hahne, 2008; Meyers et al.,2009; Nagy et al., 2018; Weirich et al., 2017). Demirkol & Kelecioğlu (2022) found that there is a general item position effect in the PISA 2015 reading and mathematics data, and the probability of correct answers decreases when the items are located in later positions. But it is precious, especially for test developers and practitioners, to examine items that are more sensitive to item position effect (Albano, 2013, Bulut et al., 2017). Therefore, the purpose of this study is to examine the items that are more sensitive to the item position rather than a general item position and to investigate the relationship of item position effect with student characteristics. For this purpose, the interaction between individual characteristics and the effect of item position is examined on the basis of items, and a more detailed picture is tried to be provided. According to the results, when the item-position interaction is examined at the item level, the change in the positions of approximately 42% of the items in the test significantly affects the probability of answering the item correctly. Answering these items one cluster later reduces the probability of correct answers.

SES can be built using different variables. In the study, it was built by the variables of parental education level, home possessions, and highest parental occupation via principal component analysis. According to the results, the most important variable related to item position among the variables discussed in this study is the SES level of students. In approximately 34% of the items in the study, the item-position interaction is related to the SES level of the students. The relationship which occurs in these items increases the probability of correct answers. That is, there is a learning effect in students with high SES. In the graphs shows that the item position effect has less effect on students whose SES level is 1 standard deviation higher than the mean SES level. Given that this variable is continuous, it can be said that learning effects occur as the SES level of individuals increases; students become more familiar with the items or increase their attention levels during the test. While there are studies in the literature that support

a relationship between the item position effect and SES (Nagy et al., 2018), there are also studies in which there is no stable relationship between the item position effect and SES (Wu et al., 2019).

When examining how the probability of answering the item correctly changes according to the item position for different gender groups, item-position interactions are associated with gender in 23% of the items. In the study, it is found that male students are more affected by item position than female students. This result is especially important for education politicians. The reasons for this difference between male and female students should be investigated, and necessary studies should be focused on. In studies examining the relationship between the item position effect and gender, it is found that male students are more affected by the item position effect than female students (Qian, 2014; Nagy et al., 2018; Wu et al., 2019).

When the interaction of the anxiety levels of the students with the item position is examined, it is seen that this interaction is significant in 11% of the items, and students with average anxiety levels are less affected by the item position. Increasing the level of anxiety reduces the probability of correctly answering the items in the later parts of the test. Many students are a bit anxious because of their high motivation to get a "good" score. During the test, inevitably, students' anxiety levels may increase as items become too difficult or unclear for them to answer. Therefore, when moving on to the next items, the anxiety and disappointment caused by the unsolved item may decrease the student's performance and the probability of correctly answering the items (McKeachie et al., 1955; Stanley, 1961; Cronbach, 1984). Smouse and Munz (1968) stated that students with very low or high anxiety levels are affected by item position. On the other hand, some studies found that there are no significant interactions between anxiety and item position (Berger et al., 1969; Towle & Merrill, 1975).

When the relationship between the item position effect and the motivation level of the students is examined, in approximately 8% of the items, the item-position interaction is associated with the motivation level of the students, and this relationship generally increases the probability of the correct answer to the items. However, when compared with other individual characteristics investigated in this study, the least interaction with the item position effect was seen in the motivation level of the student. The motivation variable used in this study is based on the information given by students about their own motivation levels. The scores obtained from such scales have limitations that may arise from the fact that students have deviated from the real situation (Finn, 2015; Wise & Kong, 2014). In addition, although the motivation levels of the students when they start the test are important, the item position effect may be more related to the motivation levels of the student during the test (Weirich et al., 2017). In some studies, it has been found that the most important variable related to the item position effect is motivation (Qian, 2014), while in some studies, the relationship between the item position effect and the motivation levels of students is not very clear (Wu et al., 2019).

Context and position effects are possible sources of scores independent of the test structure (Brennan, 1992). Therefore, the position of an item should be included in the measurement model as a predictor in order to examine whether the item's probability of answering the item correctly depends on the item position (Leary & Dorans, 1985; Pomplun & Ritchie, 2004). Brennan (1992) stated that “models that include the probability of the existence of context effects should be developed”. Kingston and Dorans (1984) suggested that “more general models with item position parameters should be developed”. Davey and Lee (2011) stated that “a possible direction for future analysis is to use some IRT models that can incorporate item position as a predictor”. PISA 2015 reading data was used in this study. Compared to low-stakes exams such as PISA and TIMSS, students' motivation level is higher in high-stakes exams where important decisions are made for the future of the student (Wise & DeMars, 2005). For this reason, the effect of item position on high-stakes exams can be investigated in future studies.

PISA is an exam assessing 15-year-old students. In future studies, it can be examined whether the developmental characteristics of students are effective in the item position effect. For example, whether the student's developmental characteristics (child-early-adult) are related to the item position effect can be examined with a longitudinal study. Debeer and Janssen (2013) found that linearly modeling item position effects have better model-data fit than non-linear models. Therefore, in this study, the item position effect was modeled linearly. However, an item at the beginning of the test and an item at the end of the test may not be affected by the item position effect of the same level. That is, the item position

effect may increase or decrease during the test. In future studies, the item position effect can be modeled more flexibly by non-linear models.

This study has some limitations. Partially scored items in the analyses were converted to dichotomous scoring. The analysis method used in the study is suitable for dichotomously scored items. When the studies on the item position effect were examined, it was seen that the partially scored items were converted to dichotomous scoring, and the analyses were carried out in this way (Debeer & Janssen, 2013; Hartig & Buchholz, 2012; Wu et al., 2019). Therefore, converting the partially scored items to dichotomous scoring is a limitation of the study. It was observed that 42% of the items in this study were significantly affected by the item position effect. This may be due to the analysis method and scoring procedures used. In future studies, results obtained using other analysis methods and scoring procedures can be compared. In addition, item difficulties and item discriminations are estimated in PISA. However, in this study, only the effect of item position on item difficulty was investigated. In future studies, the effect of the item position on item difficulty and the effect on item discrimination can be examined.

