International Journal of Assessment Tools in Education
2022, Vol. 9, No. 3, 772–786

https://doi.org/10.21449/ijate.1173356

Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate ***Research Article***

A novel approach for calculating the item discrimination for Likert type of scales

Umit Celen[1,*], Eren Can Aybek[2]

[1]Amasya University, Faculty of Education, Department of Educational Sciences, Amasya, Türkiye
[2]Pamukkale University, Faculty of Education, Department of Educational Sciences, Denizli, Türkiye

ARTICLE HISTORY

Received: May. 21, 2022

Accepted: Sep. 04, 2022

Keywords:

Item discrimination index,
Likert-type scale,
Exploratory factor analysis,
Slope coefficient,
Monte-Carlo simulation.

Abstract: Item analysis is performed by developers as an integral part of the scale development process. Thus, items are excluded from the scale depending on the item analysis prior to the factor analysis. Existing item discrimination indices are calculated based on correlation, yet items with different response patterns are likely to have a similar item discrimination index. This study proposed a new item discrimination index that can be used in Likert type of scales and examined its effect on factor analysis results. For this purpose, simulative datasets were generated, and items were excluded from the analysis according to the .20, .30 and .35 item discrimination index criteria, and exploratory factor analysis was performed for a single factor. Accordingly, it was found that more variance could be explained by a single factor with fewer items compared to other discrimination indices when the .20 criterion of the slope coefficient was used as suggested in this study. Similar findings were obtained using the .35 criterion with other discrimination indices. In this context, it is recommended to use the slope coefficient as an additional discrimination index calculation method in the scale development process.

1. INTRODUCTION

Although validity and reliability are the features related to the scores obtained with the measurement tool, and not the measurement tool itself, the qualities of the measurement tool affect the validity and reliability of the scores obtained with that instrument. Qualities such as having items that include the characteristic to be measured and being prepared in accordance with the guidelines of item writing can be examined with the help of expert opinion. On the other hand, statistical methods are used to measure the difficulty levels of the items or whether they can distinguish among the individuals who more or less have the characteristic to be measured. According to the classical test theory, the correct answer rate of an item by the group constitutes the item difficulty while a wide variety of statistics have been developed to detect item discrimination. Long and Sandiford reported that 23 different methods were defined to calculate the item discrimination index even in 1935 (as cited in Oosterhof, 1976). Kelley (1939) presented that among these methods, the most appropriate findings could be obtained in

*CONTACT: Umit Celen umit.celen@amasya.edu.tr Amasya University, Faculty of Education, Department of Educational Sciences, Amasya, Türkiye

e-ISSN: 2148-7456 /© **IJATE** 2022

the method based on the comparison of the lower and upper groups when the group sizes were 27% and Johnson (1951), on the other hand, made corrections in this formula and suggested the formula for the lower - upper 27% groups method used today. In addition, methods based on the correlation between the item and the total score are also frequently used to determine item discrimination. However, this way of calculating the correlation also gave rise to different methods.

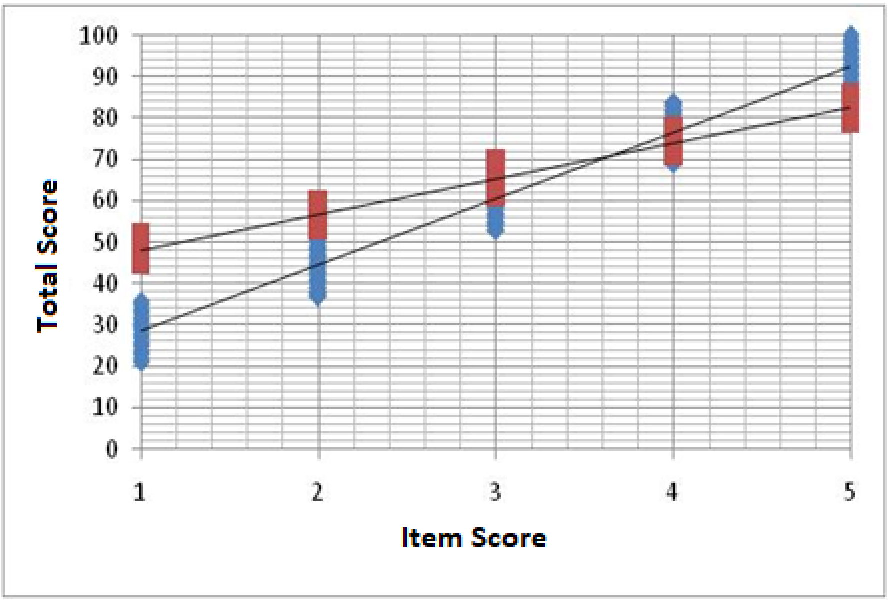
For dichotomous items, the item score has two categories, while the total score is a continuous variable. For this reason, biserial or point biserial correlation coefficients are used to calculate the correlation between a two-category discontinuous variable and a continuous variable (Popham, 2014). The main difference between these two correlation coefficients is that the point biserial correlation coefficient assume that the variable (item score) is true categorical in nature, whereas the biserial correlation coefficient assumes that the categorical variable actually has a continuous nature but has been artificially made discontinuous (Crocker & Algina, 2008). On the other hand, Guilford (1965) suggested using the point biserial correlation coefficient as it provides more information about the contribution of each item to the predictive validity of the test. Henrysson (1971) stated that the biserial correlation coefficient can be used if the total score is normally distributed (as cited in Oosterhof, 1976).

Although these methods suggested for item discrimination index were first generated for items scored 1-0, they are also used for items scored in polytomous categories (e.g. Likert type). However, in this case, the item-total correlation is calculated with the Pearson product moments correlation coefficient instead of the biserial or point biserial correlation coefficient and since there are no correct and incorrect answers, the upper-lower groups method is calculated by taking into account the difference between the item score averages of the upper and lower groups. However, since the total score obtained from the test also includes the item score whose discrimination is to be calculated, the correlation coefficient calculated between the item and the total score gives an overestimate of discrimination. For this reason, the correlation coefficient between the item and the total score obtained from the other items in the test (item-rest correlation) is also used in the calculation of item discrimination. On the other hand, polyserial correlation also could be used instead of Pearson correlation, if one assumes item score as ordinal and total score as continuous (Moses, 2017).

Either by using a correlation-based or group comparison-based item discrimination coefficient, the main purpose of an item discrimination index is to show whether individuals more or less exhibiting the measured trait also respond to the item in a similar way. However, it is stated (Livingston & Dorans, 2004) that a graphical method should also be followed in the examination of item discrimination. In this context, it is seen that when the items with the same item discrimination index are examined graphically, they distinguish individuals differently. Figure 1 provides a sample graphic. As Figure 1 shows, the two items indicated by blue and red dots distinguish individuals differently. However, the correlation coefficient between the item and the total score for both items is obtained as .98. This case reveals the necessity of considering different methods in conjunction when item discrimination index is calculated.

On the other hand, the test development process includes performing exploratory factor analysis to obtain proof of construct validity after the implementation of the draft items and investigating the common structure under which the items are joined (Tabachnick & Fidell, 2013). Before exploratory factor analysis, item discrimination is performed to exclude the items with low discrimination from the analysis at the very beginning.

