This version: May 10, 2013

A nonparametric item characteristic curve using the Gini's Mean Difference

Shlomo Yitzhaki, Rinat Itzhaki and Taina Pudalov

Shlomo Yitzhaki - Corresponding author. Hebrew University and Central Bureau of Statistics, Israel, E-mail: shlomo.yitzhaki@huji.ac.il

Rinat Itzhaki, Henrietta Szold Institute. E-mail: ritzhaki@szold.org.il

Taina Pudalov, Central Bureau of Statistics, Israel. E-mail: tainap@cbs.gov.il

* This report represents the opinion of the authors. It does not necessarily represent the institutes in which the authors are employed in.

Abstract

The Item Characteristic Curve describes the relationship between the probability of correctly answering a question and the ability presumed to be required to answer the question. Since ability is a latent variable, distributional assumptions must be imposed on it in order to estimate such relationship. In this paper we overcome the need to impose an assumption on the distribution of abilities by using the properties of concentration curves and Gini's Mean Difference (GMD). This enables investigation of whether the probability of correctly answering a question is monotonically related to a specific ability. The method is also useful for classifying abilities. This paper details the properties of the technique and provides examples of its application.

Keywords: Item characteristic curve, Gini correlation, Gini Regression, LMA curve

1. Introduction

Item Response Theory (hereafter IRT) was developed between the 1950s and the 1970s. It became mainstream theory in the field of education with Lord and Novick's publication (1968). Since then, a number of extensions and methods of estimation have developed. However, the basic framework remains the same. The theory imposes a structure on data that reflects a certain distribution of abilities among examinees and represents a given structure of the relationship between the probability of success in answering a question and the ability presumed to be required to answer the question. The curve depicting such relationship is referred to as the ICC (Item Characteristic Curve). For excellent reviews of developments of this theory, see Lee, Wollack, and Douglas (2009) and Yang (2007).

The aim of the current paper is to present a new methodology for estimating properties of the ICC. Since ability is a latent variable, it is assumed to follow a given distribution. We circumvent this problem by estimating the response to a question as a function of the cumulative distribution of ability, which is known to follow the uniform distribution.[1] In addition, instead of relying on the individual responses to a question, we use the cumulative responses as a function of the cumulative distribution of ability. This has two effects: (a). Smoothing the curve, so that its properties will reflect those of the ICC. (b). Enabling estimation of the curve from observable events, and thus eliminating the need to assume a specific distribution of ability. In contrast to existing methods that estimate a parameter, or several parameters, as with non-parametric estimates (Bolt, 2001; Sijtsma & Molenaar, 1987; Ramsay, 1995), the proposed method estimates the ICC by a curve.

An additional advantage of the new approach is that it enables decomposition of the regression coefficient of the ICC to reflect the contribution of different levels of ability. This decomposition reveals whether the contribution of each range of ability to the regression coefficient is with the same sign. Large ranges of abilities with negating contributions would indicate that a given question does not actually examine the trait intended. This could be true even in cases in which the overall relationship is positive and statistically significant.[2] Failure to find a monotonic non-negative correlation implies that the ability needed to answer a question is not the ability examined. Such information can also be useful in classifying abilities. Unlike other alternative non-parametric approaches, the proposed method does not require arbitrary window selection for smoothing the

[1] This property is referred to as the probability integral transformation (DeGroot, 1975, pp127-128).
[2] This can be viewed as a non-parametric version of Linacre (1995, 1998).

curve, a procedure that can blur the graph and hide some non-monotonic sections. Finally, no assumption on the distribution of ability is required.

The methodology presented is based on techniques that have been developed and demonstrated in the fields of income distribution and finance (Yitzhaki & Schechtman, 2013). Since the derivation of the method has been documented elsewhere, we target the current presentation to the reader who is interested in applying the methodology to the area of education, and refer the reader interested in the basic tenets or evidence of the methodology to the original papers and the recently published book on the subject (Yitzhaki & Schechtman, 2013).

We emphasize that we are not the first to advocate the use of non-parametric estimation methods. Lord (1970), Holland and Rosenbaum (1986), Ramsay (1991), Douglas (1997), Junker and Sijtsma (2001), Stout (2001), Lee et al. (2009) are among those who suggested and attempted non-parametric approaches.[3] However, the curves estimated by these non-parametric approaches require the quantification of ability and the determination of arbitrary "windows" for averaging data in order to achieve a smooth curve.

The structure of the current paper is as follows: Section 2 describes differences between measurements of abilities and those of such quantitative variables as height. Section 3 presents the curves that will be used in this article. The definitions and the curves are presented in subsection 3.1 while subsection 3.2 connects the curves to regression coefficients, and subsection 3.3 presents the contribution of ranges of abilities to the regression coefficients. Since the curves were developed elsewhere, we have added a brief subsection, 3.4 for the practitioner. The practitioner can skip subsections 3.1, 3.2 and 3.3, and focus on sections 3.4. Section 4 compares the properties of the suggested methodology with TestGraf (Ramsay,2000), the most popular method that is similar to the suggested methodology, while subsection 4.1 illustrates the representation of different types of ICC in the suggested approach. Section 5 discusses estimation and testing. Section 6 illustrates the proposed methodology using data from exams administered at The Henrietta Szold Institute in Jerusalem in the years 2006-2008. Section 7 presents concluding remarks, with suggestions for further research.