Declarations

Author Contribution: Sinem Demirkol-Conceptualization, methodology, analysis, writing & editing, visualization. Hülya Kelecioğlu-Conceptualization, methodology, writing & editing, supervision.

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

Ethical Approval: Secondary data were used in this study. Therefore, ethical approval is not required.

References

Albano, A. D. (2013). Multilevel modeling of item position effects. *Journal of Educational Measurement, 50*(4), 408–426. https://doi.org/10.1111/jedm.12026

Alexandrowicz, R., & Matschinger, H. (2008). Estimating item location effects by means of a generalized logistic regression model. *Psychology Science Quarterly, 50*(1), 64-74. https://www.psychologie-aktuell.com/fileadmin/download/PschologyScience/1-2008/06_Alexandrowicz.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME]. (2014). *Standards for educational and psychological testing*. American Psychological Association.

Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). American Council on Education.

Atar, B. (2011). An application of descriptive and explanatory item response models to TIMSS 2007 Turkey mathematics data. *Education and Science, 36*(159), 256-259. http://egitimvebilim.ted.org.tr/index.php/EB/article/view/811

Atar, B., & Cobanoglu Aktan, D. (2013). Person explanatory item response theory analysis: Latent regression two parameter logistic model. *Education and Science, 38*(168), 59–68. http://egitimvebilim.ted.org.tr/index.php/EB/article/view/942

Bates, D., Maechler, M., Bokler, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berger, V. F., Munz, D. C., Smouse, A. D., & Angelino, H. (1969). The effects of item difficulty sequencing and anxiety reaction type on aptitude test performance. *Journal of Psychology, 71*(2), 253–258. https://doi.org/10.1080/00223980.1969.10543091

Brennan, R. L. (1992). The context of context effects. *Applied Measurement in Education, 5*(3), 225–264. https://doi.org/10.1207/s15324818ame0503_4

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, O. (2021). *eirm: Explanatory item response modeling for dichotomous and polytomous item responses, R package* (version 0.3.0) [Computer software]. http://CRAN.R-project.org/package=eirm

Bulut, O., Quo, O., & Gierl, M. (2017). A structural equation modeling approach for examining position effects in large-scale assessments. *Large Scale in Assessments in Education, 5*(8), 1-20. https://doi.org/10.1186/s40536-017-0042-x

Butler, J., & Adams, R. J. (2007). The impact of differential investment of student effort on the outcomes of international studies. *Journal of Applied Measurement, 8*(3), 279–304. https://pubmed.ncbi.nlm.nih.gov/17804895/

Christiansen, A., & Janssen, R. (2020). Item position effects in listening but not in reading in the European Survey of Language Competences. *Educational Assessment, Evaluation and Accountability, 33*(3), 49–69. https://doi.org/10.1007/s11092-020-09335-7

Cronbach, L. J. (1984). *Essentials of psychological testing*. Harper and Row.

Davey, T., & Lee, Y. H. (2011). *Potential impact of context effects on the scoring and equating of the multistage GRE revised general test* (RR-11-26). ETS.

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Statistics for Social Science and Public Policy. Springer.

Debeer, D., & Janssen, R. (2013). Modeling item-position effects within an IRT framework. *Journal of Educational Measurement, 50*(2), 164-185. https://doi.org/10.1111/jedm.12009

DeMars, C. (2016). *Madde tepki kuramı* (Çev. Ed. H. Kelecioğlu). Nobel.

Demirkol, S., & Kelecioğlu, H. (2022). Investigating the effect of item position on person and item parameters: PISA 2015 Turkey sample. *Journal of Measurement and Evaluation in Education and Psychology, 13*(1), 69-85. https://doi.org/10.21031/epod.958576

Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children, 15*(1), 35-54. https://doi.org/10.1353/foc.2005.0004

Eklöf, H. (2007). Test-taking motivation and mathematics performance in TIMSS 2003. *International Journal of Testing, 7*(3), 311–326. https://doi.org/10.1080/15305050701438074

Embretson, S. E., & Reise, S. P. (2000). *Item response theory* (1st ed.). Psychology Press. https://doi.org/10.4324/9781410605269

Finn, B. (2015). *Measuring motivation in low-stakes assessments* (RR-15–19). Educational Testing Service. https://doi.org/10.1002/ets2.12067

Grandy, J. (1987). *Characteristics of examinees who leave questions unanswered on the GRE general test under rights-only scoring* (RR-87-38). Educational Testing Service. https://doi.org/10.1002/j.2330-8516.1987.tb00242.x

Guertin, W. H. (1954). The effect of instructions and item order on the arithmetic subtest of the Wechsler-Bellevue. *Journal of Genetic Psychology, 85*(1), 79–83. https://doi.org/10.1080/00221325.1954.10532863

Hahne, J. (2008). Analyzing position effects within reasoning items using the LLTM for structurally incomplete data. *Psychology Science Quarterly, 50*(3), 379-390. http://www.fyhe.com.au/past_papers/2006/Papers/Taylor.pdf

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Theory and applications*. Kluwer-Nijhoff.