Figure 1. *Response plot for two different items.*



1.1. Current Study

This study proposed a new item discrimination index to be used alongside the existing item discrimination indices. Equation 1 provides this coefficient, called the slope coefficient.

$$Slope\ Coefficient = \frac{\bar{X}_n - \bar{X}_m}{n - m} \qquad (1)$$

In this equation, $\bar{X}_n$ is the total mean score of the participants choosing the highest category; $\bar{X}_m$ is the total mean score of the participants choosing the lowest category; k is the number of items in the scale; n represents the point value of the highest category and m represents the point value of the lowest category. When calculating the average score of individuals, some researchers reduce the individual's score to the response category range by dividing the total scores by the number of items, instead of the average of the total score obtained from the scale. In this situation, for example, the individual's score from the scale is obtained in the range of 1-5, for a 5-point Likert scale. In this case, there is no need to use the k value in the equation.

The present study aimed to examine the effect of *slope coefficient (sc)* on the total variance explained by the exploratory factor analysis and in this context, to compare the performance of sc with other item discrimination indices.

2. METHOD

2.1. Data

The research data were generated in R (R Core Team, 2022) using the *genPolyMatrix* function of the *catR v3.16* (Magis & Raiche, 2012) package. The *catR* package generates data based on Item Response Theory (IRT). Although this research was carried out according to the Classical Test Theory (CTT), this package was preferred because of its convenience in producing the item pool and response pattern.

Both the item pool and the sample size were controlled during data generation. Accordingly, the item pool size was 10, 30, 50 and 100, respectively and sample size was assigned as 50, 100, 250, 500 and 1000, respectively. Hence, a total of 20 different response patterns were generated simulatively, in four different item pools and five different samples. In addition, the number of replications was determined to be 100 and each simulation was repeated 100 times. The response category was chosen as 5. Since the item parameters were generated according to

IRT with *catR*, the item discrimination indices of the items would be quite high. To prevent this, the item discrimination *a* parameters of 30% of the generated items were corrected. For this, parameter *a* was randomly assigned from a normal distribution with a mean of 0.3 and a standard deviation of 0.05. Then, for each item pool, the response pattern with the specified sample size was generated using the *genPattern* function. As a result, a total of 2000 data files with a total of 20 conditions and 100 replications with different item pools and sample sizes were examined in the study.

2.2. Data Analysis

Simulative data were examined in the study and five different coefficients were calculated for item discrimination: Polyserial correlation coefficient, item-total correlation coefficient, item-rest correlation coefficients upper-lower 27% groups method and slope coefficient. The polyserial correlation coefficients were calculated by using the *polyserial* function in the *psych v2.2.5* (Revelle, 2022) package and the item discrimination indices found via item-total, item-rest, and upper-lower 27% groups method were calculated by using the *ItemAnalysis* function in the *ShinyItemAnalysis v1.4.1* (Martinkova & Drabinova, 2018) package. The slope coefficient, constituting the essence of the research, was calculated by transferring Equation-1 to R.

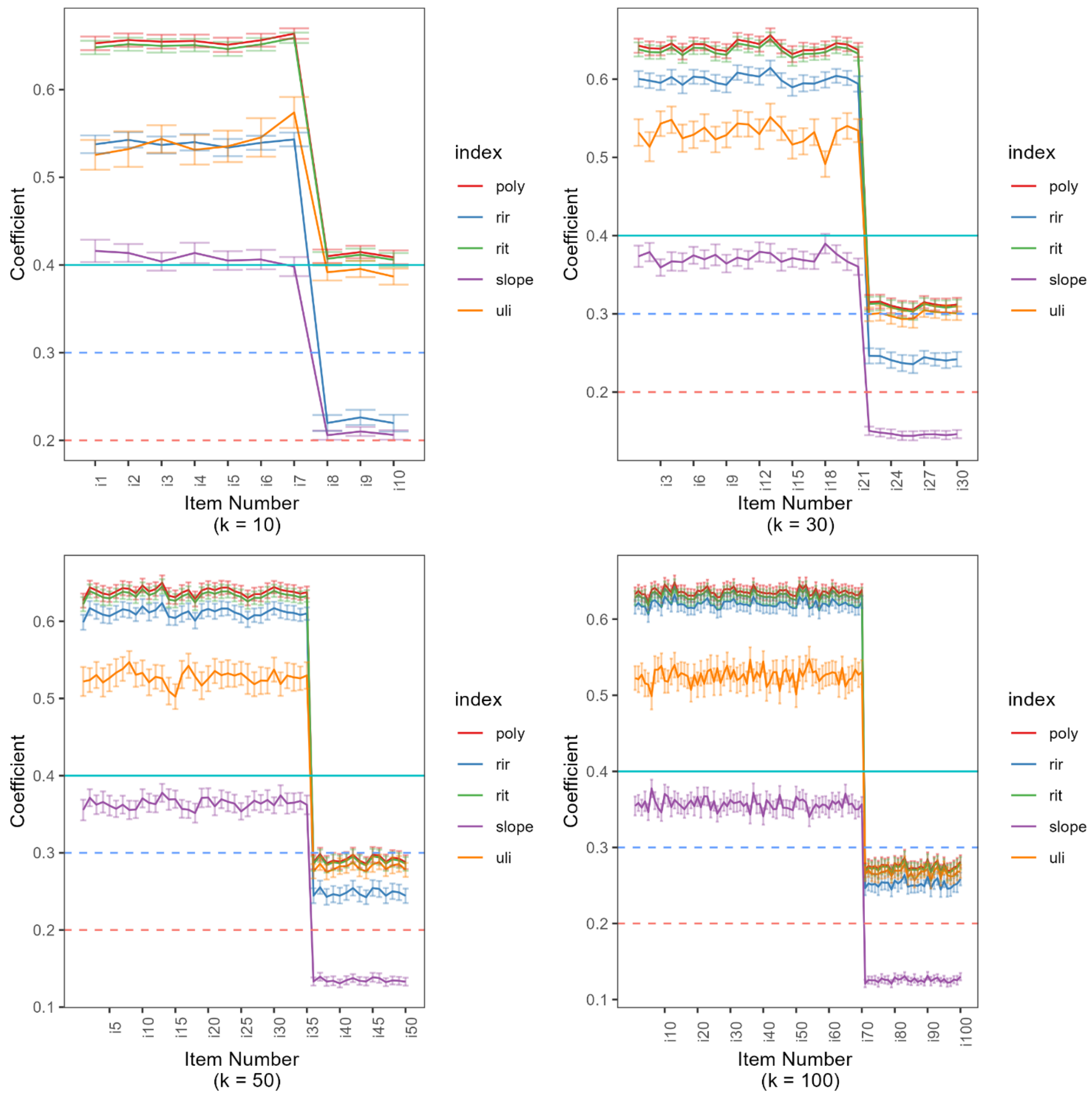
The relevant coefficients were separately calculated for each replication and then the item discrimination index averages and 95% confidence interval values for each item were visualized with the help of the *dplyr v1.0.9* (Wickham et.al., 2022) and *ggplot2 v3.3.6* (Wickham, 2016) packages. Then, .20, .30 and .35 values were accepted as criteria, respectively, and the items below the criteria were excluded from the data set and factor analysis was performed with the remaining items. The number of items included in the factor analysis and the variance rates explained by these items in a single factor were reported by calculating the mean and 95% confidence intervals. The R script used in data generation and analysis can be accessed via https://www.github.com/anonym [The full URL will be provided if the manuscript is approved. URL is hidden for the purpose of anonymity].