2. Characteristics of measurements in the field of education

Ability is a latent variable. Hence, in measuring ability one has to deal with two separate issues: (a). that the ability demonstrated in the exam is based on one-dimensional ability. (b). that the

[3] Xu and Douglas (2006) present a simulation study in which they examined different algorithms. However, it is not clear how their approach compares with the one presented in this paper.

responses to the question that is analyzed are a function of the ability demonstrated in the exam. This distinction does not arise in quantitative variables that are used by regression techniques in estimating ICCs. To see those points we compare measurement of the quantitative variable, height, with that of ability in a specific area. Height can be measured directly, and can include random or non-random measurement error. Also, height is well defined, and no one confuses it with another variable that may be correlated with it, like weight. In contrast, in measuring ability as demonstrated in an exam, we have to be clear that the ability we intend to measure and we actually measure are the same ability.

Measuring one-dimensional ability is like measuring height of examinees who are behind a screen. Imagine that the administrator asks who is taller than 150 centimeters and each examinee truthfully answers yes or no. The administrator then picks another height and the process is repeated. The "height" of a person would then be determined by the number of questions that were answered positively. An individual's "height" and the average "height" of the sample would depend on the "difficulty" of the distribution of questions about height. In the case of height, the "difficulty" can be measured by the distance in centimeters between consecutive questions. But, since there is no independent yardstick for measuring the difficulty of the distribution of questions about ability, difficulty and ability must be assessed by the same scale, the number of examinees answered successfully. Two essential differences between assessments of height and ability are apparent here: (a). While height is well defined, the criteria for specific abilities are not a-priori clear. We refer to this issue as the classification of abilities. (b). While height is a quantitative variable, ability, even one-dimensional ability, does not have a unit of measurement. As a result, ability must be considered an ordinal variable. These distinctions play an important role in the proposed methodology.[4]

Concerning the first issue, we argue that since we are dealing with a latent variable, the only way to distinguish whether we are dealing with one dimensional ability is if the traits that causing the demonstrated ability are not monotonically correlated in the population of the examinees. That is, unlike the case of weight and height that we can distinguish between them, and the contribution of each component can be distinguished by decomposing the variance of the sum, in the case of different abilities we can only distinguish between them if the marginal distributions of the different abilities are not monotonically correlated in the population.

Our approach to the first difference described above (a), regarding whether ability is general or specific, is to investigate whether the correlation between the ability demonstrated in an exam and the

[4] For the implications of this distinction on the robustness of conclusions on average ability see Yitzhaki & Eisenstaedt (2003), Schechtman et al. (2008b), and Schechtman & Yitzhaki (2009).

ability to answer a specific question of the exam is monotonic. If so, then both abilities are indistinguishable, and can be referred to as the same ability. If, on the other hand, the correlation between these abilities is not monotonically increasing, i.e., either the correlation is negative or ranges of the ability demonstrated in the exam have negating signs, then we can conclude that the ability required to answer the specific question is not the same as the ability demonstrated in the exam. This approach resembles that delineated by Linacre (1995, 1998).

To understand the implication of the second difference between ability and height described above (b), i.e., the issue of quantitative vs. ordinal measurement, imagine the difference between the direct measurement of height (i.e., quantitative) and the measurement of the "behind the screen" height (hereafter BTS), the latter corresponding to assessment of ability. Consider an exam composed of twenty questions, in which examinees are asked whether their height is 150, 155, 160,….centimeters. BTS measurements are equivalent to a particular way of binning (Wainer, Gessaroli, & Verdi 2006). That is, all individuals whose heights are between 150 and 155 cm, whether they are 151 or 154 cm tall, will be classified with the lower height, i.e. 150 cm. As such, the continuous variable of ability is converted into a discrete distribution, with increments defined by the difficulty distribution of the questions. However, unlike the example of BTS height, there is no quantitative scale for measuring ability. Thus, the assumption of a specific distribution of ability implies converting an ordinal variable of ability into a quantitative one. The implication of this statement is that non-parametric estimates of ability do not exist. This is because a distributional assumption must be assumed to quantify ability.

Another way of understanding the implication of the difference between the direct and the BTS measurement is to examine the implications of asking about 151 rather than 155 centimeters. It is as if we would change the step function converting ability into scores of the exam. The more questions that are added in a specific height range, the greater the slope of the ICC in that range. Adding a question to the exam, or changing the difficulty distribution can be interpreted as applying a monotonic non-decreasing transformation to the explanatory variable (i.e., ability as demonstrated in the exam) in the ICC. Therefore, the application of a monotonic non-decreasing transformation on the measurement of ability is of interest. This highlights the advantage of the proposed methodology: the relevant question is whether there is a monotonic non-decreasing transformation that can change the sign of the estimated slope of the ICC.