Hartig, J., & Buchholz, J. (2012). A multilevel item response model for item position effects and individual persistence. *Psychological Test and Assessment Modeling, 54*(4), 418-431. https://www.proquest.com/openview/075b5103d499407933e3c62cca521618/1?pq-origsite=gscholar&cbl=43472

Hohensinn, C., Kubinger, K., Reif, M., Schleicher, E., & Khorramdel, L. (2011). Analysing item position effects due to test booklet design within large-scale assessment. *Educational Research and Evaluation, 17*(6), 497-509. https://doi.org/10.1080/13803611.2011.632668

Kingston, N. M., & Dorans, N. J. (1982). The effect of the position of an item within a test on item responding behavior: An analysis based on item response theory (RR-79-12bP). Educational Testing Service. https://doi.org/10.1002/j.2333-8504.1982.tb01308.x

Kingston, N. M., & Dorans, N. J. (1984). Item location effects and their implications for IRT equating and adaptive testing. *Applied Psychological Measurement, 8*(2), 147–154. https://doi.org/10.1177/014662168400800202

Leary, L. F., & Dorans, N. J. (1985). Implications for altering the context in which test items appear: A historical perspective on an immediate concern. *Review of Educational Research, 55*(3), 387–413. https://www.jstor.org/stable/1170392

MacNicol, K. (1956). *Effects of varying order of item difficulty in an unspeeded verbal test*. Unpublished manuscript, Educational Testing Service.

Meyers, J. L., Miller, G. E., & Way, W. D. (2009). Item position and item difficulty change in an IRT-based common item equating design. *Applied Measurement in Education, 22*(1), 38-60. https://doi.org/10.1080/08957340802558342

McKeachie, W. J., Pollie, D., & Speisman, J. (1955). Relieving anxiety in classroom examinations. *Journal of Abnormal and Social Psychology, 50*(1), 93-98. https://doi.org/10.1037/h0046560

Mollenkopf, W. G. (1950). An experimental study of the effects on item-analysis data of changing item placement and test time limit. *Psychometrika, 15*(3), 291–315. https://doi.org/10.1007/BF02289044

Nagy, G., Nagengast, B., Frey, A., Becker, M., & Rose, N. (2018). A multilevel study of position effects in PISA achievement tests: Student- and school-level predictors in the German tracked school system. *Assessment in Education: Principles, Policy & Practice, 26*(4), 422–443. https://doi.org/10.1080/0969594X.2018.1449100

Organisation for Economic Co-operation and Development [OECD]. (2017). *PISA 2015 technical report.* OECD. https://www.oecd.org/pisa/data/2015-technical-report/

Pomplun, M., & Ritchie, T. (2004). An investigation of context effects for item randomization within testlets. *Journal of Educational Computing Research, 30*(3), 243–254. https://doi.org/10.2190/Y4FU-45V7-74UN-HW4T

Qian, J. (2014). An investigation of position effects in large-scale writing assessments. *Applied Psychological Measurement, 38*(7), 518–534. https://doi.org/10.1177/0146621614534312

Rose, N., Nagy, G., Nagengast, B., Frey, A., & Becker, M. (2019). Modeling multiple item context effects with generalized linear mixed models. *Frontiers in Psychology, 10*(248), 1-13. https://doi.org/10.3389/fpsyg.2019.00248

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Smouse, A. D., & Munz, D. C. (1968). The effects of anxiety and item difficulty sequence on achievement testing scores. *Journal of Psychology, 68*(2), 181-184. https://doi.org/10.1080/00223980.1968.10543421

Stanley, J. C. (1961). Studying status vs. manipulating variables. In R. O. Collier & S. M. Elam (Eds.), *Research design and analysis*. Phi Delta Kappan.

Taylor, J. A. (2005). Poverty and student achievement. *Multicultural Education, 12*(4), 53-55. https://www.proquest.com/openview/3c9d66c77504e21370c0b718cf66e27f/1?pq-origsite=gscholar&cbl=33246

Towle, N. J., & Merrill, P. F. (1975). Effects of anxiety type and item-difficulty sequencing on mathematics test performance. *Journal of Educational Measurement, 12*(4), 241–249. https://www.jstor.org/stable/1434151

Weirich, S., Hecht, M., Penk, C., Roppelt, A., & Böhme, K. (2017). Item position effects are moderated by changes in test-taking effort. *Applied Psychological Measurement, 41*(2), 115–129. https://doi.org/10.1177/0146621616676791

Whitely, E., & Dawis, R. (1976). The influence of test context on item difficulty. *Educational and Psychological Measurement, 36*(2), 329–337. https://doi.org/10.1177/001316447603600211

Wise, L. L., Chia, W. J., & Park, R. (1989, March 27-31). *Item position effects for test of word knowledge and arithmetic reasoning*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Wise, S. L., & DeMars, C. E. (2005). Examinee motivation in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*(1), 1–17. https://doi.org/10.1207/s15326977ea1001_1

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*(2), 163-183. https://doi.org/10.1207/s15324818ame1802_2

Wu, Q., Debeer, D. Buchholz, J., Hartig, J., & Janssen, R. (2019). Predictors of individual performance changes related to item positions in PISA assessments. *Large Scale Assessment in Education, 7*(5), 1-20. https://doi.org/10.1186/s40536-019-0073-6

Yen, W. M. (1980). The extent, causes and importance of context effects on item parameters for two latent trait models. *Journal of Educational Measurement, 17*(4), 297–311. http://www.jstor.org/stable/1434871