3. RESULTS

The item discrimination indices were the first findings obtained as a result of data analysis. For ease of interpretation, the calculated item discrimination indices for each item pool and sample size were plotted, but in order not to disrupt the flow of the text, only the plot for the sample size of 1000 was provided in Figure 2. Appendix lists the plots obtained for all sample sizes. The mean value of 100 replications for each item was taken as the basis for obtaining the graphs, but the 95% confidence intervals were also presented in the graph.

Accordingly, for each item pool size, item discrimination values were found to decrease in the last 30% of the item pool. While creating the simulation conditions, the last 30% of each item pool was manipulated and deliberately reduced. Therefore, this was an expected result. On the other hand, the slope coefficient for each item pool size generated significantly lower values than the other item discrimination indices, except when the item pool size was 10. This can be seen from the fact that the confidence intervals did not intersect. In addition, when the number of items exceeds 50, the item discrimination indices obtained by the item-total, item-rest, and upper-lower methods were quite similar to each other; polyserial correlation coefficient was significantly lower and the slope coefficient provided the lowest value among the five item discrimination indices.

Figure 2. *Item discrimination indices for n = 1000 in different item pool sizes.*



poly: polyserial correlation; rir: item-rest correlation; rit: item-total correlation; uli: upper-lower 27% groups; slope: slope coefficient.

However, item discrimination indices provided similar results for items specifically when the number of items was 100, in the last 30% of the item pool and expected to be non-discriminatory, excluding the slope coefficient. Another finding that emerged after examining the graphics in Figure-2 and Annex-1 showed that the slope coefficient generally generated lower values; on the other hand, the remaining four item discrimination indices and items expected to have poor discrimination values were above .20. Then, using .20, .30 and .35 criteria, respectively, items with an item discrimination index below this criterion were excluded from the test, and exploratory factor analysis was performed with the remaining items. Tables 1, 2, and 3 provide the mean number of remaining items, the mean of variance explained by a single factor, and the 95% confidence intervals for both values.

Table 1. *The number of items included in the EFA and the mean ratio of variance explained by a single factor for item discrimination index criterion .20*

		k = 10		k = 30		k = 50		k = 100	
n	Method	k_r [95% CI]	Var [95% CI]	k_r [95% CI]	Var [95% CI]	k_r [95% CI]	Var [95% CI]	k_r [95% CI]	Var [95% CI]
50	SC	8.72 [8.53-8.92]	.36 [.35-.37]	22.84 [22.45-23.23]	.34 [.33-.34]	36.89 [36.32-37.46]	.40 [.40-.41]	72.56 [71.65-73.47]	.40 [.40-.41]
	Poly	9.90 [9.84-9.96]	.32 [.31-.33]	28.32 [28.08-28.56]	.33 [.32-.34]	46.77 [46.42-46.12]	.34 [.33-.35]	91.47 [90.94-92.00]	.34 [.34-.35]
	ULI	9.68 [9.57-9.79]	.33 [.32-.34]	27.67 [27.04-27.94]	.34 [.33-.34]	45.75 [45.40-46.10]	.34 [.34-.35]	89.73 [89.12-90.34]	.35 [.34-.35]
	RIT	9.90 [9.84-9.96]	.32 [.31-.33]	28.26 [28.02-28.50]	.33 [.32-.34]	46.61 [46.26-46.96]	.34 [.33-.35]	91.28 [90.75-91.82]	.34 [.34-.35]
	RIR	8.80 [8.61-8.99]	.36 [.35-.37]	26.71 [26.40-27.02]	.35 [.34-.36]	45.12 [44.69-45.55]	.35 [.34-.36]	89.72 [89.13-90.31]	.35 [.34-.35]
100	SC	8.78 [8.60-8.96]	.35 [.34-.36]	22.35 [22.09-22.62]	.39 [.39-.40]	36.81 [36.46-37.17]	.40 [.39-.40]	71.80 [71.25-72.35]	.40 [.40-.41]
	Poly	9.97 [9.94-10.0]	.31 [.30-.32]	28.99 [28.84-29.14]	.32 [.31-.32]	47.42 [47.09-47.76]	.33 [.32-.33]	93.55 [93.21-94.11]	.33 [.32-.33]
	ULI	9.95 [9.91-9.99]	.31 [.30-.32]	28.39 [28.19-28.59]	.32 [.32-.33]	46.80 [46.48-47.13]	.33 [.32-.33]	92.26 [91.76-92.76]	.33 [.33-.34]
	RIT	9.97 [9.94-10.0]	.31 [.30-.32]	28.96 [28.80-29.12]	.32 [.31-.32]	47.31 [46.98-47.64]	.33 [.32-.33]	93.35 [92.89-93.81]	.33 [.33-.33]
	RIR	8.86 [8.69-9.03]	.31 [.30-.32]	27.18 [26.90-27.46]	.34 [.33-.34]	45.59 [45.20-45.98]	.34 [.33-.34]	91.28 [90.74-91.83]	.34 [.33-.34]
250	SC	8.84 [8.66-9.02]	.34 [.33-.34]	21.72 [21.54-21.90]	.40 [.39-.40]	35.56 [35.39-35.73]	.40 [.39-.40]	70.79 [70.56-71.03]	.40 [.40-.40]
	Poly	10.0 [10.0-10.0]	.30 [.29-.31]	29.52 [29.31-29.65]	.31 [.30-.31]	48.55 [48.31-48.79]	.31 [.31-.32]	95.44 [95.05-95.83]	.32 [.31-.32]
	ULI	9.99 [9.97-10.0]	.30 [.29-.31]	29.15 [28.97-29.33]	.31 [.31-.32]	47.84 [47.57-48.11]	.32 [.31-.32]	94.09 [93.60-94.58]	.32 [.32-.32]
	RIT	10.0 [10.0-10.0]	.30 [.29-.31]	29.50 [29.37-29.63]	.31 [.30-.31]	48.43 [48.19-48.67]	.31 [.31-.32]	95.23 [94.82-95.65]	.32 [.31-.32]
	RIR	9.00 [8.84-9.16]	.33 [.33-.34]	27.56 [27.28-27.84]	.33 [.32-.33]	46.23 [45.90-46.56]	.33 [.32-.33]	92.77 [92.21-93.33]	.32 [.32-.33]
500	SC	8.90 [8.73-9.07]	.33 [.33-.34]	21.44 [21.31-21.58]	.40 [.39-.40]	35.30 [35.19-35.41]	.40 [.40-.40]	70.25 [70.12-70.39]	.40 [.40-.40]
	Poly	10.0 [10.0-10.0]	.30 [.30-.30]	29.85 [29.77-29.93]	.30 [.30-.31]	49.23 [49.06-49.40]	.31 [.30-.31]	97.00 [96.65-97.35]	.31 [.31-31]
	ULI	9.99 [9.97-10.0]	.30 [.30-.30]	29.65 [29.53-29.77]	.31 [.30-.31]	48.76 [48.57-48.95]	.31 [.31-.31]	96.05 [95.64-96.46]	.31 [.31-.31]
	RIT	10.0 [10.0-10.0]	.30 [.30-.30]	29.83 [29.75-29.91]	.30 [.30-.31]	49.18 [49.00-49.36]	.31 [.30-.31]	96.86 [96.51-97.21]	.31 [.31-.31]
	RIR	9.11 [8.96-9.26]	.33 [.32-.33]	28.03 [27.81-28.26]	.32 [.32-.32]	46.88 [46.55-47.21]	.32 [.32-.32]	94.54 [94.08-95.00]	.32 [.31-.32]
1000	SC	8.76 [8.59-8.93]	.34 [.33-.34]	21.30 [21.18-21.42]	.40 [.39-.40]	35.29 [35.18-35.40]	.40 [.39-.40]	70.18 [70.10-70.26]	.40 [.40-.40]
	Poly	10.0 [10.0-10.0]	.30 [.29-.30]	29.88 [29.81-29.95]	.30 [.30-.30]	49.56 [49.43-49.69]	.30 [.30-.31]	98.01 [97.73-98.29]	.31 [.30-.31]
	ULI	10.0 [10.0-10.0]	.30 [.29-.30]	29.77 [29.68-29.86]	.30 [.30-.31]	49.31 [49.17-49.47]	.30 [.30-.31]	97.46 [97.14-97.79]	.31 [.30-.31]
	RIT	10.0 [10.0-10.0]	.30 [.29-.30]	29.88 [29.81-29.95]	.30 [.30-.30]	49.53 [49.40-49.66]	.30 [.30-.31]	97.91 [97.62-98.20]	.31 [.30-.31]
	RIR	9.06 [8.90-9.22]	.33 [.33-.34]	28.30 [28.04-28.56]	.32 [.31-.32]	47.41 [47.13-47.69]	.31 [.31-.32]	95.50 [95.03-95.97]	.31 [.31-.31]