We now turn to the ICC. An exam is defined as a set of questions intended to assess ability based on a given trait. Two matters need to be determined, whether the question analyzed by the ICC assesses the trait tested by the exam, and if so, the level of ability the question refers to. Regarding the former, a question assesses a particular trait if the correct responses to the question form a monotonic non-decreasing function of the ability demonstrated in the exam. A parameter cannot provide an answer to this question because a parameter imposes a structure on the shape of the curve. The methodology proposed in the current paper enables the researcher to identify ranges of ability with negating contributions to the slope of the ICC.

Once we determine that a question does assess a particular trait, we can consider the discrimination property of the question. An ideal question is one in which up to a certain level of ability the proportion of correct responses is zero, and from that level the proportion of correct responses is 100 percent. The more difficult the question, the higher the level of ability at which the "jump" occurs. This issue is not discussed in this paper because it requires the use of concepts that require a totally different kind of analysis. The interested reader is referred to Yitzhaki (2013).

Formally, the basic elements of the IRT model are: Let $S_{jh}(d_j, \theta_h)$ be a random variable representing the success of an examinee h, (h=1,…,H), with ability θ in answering a question of difficulty d_j, (j=1,…,J). It is convenient to think of S_{jh} as a binary variable with 1 representing a correct answer and zero an incorrect one. S_j represents the answers to question j. $\bar{S}_h$ represents the realized ability of examinee h in the exam. Formally, the realized ability of examinees is described by (2.1) while the difficulty of the question by (2.2).

$$\bar{S}_h = (1/J)\sum_{j=1}^{J} S_{jh}(d_j, \theta_h)\,. \tag{2.1}$$

$$\bar{S}_j = (1/H)\sum_{h=1}^{H} S_{jh}(d_j, \theta_h)\,. \tag{2.2}$$

The expected values of the random variables S_h and S_h have the following properties:

$$\frac{\partial E\{\bar{S}_i\}}{\partial d_j} \le 0, \quad \text{for } i = \{h, j\} \text{ and all } j; \tag{2.3}$$

$$\frac{\partial E\{\bar{S}_i\}}{\partial \theta} \ge 0, \quad \text{for } i = \{h, j\}\ , \tag{2.4}$$

where E{} represents the expectation operator. In empirical work, $\bar{S}_h$ substitutes for θ_h .

3. The LMA and NLMA curves.[5]

We present in this section the curves proposed to substitute the ICC in order to find out whether the question analyzed belongs to the ability examined in the exam. [6] We refer to them as

[5] This section is based on Yitzhaki & Schechtman (2012b, 2013).

LMA (Line of independence Minus the Absolute concentration curve) and NLMA (Normalized LMA). The curves are based on extension of the Lorenz curve that has been used for more than a century in the area of income distribution. The Lorenz curve defines a measure of variability, the Gini's Mean Difference, (GMD or simply Gini), which is similar to the variance. The GMD can be extended to define equivalents of the Pearson correlation coefficient and the Ordinary Least Squares regression coefficient. The Gini has decomposition properties that nest the decomposition of the variance as a special case.[7] The use of curves to describe variability and correlation enhances understanding of the contribution of different sections to the overall parameter. Hence, in the discussion that follows, we will be shifting our discussion from curves to parameters and vice versa.

3.1 Definitions and properties of the curve

Our basic assumption is that the probability of success in answering a question that is relevant for a given exam should be a monotonically non-decreasing function of the trait examined by the exam. The meaning of monotonically non-decreasing is that the expected score of examinees in answering the question is a non-decreasing function of ability demonstrated in the exam at every point. A failure of the function to be monotonically non-decreasing indicates that a question does not belong to the trait examined. This is evident by the fact that, if the function is a constant, the probability of success is statistically independent of the trait evaluated by the exam.

Presumably, both responses to a question and an exam are provided, and the task is to discover whether the question is appropriate for the trait examined. To avoid spurious correlation, the scale of ability should not include the examined question.[8] The LMA curve is the difference between two curves. We will first define the curves and then present their properties.

Assume that the observations are ordered according to the ability demonstrated in the exam. Then

(3.1) $\mathrm{LOI(P)} = \mu_j \mathrm{P}$, for $0 \le \mathrm{P} \le 1$,

where $\mu_j = E\{S_j\}$ is the expected value of the success of the examinees in answering the j-th question, and P= F(θ) is a specific value of F(θ), the cumulative distribution of ability as demonstrated in the exam. The LOI, which is an abbreviation for **L**ine **O**f **I**ndependence (represented by OCB in Figure 1a), represents the cumulative success of the P least able examinees, in the case that the probability

[6] Note, however, that we do not deal with the discrimination properties of the question, an issue that requires a different concept, and hence is dealt in Yitzhaki (2013).
[7] For a review of the properties of the Gini and a comparison of its properties to the properties of the variance, see Yitzhaki (2003) or Yitzhaki & Schechtman (2013).
[8] In the case that all questions have equal weight, this correction is not important.

to successfully answer the question would be statistically independent of the ability of the examinees. In empirical work, μ is substituted by $\bar{S}_j$ while F(θ) by F($\bar{S}_h$).