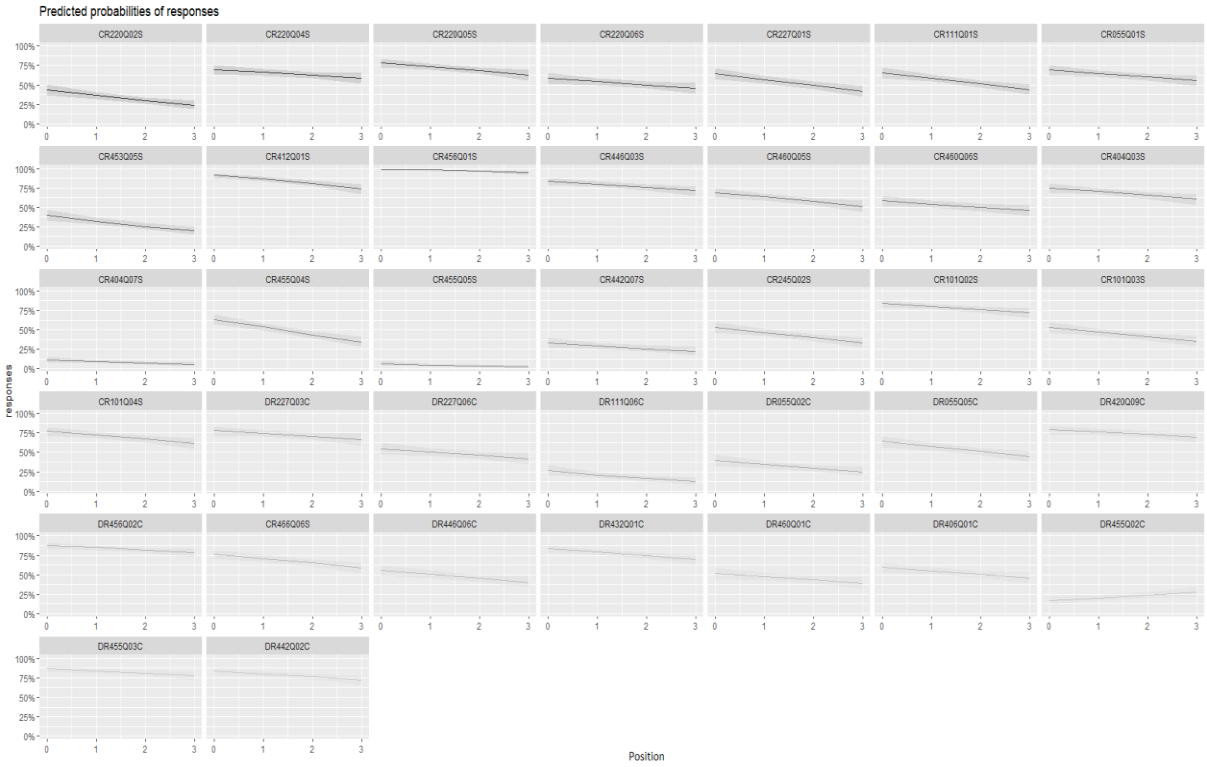
Appendix A

Item No	Item Code	Item-Position Interactions	Item-Position Gender Interactions	Item-Position SES Interactions	Item-Position Anxiety Interactions	Item-Position Motivation Interactions
1	CR067Q01S	-0.101 (0.087)	-0.068 (0.085)	0.003 (0.034)	-0.022 (0.047)	-0.015 (0.044)
2	CR102Q07S	0.014 (0.084)	-0.061 (0.083)	0.009 (0.033)	0.024 (0.046)	0.020 (0.042)
3	CR220Q02S	-0.307 (0.081)***	-0.011 (0.079)	0.144 (0.034)***	-0.086 (0.043)*	-0.012 (0.040)
4	CR220Q04S	-0.162 (0.079)*	-0.167 (0.078)*	-0.025 (0.031)	-0.051 (0.042)	-0.033 (0.040)
5	CR220Q05S	-0.254 (0.083)**	-0.093 (0.080)	0.099 (0.031)**	-0.098 (0.044)*	-0.039 (0.041)
6	CR220Q06S	-0.184 (0.077)*	-0.040 (0.075)	0.000 (0.030)	-0.074 (0.041)	-0.034 (0.039)
7	CR227Q01S	-0.310 (0.083)***	-0.225 (0.081)**	0.050 (0.032)	-0.049 (0.043)	-0.113 (0.043)**
8	CR227Q02S	-0.111 (0.111)	-0.029 (0.114)	-0.005 (0.044)	-0.095 (0.062)	-0.088 (0.060)
9	CR111Q01S	-0.298 (0.079)***	-0.153 (0.079)*	0.047 (0.031)	-0.088 (0.042)*	-0.056 (0.041)
10	CR055Q01S	-0.196 (0.080)*	-0.063 (0.078)	0.089 (0.032)**	-0.057 (0.041)	-0.070 (0.041)
11	CR453Q01S	-0.043 (0.080)	-0.050 (0.078)	-0.020 (0.033)	-0.053 (0.044)	0.118 (0.046)*
12	CR453Q05S	-0.336 (0.082)***	-0.182 (0.083)*	0.111 (0.035)**	-0.105 (0.049)*	-0.119 (0.046)**
13	CR412Q01S	-0.444 (0.101)***	-0.146 (0.093)	0.092 (0.039)*	-0.108 (0.050)*	-0.028 (0.052)
14	CR412Q05S	-0.106 (0.076)	-0.155 (0.076)*	0.010 (0.031)	-0.021 (0.043)	-0.004 (0.042)
15	CR412Q06S	-0.152 (0.077)	-0.141 (0.077)	0.048 (0.032)	0.037 (0.045)	-0.002 (0.043)
16	CR437Q01S	0.039 (0.079)	-0.038 (0.079)	0.065 (0.033)*	-0.059 (0.047)	0.004 (0.044)
17	CR437Q06S	-0.116 (0.076)	-0.140 (0.075)	0.033 (0.031)	0.008 (0.043)	0.030 (0.042)
18	CR456Q01S	-0.549 (0.192)**	-0.352 (0.154)*	0.008 (0.067)	-0.173 (0.086)*	-0.008 (0.091)
19	CR466Q03S	0.153 (0.146)	0.114 (0.149)	-0.048 (0.056)	0.095 (0.093)	0.051 (0.083)
20	CR446Q03S	-0.238 (0.087)**	-0.038 (0.086)	0.010 (0.033)	-0.095 (0.049)	-0.054 (0.047)
21	CR432Q06S	-0.169 (0.302)	-0.049 (0.343)	-0.135 (0.107)	-0.101 (0.214)	-0.176 (0.162)
22	CR460Q05S	-0.249 (0.079)**	-0.024 (0.080)	0.045 (0.030)	-0.039 (0.046)	-0.083 (0.043)
23	CR460Q06S	-0.172 (0.077)*	-0.076 (0.079)	0.027 (0.030)	-0.035 (0.046)	-0.061 (0.043)
24	CR424Q02S	0.045 (0.087)	0.031 (0.090)	-0.003 (0.035)	0.016 (0.048)	0.031 (0.047)
25	CR424Q03S	-0.012 (0.077)	-0.007 (0.081)	-0.056 (0.031)	0.038 (0.042)	0.062 (0.041)