SC: Slope coefficient; Poly: Polyserial correlation; ULI: Upper-Lower index; RIT: Item-total correlation; RIR: Item-rest correlation.

When the criterion for the item discrimination index was set to .20, it was observed that a similar number of items remained in the test with the help of the slope coefficient for 10 items and the item-rest correlation and a similar variance was explained by a single factor. Other item discrimination indices left more items in the test but explained a lower rate of variance. However, in cases where the number of items in the pool exceeded 30 and the sample size was 100 or more, the slope coefficient excluded more items than the others, and accordingly, a higher rate of variance could be explained by a single factor. For example, when the sample size was 1000 and the item pool size was 100; an average of 70.18 [70.10-70.39] items were included in the factor analysis with the slope coefficient, while an average of 95.50 [95.03-95.97] items were included in the factor analysis according to the item-rest correlation.

Table 2. *The number of items included in the EFA and the mean ratio of variance explained by a single factor for item discrimination index criterion .30*

		k = 10		k = 30		k = 50		k = 100	
n	Method	k_r [95% CI]	Var [95% CI]	k_r [95% CI]	Var [95% CI]	k_r [95% CI]	Var [95% CI]	k_r [95% CI]	Var [95% CI]
50	SC	6.73 [6.47-6.99]	.42 [.41-.43]	16.63 [15.91-17.35]	.44 [.43-.44]	27.19 [26.11-28.26]	.45 [.44-.45]	52.66 [50.50-54.83]	.45 [.44-.45]
	Poly	9.47 [9.34-9.60]	.34 [.33-.34]	26.11 [25.82-26.40]	. .35 [.35-.36]	42.85 [42.40-43.30]	.36 [.36-.37]	83.57 [82.89-84.25]	.37 [.36-.38]
	ULI	8.88 [8.68-9.08]	.35 [.34-.36]	24.46 [24.10-24.82]	.37 [.36-.37]	40.42 [39.89-40.99]	.38 [.37-.38]	79.70 [78.89-80.51]	.38 [.37-.38]
	RIT	9.43 [9.30-9.56]	.34 [.33-.35]	25.97 [25.67-26.27]	.36 [.35-.36]	42.60 [42.16-43.04]	.37 [.36-.37]	83.03 [82.37-83.69]	.37 [.37-.38]
	RIR	7.78 [7.60-7.96]	.39 [.38-.40]	24.33 [24.02-24.64]	.37 [.37-.38].	40.52 [40.05-40.99]	.38 [.37-.39]	81.19 [82.37-83.69]	.38 [.37-.38]
100	SC	6.77 [6.61-6.93]	.41 [.40-.42]	17.99 [17.45-18.53]	.42 [.42-.43]	28.23 [27.26-29.20]	.43 [.43-.43]	55.32 [53.48-57.16]	.43 [.43-.44]
	Poly	9.81 [9.72-9.90]	.31 [.31-.32]	26.21 [25.93-26.49]	.35 [.34-.35]	42.64 [42.25-43.03]	.36 [.35-.36]	82.55 [81.99-83.11]	.37 [.36-.37]
	ULI	9.33 [9.17-9.49]	.33 [.32-.34]	25.31 [25.04-25.58]	.36 [.35-.36]	41.42 [41.00-41.84]	.36 [.36-.37]	80.25 [79.54-80.96]	.37 [.36-.37]
	RIT	9.75 [9.65-9.85]	.32 [.31-.32]	26.09 [25.80-26.38]	.35 [.34-.36]	42.44 [42.05-42.83]	.36 [.35-.36]	82.10 [81.55-82.65]	.37 [.36-.37]
	RIR	7.62 [7.47-7.77]	.39 [.38-.40]	23.66 [23.39-23.93]	.38 [.37-.38]	40.23 [39.80-40.66]	.37 [.37-.38]	80.13 [79.59-80.67]	.37 [.37-.38]
250	SC	6.83 [6.72-6.94]	.40 [.39-.41]	18.47 [18.07-18.87]	.41 [.41-.42]	29.23 [28.52-29.94]	.42 [.41-.42]	56.35 [55.02-57.68]	.42 [.42-.42]
	Poly	9.93 [9.88-9.98]	.30 [.30-.31]	26.17 [25.85-26.49]	.34 [.34-.35]	42.07 [41.69-42.45]	.35 [.35-.36]	80.98 [80.34-81.62]	.36 [.36-.37]
	ULI	9.72 [9.62-9.82]	.31 [.30-.31]	25.23 [24.92-25.54]	.35 [.35-.36]	40.84 [40.39-41.29]	.36 [.35-.36]	79.06 [78.45-79.67]	.37 [.36-.37]
	RIT	9.92 [9.87-9.92]	.30 [.30-.31]	26.00 [25.69-26.31]	.34 [.34-.35]	41.78 [41.38-42.18]	.35 [.35-.36]	80.55 [79.92-81.18]	.36 [.36-.37]
	RIR	7.43 [7.30-7.56]	.38 [.38-.39]	22.86 [22.60-23.12]	.38 [.38-.39]	38.76 [38.42-39.10]	.38 [.37-.38]	77.51 [76.96-78.06]	.37 [.37-.38]
500	SC	6.92 [6.85-6.99]	.40 [.40-.41]	18.97 [18.64-19.30]	.41 [.41-.41]	29.64 [29.10-30.18]	.41 [.41-.42]	57.64 [56.61-58.67]	.41 [.41-.42]
	Poly	9.97 [9.94-10.0]	.30 [.30-.30]	26.46 [26.18-26.74]	.34 [.33-.34]	41.75 [41.40-42.10]	.35 [.35-.35]	80.39 [79.80-80.98]	.36 [.36-.36]
	ULI	9.85 [9.77-9.93]	.30 [.30-.31]	25.64 [25.34-25.94]	.34 [.34-.35]	40.43 [40.03-40.83]	.36 [.36-.36]	78.15 [77.57-78.73]	.37 [.37-.37]
	RIT	9.95 [9.91-9.99]	.30 [.30-.31]	26.34 [26.05-26.63]	.34 [.33-.34]	41.44 [41.07-41.81]	.35 [.35-.36]	80.03 [79.46-80.67]	.36 [.36-.37]
	RIR	7.20 [7.12-7.28]	.30 [.30-.30]	22.57 [22.30-22.84]	.38 [.38-.39]	37.73 [37.43-38.03]	.38 [.38-.38]	76.34 [75.79-76.89]	.38 [.37-.38]
1000	SC	6.96 [6.91-7.01]	.40 [.39-.40]	18.90 [18.60-19.20]	.41 [.41-.41]	29.70 [29.13-30.27]	.41 [.41-.41]	57.85 [57.01-58.69]	.41 [.41-.41]
	Poly	9.99 [9.97-10.0]	.30 [.29-.30]	26.38 [26.09-26.67]	.34 [.33-.34]	41.29 [40.90-41.68]	.35 [.35-.36]	79.20 [78.64-79.76]	.36 [.36-.37]
	ULI	9.86 [9.79-9.93]	.30 [.30-.30]	25.24 [24.91-25.57]	.35 [.34-.35]	40.21 [39.84-40.58]	.36 [.36-.36]	77.02 [76.49-77.55]	.37 [.37-.37]
	RIT	9.99 [9.97-10.0]	.30 [.29-.30]	26.16 [25.86-26.46]	.34 [.33-.34]	41.08 [40.70-41.46]	.35 [.35-.36]	78.62 [78.09-79.15]	.37 [.36-.37]
	RIR	7.11 [7.05-7.17]	.39 [.39-.40]	22.02 [21.83-22.21]	.39 [.38-.39]	37.39 [37.10-37.68]	.38 [.38-38]	75.14 [74.68-75.60]	.38 [.38-.38]