The second curve is the Absolute Concentration Curve (ACC) of the answers:

(3.2) $$\mathrm{ACC}(P) = \int_0^P E(S_j|\theta)dF(\theta), \text{ for } 0 \le P \le 1.$$

The ACC (represented by OAB on Figure 1a) describes the cumulative expected success in answering the question of the P least able examinees.[9]

The LMA curve is defined as the vertical difference between the LOI and the ACC curves:

(3.3) $$\mathrm{LMA}(P) = \mathrm{LOI}(P) - \mathrm{ACC}(P), \text{ for } 0 \le P \le 1.$$

Figures 1a and b illustrate the construction of an LMA curve. In Figure 1a the LOI and the ACC are plotted. The horizontal axis represents the cumulative distribution of θ, and the vertical axis the cumulative probability of correct answers for the case of statistical independence between the probability of answering correctly and ability (LOI), and the actual cumulative probability of a correct answer as a function of P=F(θ). The vertical difference between the two curves is presented in Figure 1b, the LMA curve. Note the inverse of P=F(θ) by $\theta_P = F^{-1}(P)$. θ_P is the inverse of the cumulative distribution of ability.

Figure 1a: The LOI and ACC curves

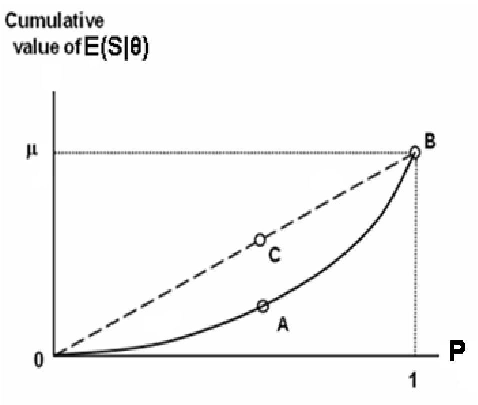
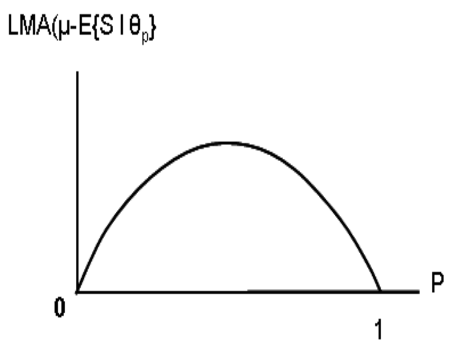


Figure 1b: The LMA curve



The properties of the LMA curve that are used in this paper are the following:

(a) The curve starts at (0,0) and ends at (1,0). It can take any shape depending on properties of $E(S_j|\theta_P)$.[10]

[9] For definitions and properties of Absolute Concentration Curves, see Yitzhaki & Olkin (1991). For the use of the LMA in examining monotonicity, see Yitzhaki & Schechtman (2012).

(b) The derivative (slope) of the curve is equal to $\mu_j - E(S_j|\theta_P)$; that is, the LMA curve has a positive slope if the probability of success in answering a question is lower than the expected probability of success of all examinees, is horizontal if the probability of success is equal to the expected success, and declines if $E(S_j|\theta_P) > \mu$. [11]

(c) Property (b) implies that one can identify the quality of discriminatory power of a question by the peaks in the LMA curve. Provided that $E(S_j|\theta_P)$, i.e., the ICC, is monotonically non-decreasing in relation to ability, then the LMA curve first increases until reaching a peak, and then declines. The steeper the curve around the peak, the higher the discriminatory power of the question.

(d) The LMA curve is concave (convex) if and only if $E(S_j|\theta_P)$ is monotonically increasing (decreasing) with θ.[12]

(e) The LMA curve is not affected by a monotonic increasing transformation of θ, that is, by the scale by which θ is measured, or by the assumption imposed on the distribution of θ.

The properties listed above enable defining a question that belongs to the trait examined, as one whose LMA increases first and then declines, and is concave everywhere. Non-random deviations from this pattern indicate either inappropriateness of the question to the trait examined or some deficiency in phrasing the question.

The LMA curve has several additional properties that enable its being related to the familiar concepts of regression coefficients and correlations. These properties are useful in relating the curve to the IRT model. Before proceeding we will define and present the regression and correlation coefficients.