26	CR424Q07S	-0.163 (0.086)	-0.051 (0.089)	-0.038 (0.036)	-0.003 (0.047)	-0.024 (0.045)
27	CR404Q03S	-0.228 (0.080)**	-0.116 (0.082)	0.031 (0.032)	-0.052 (0.044)	-0.052 (0.042)
28	CR404Q06S	-0.118 (0.080)	-0.081 (0.085)	0.002 (0.033)	0.045 (0.044)	-0.007 (0.043)
29	CR404Q07S	-0.242 (0.117)*	-0.019 (0.120)	0.111 (0.054)*	0.048 (0.063)	-0.100 (0.062)
30	CR455Q04S	-0.404 (0.078)***	-0.169 (0.080)*	0.095 (0.032)**	-0.103 (0.044)*	-0.091 (0.041)*
31	CR455Q05S	-0.368 (0.157)*	-0.119 (0.165)	0.072 (0.068)	-0.098 (0.088)	-0.266 (0.078)***
32	CR083Q01S	-0.085 (0.077)	-0.174 (0.078)*	0.018 (0.031)	0.020 (0.041)	-0.058 (0.041)
33	CR083Q03S	-0.147 (0.077)	-0.143 (0.076)	0.038 (0.031)	0.017 (0.041)	0.089 (0.043)*
34	CR083Q04S	-0.072 (0.076)	-0.009 (0.075)	-0.008 (0.031)	0.057 (0.041)	-0.054 (0.041)
35	CR442Q07S	-0.177 (0.083)*	-0.150 (0.084)	0.012 (0.033)	-0.017 (0.044)	-0.013 (0.046)
36	CR245Q01S	-0.120 (0.076)	0.049 (0.075)	0.083 (0.031)**	-0.032 (0.041)	-0.025 (0.041)
37	CR245Q02S	-0.276 (0.077)***	-0.198 (0.077)*	0.029 (0.031)	0.006 (0.041)	-0.018 (0.041)
38	CR101Q01S	0.006 (0.080)	0.013 (0.079)	0.014 (0.033)	-0.008 (0.043)	-0.025 (0.044)
39	CR101Q02S	-0.229 (0.086)**	-0.043 (0.084)	0.054 (0.034)	0.006 (0.046)	0.003 (0.045)
40	CR101Q03S	-0.240 (0.077)**	-0.092 (0.076)	0.036 (0.031)	-0.013 (0.041)	-0.004 (0.041)
41	CR101Q04S	-0.250 (0.080)**	-0.041 (0.079)	0.082 (0.032)**	-0.105 (0.043)*	-0.033 (0.042)
42	CR101Q05S	0.025 (0.084)	-0.047 (0.084)	-0.028 (0.033)	-0.020 (0.045)	0.063 (0.046)
43	DR219Q01EC	-0.092 (0.088)	0.003 (0.087)	0.076 (0.038)*	0.024 (0.051)	-0.061 (0.048)
44	DR219Q01C	0.013 (0.087)	0.203 (0.088)*	0.054 (0.037)	-0.055 (0.049)	0.014 (0.049)
45	DR219Q02C	0.126 (0.093)	0.018 (0.090)	0.039 (0.037)	0.035 (0.052)	-0.003 (0.048)
46	DR067Q04C	-0.048 (0.080)	-0.073 (0.082)	0.028 (0.033)	-0.009 (0.046)	-0.000 (0.044)
47	DR067Q05C	-0.157 (0.080)	-0.210 (0.083)*	0.087 (0.034)*	0.027 (0.047)	-0.022 (0.044)
48	DR102Q04C	-0.095 (0.114)	-0.068 (0.126)	0.045 (0.050)	-0.109 (0.073)	-0.062 (0.065)
49	DR102Q05C	-0.190 (0.106)	-0.022 (0.115)	0.125 (0.052)*	-0.064 (0.065)	-0.115 (0.061)
50	CR220Q01S	-0.290 (0.184)	-0.193 (0.231)	0.140 (0.096)	-0.167 (0.120)	0.052 (0.103)
51	DR227Q03C	-0.189 (0.095)*	-0.102 (0.095)	0.028 (0.038)	0.079 (0.054)	-0.080 (0.053)
52	DR227Q06C	-0.172 (0.085)*	-0.002 (0.085)	0.120 (0.036)***	0.001 (0.045)	-0.006 (0.047)