SC: Slope coefficient; Poly: Polyserial correlation; ULI: Upper-Lower index; RIT: Item-total correlation; RIR: Item-rest correlation.

The rate of variance explained by a single factor was 40% on average in the slope coefficient while it was 31% in item-rest correlation. In other words, when the item discrimination index measure was taken as .20 before factor analysis, the slope coefficient evaluated 30% of the item pool as non-discriminatory items, and the explained variance rate was approximately 9%

higher. Considering the fact that 30% of the item pool is generally consciously assigned as items with low discrimination during item production, it was seen that the findings obtained by using the slope coefficient and the .20 criterion were very close to the real situation.

Table 3. *The number of items included in the EFA and the mean ratio of variance explained by a single factor for item discrimination index criterion .35*

		k = 10		k = 30		k = 50		k = 100	
n	Method	k_r [95% CI]	Var [95% CI]	k_r [95% CI]	Var [95% CI]	k_r [95% CI]	Var [95% CI]	k_r [95% CI]	Var [95% CI]
50	SC	5.35 [5.03-5.67]	.45 [.44-46]	12.16 [11.34-12.98]	.46 [.45-.47]	19.79 [18.43-21.15]	.47 [.46-.47]	36.70 [34.27-39.13]	.47 [.46-.47]
	Poly	9.10 [8.95-9.25]	.35 [.34-.36]	24.88 [24.58-25.18]	.37 [.36-.38]	40.29 [39.81-40.77]	.38 [.38-.39]	79.18 [78.55-79.81]	.38 [.38-.39}
	ULI	8.19 [7.96-8.42]	.36 [.35-.38]	22.58 [22.16-23.00]	.38 [.37-.39]	37.06 [36.40-37.72]	.39 [.39-.40]	73.44 [72.39-74.49]	.39 [.39-.40]
	RIT	9.05 [8.89-9.21]	.35 [.34-.36]	24.73 [24.43-25.03]	.37 [.36-.38]	39.94 [39.47-40.41]	.38 [.38-.39]	78.70 [78.08-79.32]	.39 [.38-.39]
	RIR	7.28 [7.08-7.48]	.41 [.40-.42]	23.01 [22.70-23.32]	. .39 [.38-.40]	38.26 [37.80-38.72]	.40 [.39-.40]	77.00 [76.40-77.60]	.39 [.39-.40]
100	SC	5.48 [5.20-5.76]	.43 [.42-.44]	12.73 [12.02-13.44]	.45 [.44-.45]	19.96 [18.84-21.08]	.45 [.44-.45]	37.50 [35.40-39.60]	.45 [.45-.46]
	Poly	9.39 [9.27-9.51]	.33 [.32-.34]	24.41 [24.13-24.69]	.37 [.36-.38]	39.84 [39.45-40.23]	.38 [.37-.38]	77.75 [77.21-78.29]	.38 [.38-.39]
	ULI	8.72 [8.54-8.90]	.34 [.33-.35]	23.09 [22.83-23.35]	.38 [.37-.38]	37.66 [37.17-38.15]	.38 [.38-.39]	72.83 [71.96-73.70]	.39 [.38-.40]
	RIT	9.37 [9.25-9.49]	.33 [.32-.34]	24.22 [23.94-24.50]	.37 [.36-.38]	39.60 [39.22-39.98]	.38 [.37-.38]	77.32 [76.81-77.83]	.38 [.38-.39]
	RIR	7.25 [7.14-7.36]	.40 [.39-.41]	22.35 [22.11-.22.59]	.39 [.39-.40]	37.73 [37.40-38.15]	.39 [.38-.40]	75.32 [74.82-75.82]	.39 [.39-.40]
250	SC	5.76 [5.51-6.01]	.41 [.41-.42]	12.75 [12.08-13.42]	.43 [.43-.44]	19.63 [18.65-20.61]	.43 [.43-.44]	36.30 [34.69-37.91]	.44 [.43-.44]
	Poly	9.59 [9.48-9.70]	.31 [.31-.32]	23.48 [23.20-23.76]	.37 [.37-.38]	38.21 [37.90-38.52]	.38 [.37-.38]	74.41 [73.95-74.87]	.39 [.38-.39]
	ULI	9.10 [8.95-9.25]	.32 [.32-.33]	22.66 [22.37-22.95]	.38 [.37-.38]	36.66 [36.31-37.01]	.38 [.38-.39]	71.90 [71.39-72.41]	.39 [.39-.39]
	RIT	9.52 [9.40-9.64]	.31 [.31-.32]	23.28 [23.01-23.55]	.38 [.37-.38]	38.05 [37.74-38.36]	.38 [.38-.38]	74.14 [73.71-74.57]	.39 [.38-.39]
	RIR	7.11 [7.04-7.18]	.39 [.39-.40]	21.60 [21.44-21.76]	.40 [.39-.40]	36.17 [35.97-36.37]	.39 [.39-.39]	72.65 [72.31-72.99]	.39 [.39-.40]
500	SC	5.88 [5.67-6.09]	.41 [.41-.42]	12.86 [12.29-13.43]	.43 [.42-.43]	19.64 [18.78-20.50]	.43 [.43-.43]	35.77 [34.35-37.19]	.43 [.43-.43]
	Poly	9.79 [9.71-9.87]	.31 [.30-.31]	23.47 [23.19-23.75]	.37 [.37-.37]	37.22 [36.93-37.51]	.38 [.38-.39]	72.73 [72.37-73.09]	.39 [.39-.39]
	ULI	9.21 [9.06-9.36]	.32 [.31-.32]	22.31 [22.04-22.58]	.38 [.38-.38]	36.11 [35.77-36.45]	.39 [.38-.39]	70.47 [70.02-70.92]	.39 [.39-.40]
	RIT	9.76 [9.67-9.85]	.31 [.30-.31]	23.27 [22.98-23.56]	.37 [.37-.38]	36.98 [36.71-37.25]	.39 [.38-.39]	72.53 [72.19-72.87]	.39 [.39-.39]
	RIR	7.02 [6.99-7.05]	.40 [.39-.40]	21.25 [21.16-.21.34]	.40 [.39-.40]	35.58 [35.43-35.73]	.40 [.39-.40]	71.16 [70.93-71.39]	.40 [.39-.40]
1000	SC	5.96 [5.77-6.15]	.41 [.40-.41]	12.98 [12.44-13.52]	.42 [.42-.43]	19.31 [18.59-20.03]	.43 [.42-.43]	35.82 [34.78-36.86]	.43 [.43-.43]
	Poly	9.84 [9.77-9.91]	.30 [.30-.31]	22.80 [22.53-23.07]	.38 [.37-.38]	36.85 [36.59-37.11]	.39 [.38-.39]	71.78 [71.49-72.07]	.39 [.39-.39]
	ULI	9.24 [9.10-9.38]	.31 [.31-.32]	21.81 [21.53-22.09]	.38 [.38-.39]	35.43 [35.17-35.69]	.39 [.39-.39]	69.51 [69.18-69.84]	.40 [.39-.40]
	RIT	9.82 [9.74-9.90]	.30 [.30-.31]	22.65 [22.40-22.90]	.38 [.38-.38]	36.69 [36.44-36.94]	.39 [.38-.39]	71.55 [71.27-71.83]	.39 [.39-.39]
	RIR	7.01 [6.99-7.03]	.40 [.39-.40]	21.11 [21.04-21.18]	.40 [.39-.40]	35.27 [35.16-35.38]	.40 [.39-.40]	70.67 [70.50-70.84]	.40 [.39-.40]