3.2 The Ordinary Least Squares (OLS) and Gini regression coefficients

The OLS regression coefficient is well known. However, the LMA curve is directly related to a different concept of a regression coefficient – the Gini regression coefficient (Olkin & Yitzhaki, 1992; and Schechtman, Yitzhaki, & Pudalov, 2011). The properties listed below hold only with respect to the Gini regression, while others hold for both OLS and Gini regression coefficients.

Definitions:

The simple OLS regression coefficient is defined as

---

[10] Proof: The value of (3.1) and (3.2) at P=1 is μ.
[11] Proof: The derivative of (3.1) with respect to P is μ and of (3.2) is $E(S_j|\theta_P)$.
[12] Proof: trivial, see footnote 10.

(3.4) $\beta^{OLS} = \dfrac{\mathrm{cov}(S,\theta)}{\mathrm{cov}(\theta,\theta)}$.

The Gini simple regression coefficient (Olkin & Yitzhaki, 1992) is defined as

(3.5) $\beta^{G} = \dfrac{\mathrm{cov}(S,F(\theta))}{\mathrm{cov}(\theta,F(\theta))}$.

The sign of each regression coefficient is determined by the numerator.

The next set of properties enables improving adaptation of the LMA curve to IRT theory, since they demonstrate the relationship between the LMA and regression coefficients.

(f) The area enclosed between the LMA curve and the horizontal axis is equal to cov (S, F(θ)), the numerator in (3.5), see Yitzhaki (2003).

(g) If we redraw the LMA curve with θ instead of P=F(θ), plotted on the horizontal axis, then the area enclosed between the curve and the horizontal axis is equal to cov(S,θ), the numerator of the OLS. (Yitzhaki, 1998).

(h) If the curve is always above (below) the horizontal axis, then the sign of Gini and OLS regression coefficients of S on θ will be positive (negative) for all monotonic non-decreasing transformations of θ. [13]

(i) If the LMA curve intersects the horizontal axis then there exists a monotonic non-decreasing transformation of θ that can change the sign of the OLS regression coefficient (Yitzhaki, 1990).

(j) If in a given section the curve is convex (concave) then the sign of Gini and OLS regression coefficients in that section is negative (positive).[14]

Property (f) enables drawing two additional curves that will improve the content of the curve. By dividing the vertical axis by cov(θ, F(θ)) the area enclosed between the curve and the horizontal axis becomes equal to the Gini regression coefficient. This curve is referred to as the Normalized LMA (NLMA) curve. [15]

[13] See Olkin & Yitzhaki (1992) for additional properties of the Gini regression coefficient, and Schechtman, Yitzhaki & Pudalov (2011) for its properties in a multiple regression framework.

[14] Proof: truncate the distribution of ability and redraw the curve. Then use property (h).

[15] An alternative NLMA curve is derived by dividing the vertical axis by cov(S,F(S)). Then the area enclosed by the curve and the horizontal axis is equal to cov(S, F(θ))/cov(S,F(S)), which is equal to the Gini correlation coefficient. The advantage of using the alternative curve is that it is derived from observables only. However, since we are interested in imitating the ICC, it is not relevant here.

Property (g) relates the curve to the OLS regression coefficient. However, the use of θ instead of F(θ) makes the LOI (3.1) non-linear so that all the properties that are based on the convexity/concavity of the LMA curve are lost. The properties that remain are those based on intersection of the curve with the horizontal axis.

3.3 The components of a regression coefficient

Property (f) of the LMA states that the area enclosed between the LMA curve and the horizontal axis is equal to cov(S, F(θ)). From (3.5) it is evident that normalizing (i.e., dividing) the vertical axis of the LMA by cov(θ, F(θ)) yields a graphical representation of the Gini regression coefficient. We refer to the Normalized curve as NLMA curve. Having described graphically the Gini regression coefficient, we can follow Yitzhaki & Schechtman (2012) to see the contribution of different ranges of ability to the regression coefficient of the ICC.

Yitzhaki (1996) has shown that the OLS and Gini simple regression coefficients are actually a weighted average of the slopes defined between adjacent observations. The weighting scheme depends solely on the distribution of the independent variable, and on the measure of variability used. This result can be extended to decomposition of the OLS and Gini simple regression coefficients, by dividing the range of the independent variable into non-overlapping sections. In this case the weighting scheme is according to the contribution of each section to the intra and inter sections variability. In the case of the OLS, it implies that the weighting scheme can be derived by performing ANOVA on the independent variable. Assuming that the observations are partitioned into M disjoint groups, according to different levels of ability, denoted by m=1,…,M; and $p_m = n_m/n$ is the relative size of group m, then, the Gini and OLS regression coefficients can be decomposed as follows:

(3.6) $$\beta^i = \sum_{m=1}^{M} w^i_m \beta^i_m + w^i_B \beta^i_B ,$$

where i ε{OLS, Gini}. The parameters for the OLS are:

$w_m = p_m \frac{\mathrm{cov}_m(\theta,\theta)}{\mathrm{cov}(\theta,\theta)}$, $\beta_m = \frac{\mathrm{cov}_m(S,\theta)}{\mathrm{cov}_m(\theta,\theta)}$, $w_B = \frac{\mathrm{cov}_B(\bar{\theta}_{m.},\bar{\theta}_{m.})}{\mathrm{cov}(\theta,\theta)}$, $\beta_B = \frac{\mathrm{cov}_B(\bar{S}_m,\bar{\theta}_m)}{\mathrm{cov}_B(\bar{\theta}_m,\bar{\theta}_m)}$ and $\bar{S}_{m.}$ and $\bar{\theta}_{m.}$ denote the vectors of group means. The structure of the decomposition for i=Gini is identical to the one presented in (3.6), except that the terms are the equivalent ones of the Gini regression coefficients, while the weights are the shares of each range of ability in the overall measure of variability of ability, as represented by the Gini.[16] The decomposition (3.6) is based on 4 types of components: the group's weight (w_m), the group's regression coefficient (β_m), the between-group weight (w_B) and the between-group regression coefficient (β_B). We will use this decomposition in the application section.

3.4 A practitioner's guide

Most practitioners are used to the classical IRT figures. In this section we present the relationship between some theoretical Item Characteristic Curves (ICC) patterns, i.e., $E(S|\theta)$, and the representations of those curves in LMA. It is assumed that the examinee scores one point if he or she answers correctly the question, and zero otherwise. No adjustment is needed for the LMA curve in cases in which the scores are not binary.

Several properties of the LMA curve are needed for the analysis:

(a) The curve increases (decreases, remains horizontal) if and only if the probability of success in answering the question is above (below, equal) the average success of answering the question.

(b) The curve is concave (linear, convex) if and only if the probability of answering the question increases (remains constant, decreases) locally.

(c) If the curve is always above (below) the horizontal axis then the correlation between answering the question and success in the exam is positive (negative) for all possible monotonic transformations of ability, i.e., for all possible assumptions imposed on the distribution of ability.

(d) Assuming that the curve intersects the horizontal axis, then two alternatives exist for transformations of ability; one results in a positive OLS regression coefficient and the other in a negative one. In other words, two alternative assumptions can be imposed on the distribution of ability, such that one results in a negative and the other in a positive regression coefficient. Since by changing the difficulty distribution of the questions in the questionnaire is similar to imposing a different distribution of ability, this property implies that one can find an alternative exam that can change the sign of the OLS regression coefficient.

Before we proceed with examples, we will discuss the major differences between the proposed method and alternative methods, Parametric Item Response Theory (PIRT) and TestGraf .

In the proposed method, the basic tool for examining the appropriateness of a question to a given trait is analysis of the shape of an LMA curve. The basic tool in PIRT or its non-parametric

[16] The decomposition of the Gini regression coefficient follows an identical structure except that each term should be replaced by the equivalent term in the Gini framework.

version (NIRT) is a coefficient. There are two reasons for connecting the NLMA curve to the Gini regression coefficient: (a). To enable comparison between the proposed methodology and the classical IRT model. (b). To enable performance of statistical tests on the estimated parameters. The proposed method enables removing some redundant assumptions:

(a) Let $P = F(\theta)$ be the cumulative distribution of θ, i.e. the cumulative distribution of ability. Unlike θ, which is a latent variable, P is unaffected by monotonic increasing transformations of transformed θ, such that S and P are both observable from the sample. Since θ is a latent variable, and can be represented by any monotonic non-decreasing transformation of θ, there is no need to limit its assessment to a particular scale. By definition, the distribution of any cumulative distribution is uniform, between zero and one.[17] Hence, there is no need to impose a specific distribution on the data.

(b). Another redundant assumption is that the discrimination of a question is at the median of the distribution of the response. Since we expect the successful response to be a monotonic non-decreasing function of ability, we define the discriminatory point to be the mean successful responses, because there the probability of success changes from below average to above average. If the distribution of a successful answer happens to be symmetric, then the mean and the median will be identical, but as shown below, this need not be so. However, as we mentioned above, Yitzhaki (2013) suggests an alternative and an improved method for examining the discriminatory power of a question.

(c). The third redundant assumption is the need to estimate a parametric regression. This is a redundant assumption because, by imposing a monotonic relationship on the regression curve, the possibility of checking whether the relationship is actually monotonic is lost. The non-parametric item response theory (NIRT) overcomes this problem by estimating the ICC in selected "windows". The width of the window can affect the possibility of detecting non-monotonicity. The proposed methodology does not impose any structure on the data. However, it results in a regression, so that statistical tests can be performed on the estimators of the parameters. Without the need for window selection, discrimination of a monotonic relationship is left to the discretion of the researcher and the reader.

Another competing method that is also based on examining a curve is TestGraf. The comparison of properties is presented in the next section.