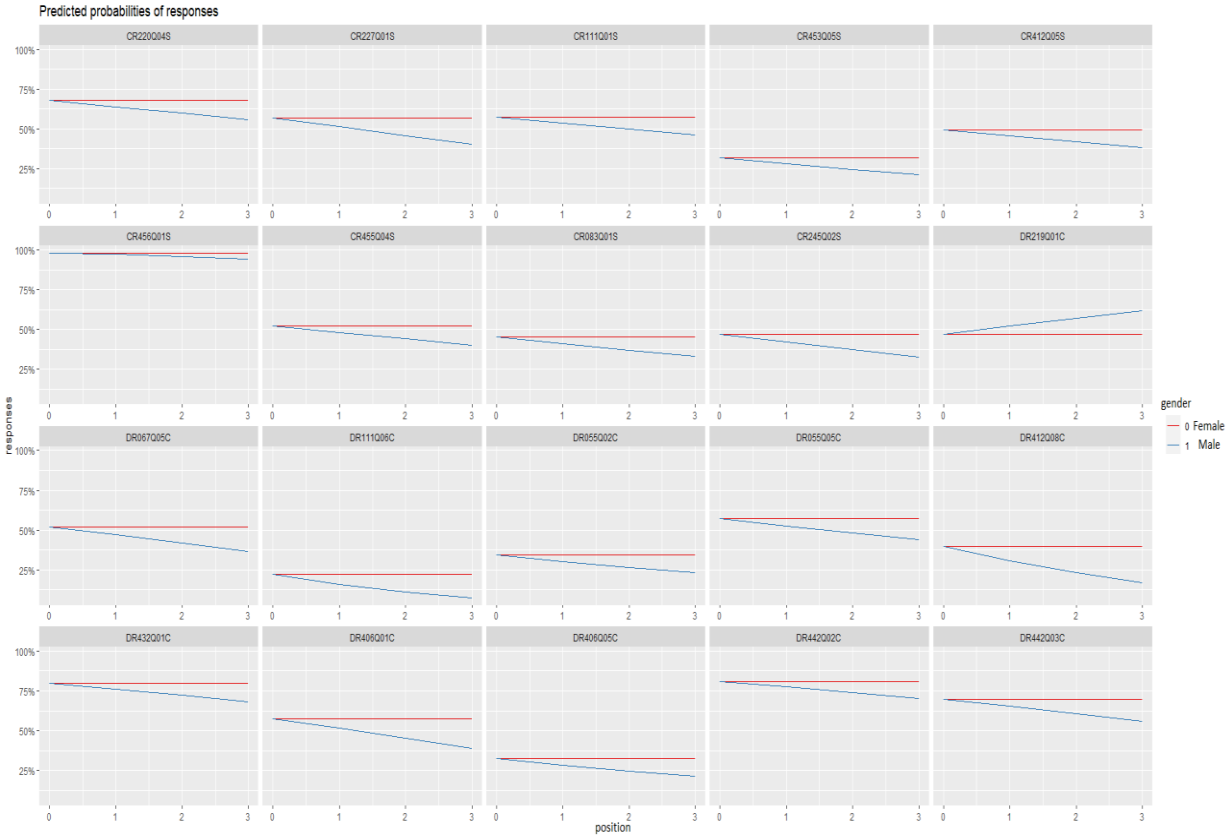
53	DR111Q02BC	0.145 (0.110)	0.002 (0.117)	0.032 (0.048)	0.141 (0.058)*	0.091 (0.063)
54	DR111Q06C	-0.312 (0.103)**	-0.428 (0.128)***	0.113 (0.048)*	0.035 (0.056)	-0.064 (0.060)
55	DR055Q02C	-0.225 (0.087)**	-0.183 (0.092)*	0.084 (0.038)*	-0.053 (0.049)	-0.025 (0.049)
56	DR055Q03C	-0.148 (0.097)	-0.111 (0.108)	0.092 (0.045)*	-0.042 (0.056)	-0.133 (0.059)*
57	DR055Q05C	-0.264 (0.081)**	-0.173 (0.084)*	0.083 (0.033)*	-0.057 (0.044)	-0.083 (0.046)
58	CR104Q01S	-0.150 (0.080)	-0.120 (0.082)	0.049 (0.034)	-0.006 (0.044)	-0.055 (0.045)
59	CR104Q02S	-0.037 (0.079)	0.140 (0.079)	0.059 (0.034)	-0.063 (0.044)	-0.012 (0.045)
60	CR104Q05S	0.143 (0.430)	0.073 (0.422)	-0.057 (0.168)	-0.053 (0.255)	-0.043 (0.253)
61	DR420Q02C	-0.008 (0.078)	0.024 (0.079)	0.003 (0.033)	-0.059 (0.044)	-0.055 (0.043)
62	DR420Q10C	-0.029 (0.082)	-0.147 (0.084)	0.126 (0.036)***	0.032 (0.050)	0.047 (0.048)
63	DR420Q06C	-0.080 (0.095)	-0.142 (0.092)	0.029 (0.039)	-0.059 (0.051)	-0.016 (0.050)
64	DR420Q09C	-0.176 (0.086)*	-0.161 (0.083)	0.055 (0.035)	0.036 (0.048)	0.047 (0.047)
65	DR453Q04C	-0.023 (0.081)	0.088 (0.085)	-0.009 (0.035)	0.124 (0.049)*	0.076 (0.047)
66	DR453Q06C	-0.060 (0.079)	-0.136 (0.081)	0.087 (0.036)*	0.026 (0.046)	0.048 (0.046)
67	DR412Q08C	-0.161 (0.086)	-0.386 (0.097)***	0.117 (0.040)**	0.022 (0.050)	0.042 (0.050)
68	DR437Q07C	-0.172 (0.095)	-0.147 (0.103)	0.074 (0.044)	0.014 (0.055)	0.054 (0.057)
69	DR456Q02C	-0.234 (0.095)*	-0.138 (0.090)	0.065 (0.037)	-0.037 (0.046)	-0.029 (0.049)
70	DR456Q06C	-0.041 (0.089)	-0.138 (0.087)	0.083 (0.036)*	0.001 (0.045)	0.052 (0.048)