SC: Slope coefficient; Poly: Polyserial correlation; ULI: Upper-Lower index; RIT: Item-total correlation; RIR: Item-rest correlation.

Before the exploratory factor analysis, by accepting the item discrimination criterion as .30, the rates of variance explained by a single factor were calculated for 100 replications as a result of removing the items with discrimination below .30 from the analysis. Table 2 provides the mean number of items included in the analysis, the mean variance explained and the 95% confidence intervals.

Table 2 presents findings similar to the .20 criterion. It was determined that fewer items were included in the analysis when the slope coefficient was used, but the rate of variance explained by a single factor was higher compared to the other methods. On the other hand, it was observed that approximately half of the items were not included in the analysis with the slope coefficient when the .30 criterion was used. For example, an average of 57.85 [57.01-58.69] items were included in the analysis with the slope coefficient in the case where the sample size was 1000 and the item pool size was 100; but when item-rest correlation was used, an average of 75.14 [74.68-75.60] items were included in the analysis. On the other hand, the explained variance rates were found to be .41 and .38, respectively. This shows that although the slope coefficient excluded approximately 20 extra items from the test, it created only a 3% change in the rate of variance that was explained.

Table 3 presents the mean number of items included in the exploratory factor analysis, mean ratio of variance explained and the 95% confidence intervals when the item discrimination index criterion was accepted as .35. The findings in Table 3 were similar to the findings obtained with the .20 and .30 criteria. On the other hand, the slope coefficient was found to eliminate most of the items in the test when the criterion was .35. For example, for a sample size of 1000 and a item bank of 100, the slope coefficient included an average of 35.82 [34.78-36.86] items in the analysis; while an average of 70.67 [70.50-70.84] items were included in the analysis with the item-rest correlation. This difference which amounted to almost half of the number of items had an effect of only 3% on the variance rate which was explained. The .35 criterion and slope coefficient were far from being ideal choices especially when content validity as considered. On the other hand, similar values obtained with the slope coefficient in the .20 criterion were obtained with other item discrimination indices in the .35 criterion.

4. DISCUSSION and CONCLUSION

A novel item discrimination index was proposed in this study as a new item discrimination index for polytomous items, and this coefficient was compared with the polyserial correlation, item-total, item-rest correlation, and upper-lower 27% groups method. This comparison was done in the context of the coefficient, in the context of the number of items included in the exploratory factor analysis when item discrimination criteria were .20, .30 and .35 before the exploratory factor analysis and in the context of percentages of variance explained by a single factor. Accordingly, it was observed that the slope coefficient generated lower values than the other coefficients in all cases. On the other hand, when the item discrimination criterion was accepted as .20 before the exploratory factor analysis, it was found that more variance was explained with fewer items compared to other indexes. However, it was observed that the values obtained with the slope coefficient in the .20 criterion were obtained with the .35 criterion when other coefficients were used. When the item discrimination criterion was accepted as .35, the slope coefficient eliminated a large number of items, which raised content validity concerns.

Although the current study did not focus on comparing other item discrimination indices within themselves, it was also found that the item discrimination indices obtained specifically by the item-total, item-rest, and upper-lower 27% groups method did not differ significantly from each other. This expected result was also demonstrated by Engelhart (1965), who compared the 10 item discrimination indices. A similar study was conducted by Beuchert and Mendoza (1979) as a Monte-Carlo study, which also reported similar item discrimination indices.

The current study proposed a new method to calculate item discrimination index for Likert-type scales. There are also current studies on calculating discrimination with different methods such as ROC curve (Cum, 2021) and fuzzy logic (Vonglao, 2017). Besides, estimation of item discrimination parameter based on Item Response Theory is also possible. However, it is difficult for researchers use these methods widely due to the complexity of the calculation of such methods.

In line with all these results, the slope coefficient is recommended as an additional item discrimination index that can be used in scale development studies. However, in cases where the slope coefficient is used, it is recommended to position the item discrimination index criterion around .20. Considering that this study only used simulative data, it is believed that working with real data sets in future studies will improve the research findings.

Acknowledgments

The preliminary results of this study have been presented in 7th Conference of Measurement and Evaluation in Education and Psychology.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

Authorship Contribution Statement

Umit Celen: Finding the problem statement and slope coefficient, reporting, data analysis, methodology. **Eren Can Aybek:** Methodology, data generation, data analysis, reporting.