[17] This is known as "the probability integral transformation". See DeGroot , 1975, p-127.

4. The properties of LMA versus TestGraf.

The most popular methodology that is similar to the one suggested in this paper is Ramsay's (1991, 2000) TestGraf. Hence a comparison of the methodologies is called for. The methods are similar in the sense that both are based on a graphical presentation, and on a smoothing procedure. The differences are the following:

(a). The smoothing procedure: TestGraf smoothing procedure is based on the kernel procedure. The suggested method is based on aggregation of responses to a question according to the ability demonstrated in the exam. There are several advantages to the proposed methodology over TestGraf that are caused by the difference in the smoothing procedures and will be listed below.

(a.1). TestGraf requires arbitrary windows selection. If the window is too wide then changing signs in the local regression coefficients can be left undetected. Small windows imply large computer time. The suggested procedure does not require window selection.

(a.2). The suggested method is simpler to use. All is needed is sorting the data according to the results of the exam, and an aggregation of the responses to a question. A program based on EXCEL program can be supplied by the authors upon request.[18]

(b). The suggested methodology enables the user to see the contribution of each section of abilities of examinees to the regression coefficient of the ICC. This property enables a search for redundant questions, i.e., those with little contribution to the ICC regression coefficient. TestGraf does not offer such a connection with a regression coefficient.

(c). The suggested procedure is part of the Gini methodology (Yitzhaki & Schechtman, 2013). The methodology can be used to perform stochastic dominance tests, which can be used to compare groups of examinees, a parameter representing the concept of stratification, which seems to be relevant for representing the discriminatory power of a question so that it is part of a methodology that can be used in the area of education (Yitzhaki, 2013). However, the interpretation and the adjustment to the area is still to be developed.

4.1 Illustrations of the proposed methodology

Following are some examples of the relationship between the ICC and the LMA curve. One possible interpretation of the figures as representing the results of TestGraf (without smoothing) and LMA.

---

[18] The program is written by Yolanda Golan.

Figure 2: The Ideal ICC at the 0.75 discrimination

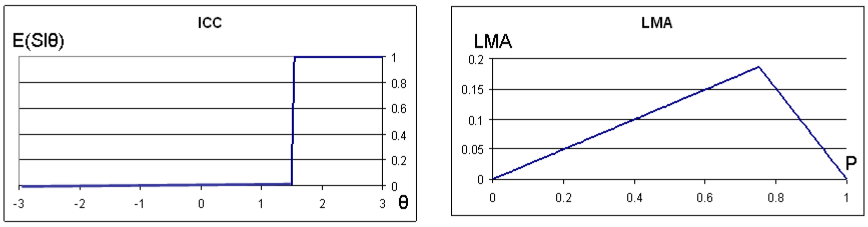


Figure 2 presents the ICC and the derived LMA of an ideal ICC with the following properties: all examinees who belong to the lowest 75% of ability do not answer the question, while the top students are all successful. This figure represents a perfect question with the location of the discrimination at the 75 percentile. As shown, the LMA curve looks like an asymmetric triangle, depending on the point of discrimination that is located at the peak of the curve. Changing the probability of success from zero and one to given constants will only change the angles of the triangle.

Figure 3: A Linear ICC

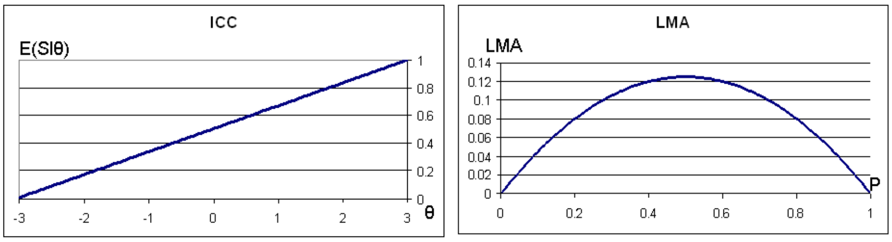


Figure 3 presents a linear ICC. Here, the probability of successfully answering the question is a linear function of ability. In this case the LMA curve looks like a bell curve with the peak at the median. At the left-hand side of the figure the curve increases, indicating that the probability of a correct response is below average. At the right hand side of the figure the curve declines, indicating that the probability of success is above average. The concavity of the curve indicates that the probability of a successful response increases with ability.

The empirical LMA curve oscillates because of random perturbations that may be due to guessing, or simply because the results are affected by random perturbations. The practitioner should be particularly aware of two properties of the LMA curve:

(a). Intersection of a curve with the horizontal axis indicates that the sign of the Pearson correlation for the relationship between answering the question and ability is dependent on the assumption for the distribution of ability. This implies that the question does not examine the ability of the examinees according to the declared purpose of the exam.

(b). Convexity of the curve in a certain region, without intersection of the horizontal axis, indicates that the correlation between answering the question and ability is positive if estimated using the full range of abilities, but negative if the assessment is restricted to the convex part of the curve.