71	DR466Q02C	-0.047 (0.088)	0.062 (0.090)	0.014 (0.038)	-0.053 (0.051)	0.067 (0.053)
72	CR466Q06S	-0.271 (0.082)**	-0.024 (0.080)	0.092 (0.033)**	-0.043 (0.042)	-0.029 (0.045)
73	DR446Q06C	-0.211 (0.078)**	-0.148 (0.079)	0.086 (0.033)**	0.016 (0.041)	-0.034 (0.043)
74	DR432Q01C	-0.274 (0.088)**	-0.209 (0.083)*	0.099 (0.035)**	-0.017 (0.044)	-0.013 (0.047)
75	DR432Q05C	0.008 (0.078)	-0.124 (0.078)	0.043 (0.032)	0.014 (0.041)	0.054 (0.044)
76	DR460Q01C	-0.169 (0.081)*	-0.154 (0.084)	0.092 (0.035)**	0.036 (0.044)	-0.015 (0.047)
77	DR404Q10AC	-0.155 (0.085)	-0.124 (0.091)	0.095 (0.037)*	-0.010 (0.046)	0.044 (0.048)
78	DR404Q10BC	-0.108 (0.084)	-0.059 (0.090)	0.103 (0.038)**	-0.051 (0.047)	0.026 (0.048)
79	DR406Q01C	-0.181 (0.079)*	-0.253 (0.082)**	0.061 (0.031)	0.063 (0.042)	0.017 (0.044)
80	DR406Q05C	-0.098 (0.083)	-0.188 (0.090)*	-0.006 (0.033)	0.016 (0.045)	-0.012 (0.047)
81	DR406Q02C	-0.019 (0.102)	-0.005 (0.109)	0.008 (0.043)	0.010 (0.057)	-0.007 (0.060)
82	DR455Q02C	-0.224 (0.090)*	-0.053 (0.094)	0.041 (0.037)	0.018 (0.047)	-0.000 (0.050)
83	DR455Q03C	-0.211 (0.093)*	-0.083 (0.095)	0.021 (0.037)	0.023 (0.049)	-0.005 (0.051)
84	CR083Q02S	-0.119 (0.078)	0.003 (0.076)	0.063 (0.032)*	-0.011 (0.044)	-0.030 (0.042)
85	DR442Q02C	-0.224 (0.088)*	-0.199 (0.083)*	0.054 (0.034)	-0.013 (0.048)	0.057 (0.047)
86	DR442Q03C	-0.165 (0.085)	-0.197 (0.085)*	0.044 (0.034)	-0.045 (0.048)	0.004 (0.047)
87	DR442Q05C	-0.184 (0.097)	-0.202 (0.113)	0.146 (0.046)**	-0.030 (0.057)	-0.069 (0.054)
88	DR442Q06C	-0.093 (0.090)	-0.157 (0.096)	0.038 (0.037)	0.085 (0.052)	0.067 (0.051)