Orcid

Umit Celen https://orcid.org/0000-0001-6376-6167
Eren Can Aybek https://orcid.org/0000-0003-3040-2337

REFERENCES

Beuchert, A.K., & Mendoza, J.L. (1979). A Monte Carlo comparison of ten item discrimination indices. *Journal of Educational Measurement, 16*(2), 109-117. http://www.jstor.org/stable/1434454

Crocker, L., & Algina, J. (2008). *Introduction to Classical and Modern Test Theory.* Cengage Learning.

Cum, S. (2021). Examining the discrimination of binary scored test items with ROC analysis. *International Journal of Assessment Tools in Education*, *8*(4), 948-958. https://doi.org/10.21449/ijate.894851

Engelhart, M.D. (1965). A comparison of several item discrimination indices. *Journal of Educational Measurement, 2,* 69-76.https://doi.org/10.1111/j.1745-3984.1965.tb00393.x

Johnson, A.P. (1951). Notes on a suggested index of item validity: The U-L index. *Journal of Educational Psychology, 42*(8), 499-504. https://doi.org/10.1037/h0060855

Kelley, T.L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*(1), 17-24. https://doi.org/10.1037/h0057123

Livingston, S.A., & Dorans, N.J. (2004). *A Graphical Approach to Item Analysis.* ETS Research Report.

Magis, D., & Raiche, G. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. *Journal of Statistical Software, 48*(8), 1-31. https://doi.org/10.18637/jss.v048.i08

Martinkova, P., & Drabinova, A. (2018). ShinyItemAnalysis for teaching psychometrics and to enforce routine analysis of educational tests. *The R Journal, 10*(2), 503-515. https://doi.org/10.32614/RJ-2018-074

Moses, T. (2017). A review of developments and applications in item analysis. In *Advancing Human Assessment* (Eds. R. E. Bennett & M. von Davier). Springer Open.

Oosterhof, A.C. (1976). Similarity of various item discrimination indices. *Journal of Educational Measurement, 13*(2), 145-150. http://www.jstor.org/stable/1434235

Popham, W. J. (2014). *Classroom Assessment: What teachers need to know?* Pearson.

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL: https://www.R-project.org/

Revelle, W. (2022) *psych: Procedures for Personality and Psychological Research*, Northwestern University, Evanston, Illinois, USA
URL: https://CRAN.R-project.org/package=psych

Tabachnick, B.G., & Fidell, L.S. (2013). *Using Multivariate Statistics.* Pearson.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.

Wickham, H., Francois, R., Henry, L., & Müller, K. (2022). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.9
URL: https://CRAN.R-project.org/package=dplyr

Vonglao, P. (2017). Application of fuzzy logic to improve the Likert scale to measure latent variables. *Kasetsart Journal of Social Sciences, 38*(3), 337-344. https://doi.org/10.1016/j.kjss.2017.01.002