Appendix B

Item-Position Interaction on Item Basis

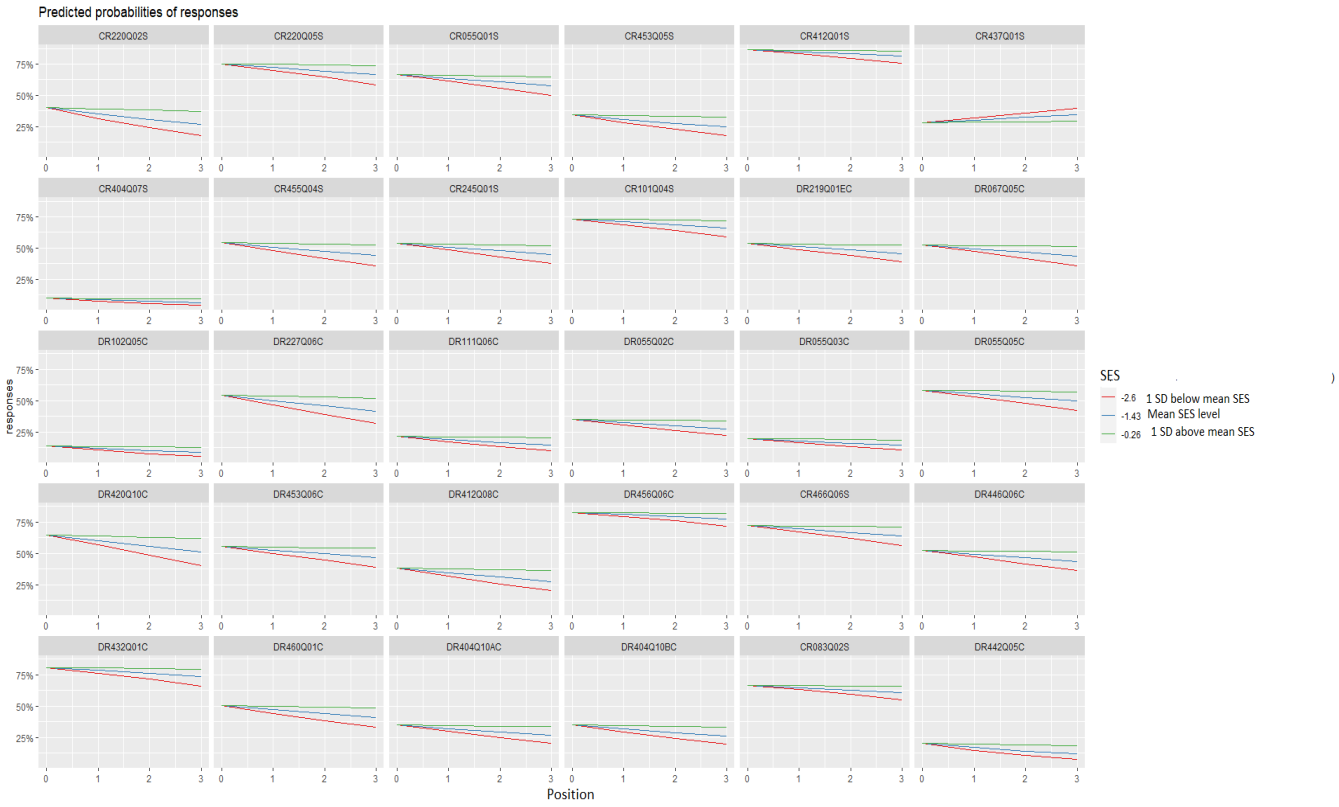


Appendix C


Predicted probabilities of responses
CR220Q04S
CR227Q01S
CR111Q01S
CR453Q05S
CR412Q05S
CR456Q01S
CR455Q04S
CR083Q01S
CR245Q02S
DR219Q01C
DR067Q05C
DR111Q06C
DR055Q02C
DR055Q05C
DR412Q08C
DR432Q01C
DR406Q01C
DR406Q05C
DR442Q02C
DR442Q03C
responses
position
gender
0 Female
1 Male

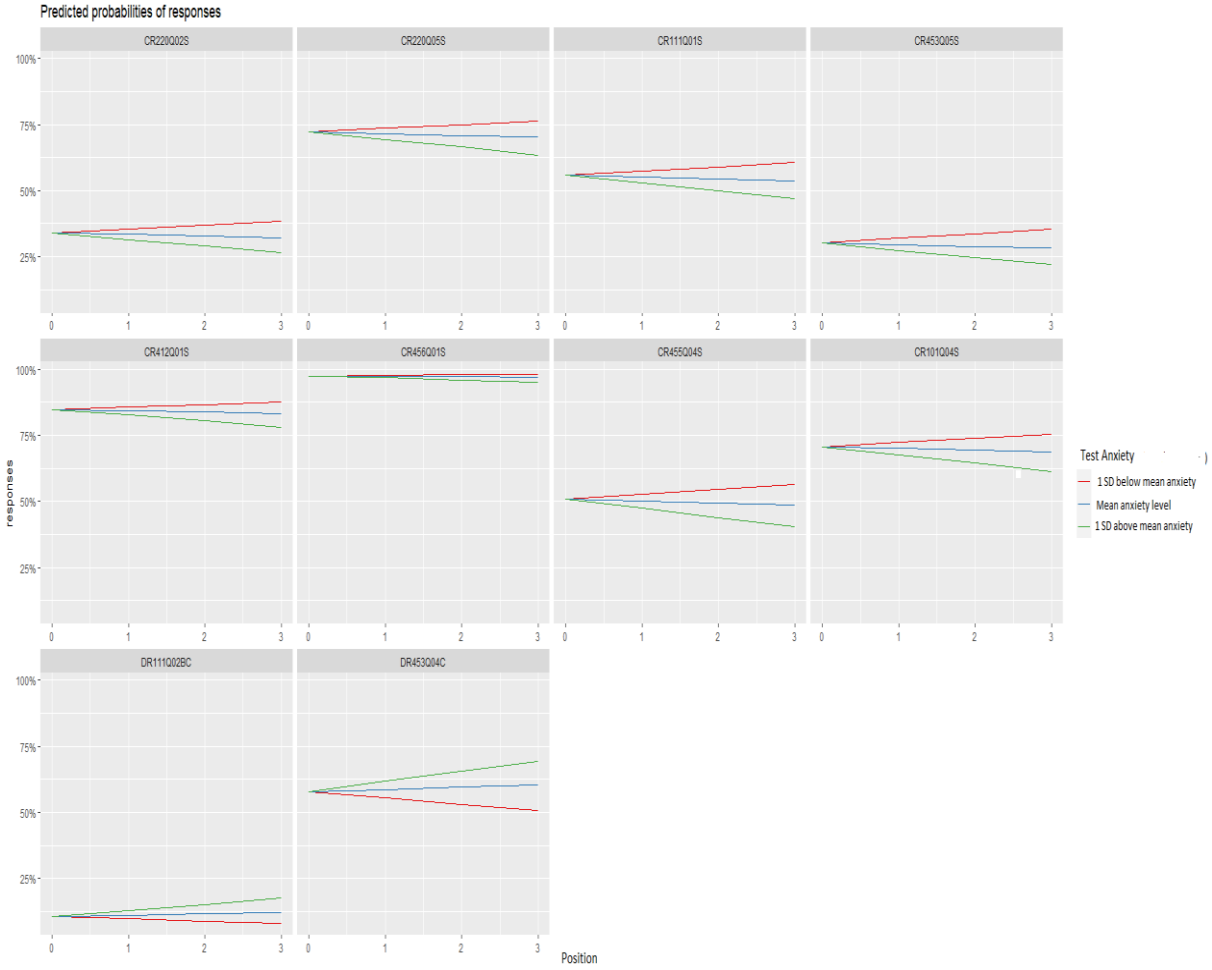
Appendix D

Item-position-SES interaction on item basis



Appendix E

Item-position-anxiety interaction on item basis



Appendix F

Item-position-motivation interaction on item basis