APPENDIX

Figure 1A. *Item discrimination indices for item pool size of 10*

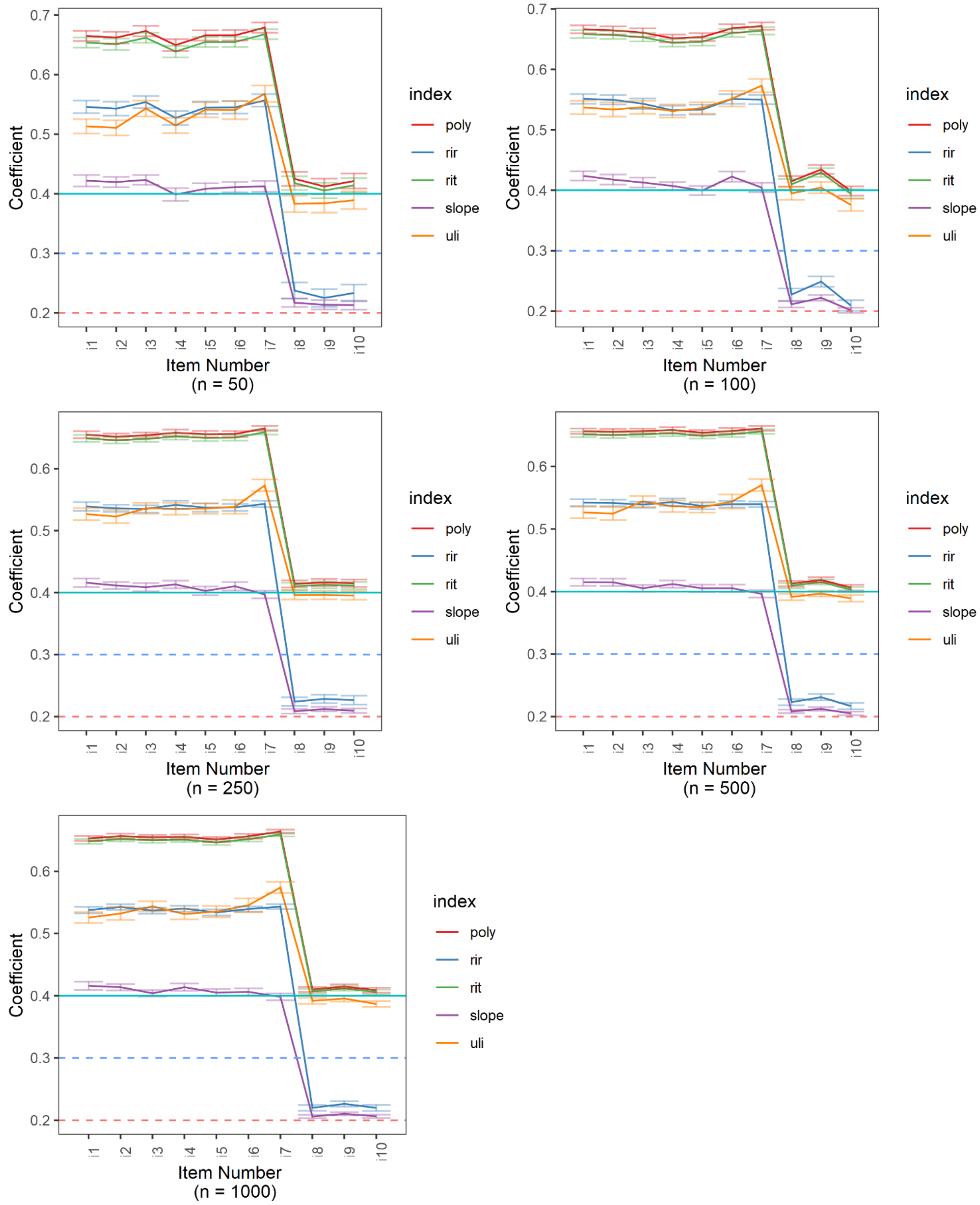


Figure 2A. *Item discrimination indices for item pool size of 30*

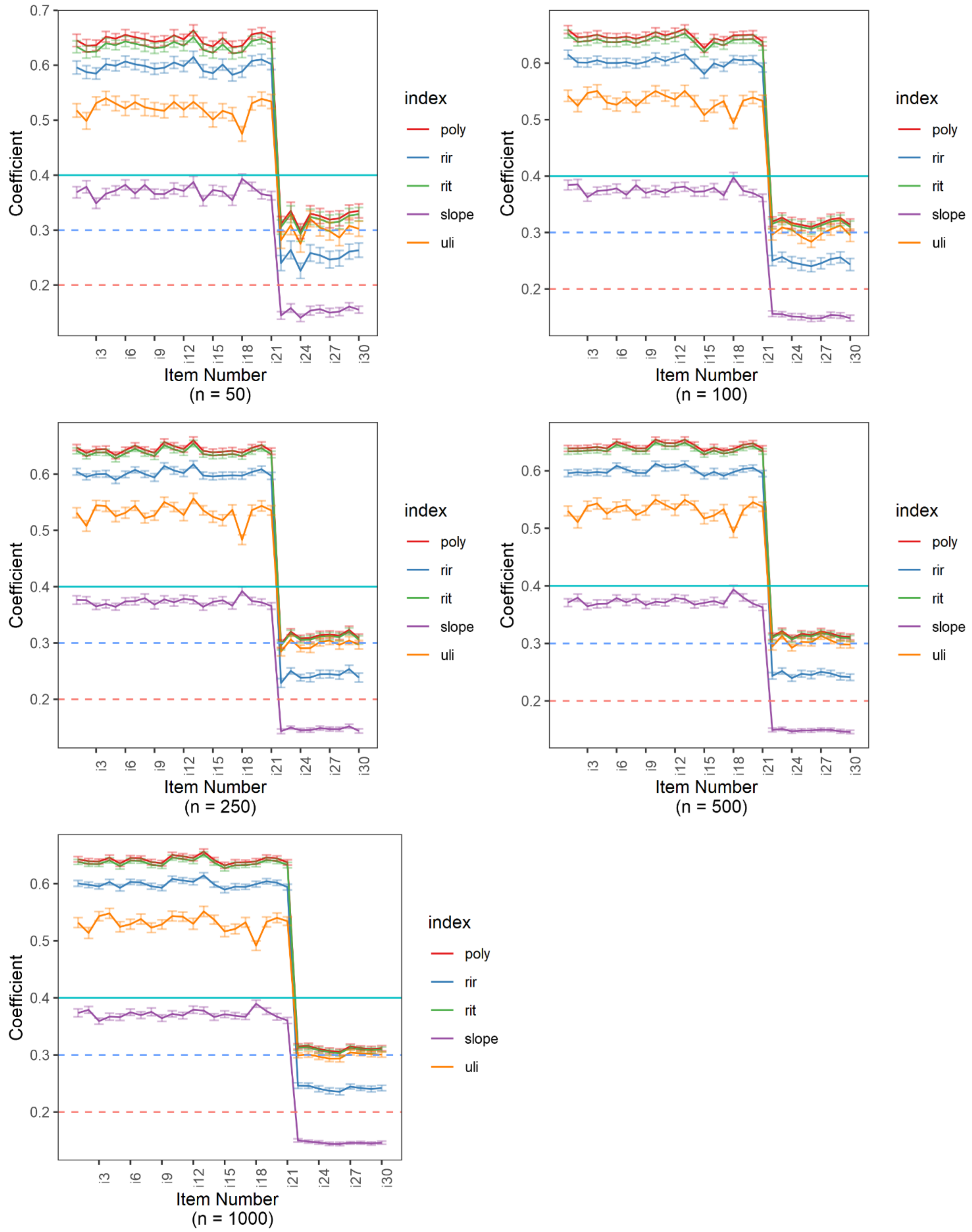


Figure 3A. *Item discrimination indices for item pool size of 50*

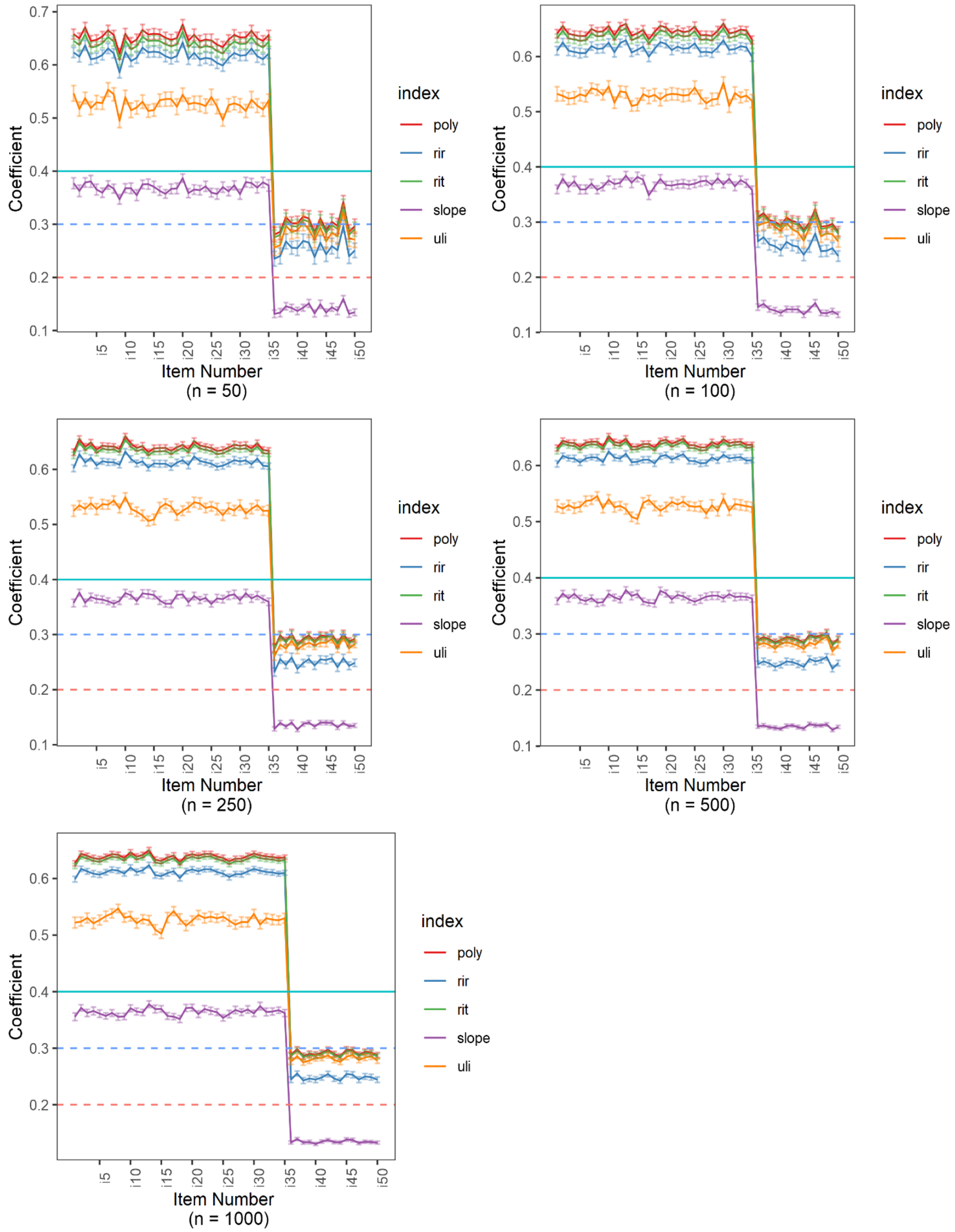


Figure 4A. *Item discrimination indices for item pool size of 100*