In both cases (a) and (b) the implication is that the question does not assess the trait intended by the exam.

Finally, the curious reader may ask why we propose the use of the LMA curve over the ICC when they rely in fact on the same data and contain the same information. The answer is based on two properties of the LMA curve: the first is that the LMA curve is a cumulative curve, and as such, it tends to reduce the effect of random perturbations.[19] The second is that using the (empirical) cumulative distribution on the horizontal axis enables overcoming the fact that the true distribution of abilities is unknown. Since the distribution of all continuous cumulative distributions between zero and one are known to be uniform, since the empirical cumulative distribution is observed, and since the larger the sample the smaller the deviation from the cumulative distribution of the population, the use of the LMA curve adds local information that improves the methodology.

5. Estimation and testing

Estimation of the parameters involved in the proposed method, as they relate to the Gini methodology were detailed in Schechtman & Yitzhaki (1987), Schechtman et al. (2008a), and Schechtman et al. (2011). The procedures developed in those papers are summarized in Yitzhaki & Schechtman (2013). The estimators are the sample's representation of the population parameters, with the empirical cumulative distribution substituting for the cumulative distribution. The above papers show the estimators to be composed of functions of U-statistics. Thus, for large samples, distributions of the estimators converge to the normal.

The estimation of the LMA curve is also based on the sample's version of the population curve. While we do not yet have means of testing concavity or convexity of the curve, we can use the Gini parameters that are based on truncated segments of the curve to test for concavity and convexity. Regarding intersection of the LMA curve with the horizontal axis, Schechtman et al.'s (2008c) proposed test for intersections of Absolute Concentration Curves can be used to determine whether the LMA curve intersects the horizontal axis. However, in evaluating the power of the test an alternative hypothesis must be assumed. Therefore, our recommendation at this stage is to either rely on large samples and consider only large sections that are based on a large number of observations, so as to be

[19] Lord, (1970) was well aware of this issue. Note the following: "The true-score approach used here leads to excessively irregular curves. Some modification producing smoother curves would be desirable". (p-40).

sure that the findings are significant; or alternatively to rely on testing the parameters of the Gini regression of truncated distribution.

6. An empirical illustration

To illustrate the methodology, we used scores of exams conducted over the years 2006-2008 by the Henrietta Szold Institute in Jerusalem, Israel. Most of the questions were found to be "good" questions, that is, they were found to test the subject matter of the exam. However, since we are interested in examining the performance of the methodology, we present examples of different types of questions, where question type is defined according to the results of the methodology.

We present results of two uses of the methodology:

(a) To determine whether successful responses to the question form a monotonic relationship with ability, as represented by the total scores in the exam.[20] We present one "positive" and one "negative" example.

(b) To determine whether a question is redundant. If the successful response to a question in one exam forms the same monotonic relationship with ability as demonstrated in another exam, then the question is redundant.

Note that the methodology used in (a) and (b) is the same, however, the conclusions with respect to the same LMA figure are opposite.

The examples presented below were selected to demonstrate the points raised in this paper. There was no attempt to present a balanced view of the exams inspected. The questions analyzed were selected from the "test of exceptions"[21], Raven test,[22] and an exam in mathematics.[23]

(a) An example of a "good" question

The question is taken from the "Test of Exceptions". It requires inductive ability and visual perception (the details are in appendix A).

[20] To prevent spurious correlation the results of the exam do not include responses to the question its properties are examined. See, Junker and Sijtsma (2000).

[21] The "Test of Exceptions" requires inductive ability and visual perception. The test includes 20 items displaying series of shapes. The examinee is requested to identify the two shapes that do not fit the series. To do so s/he has to identify the common feature or principle in the three shapes (from the 5 given) and mark the two that do not share this common feature or principle. This test was developed at the Henrietta Szold Institute.

[22] The Raven test is a well known and frequently used test of ability. A matrix of visual forms is displayed to the examinee, who is requested to complete a part of it. (Raven et al., 1995).

[23] The "Mathematical thinking test" was also developed at the Henrietta Szold Institute. It includes questions that require mathematical understanding and logic for solving problems that are not taught in school.

To examine the appropriateness of the question to the content of the exam, we plot the LMA curve of the question. The horizontal axis represents the cumulative distribution of the exam, and the vertical axis the LMA curve of the question. To increase the sample size, we analyzed the questionnaire that was used in the years 2006-2008. During this three year period, 1052 individuals participated in the exam.

Figure 4a presents what we refer to as "the empirical ICC". Each point represents {E{S(θ), θ}, that is the average success in answering the question for a given ability. The points are connected by straight lines. The graph can be misleading because frequencies of the different values of θ are missing and one can change the figure by either assuming a different distribution of θ in the population, or alternatively, by adding similar questions to the exam. The latter is essentially the application of a monotonic transformation to θ. Fluctuations along the curve are apparent.

Figure 4a: The ICC of question 6 in the exception exam

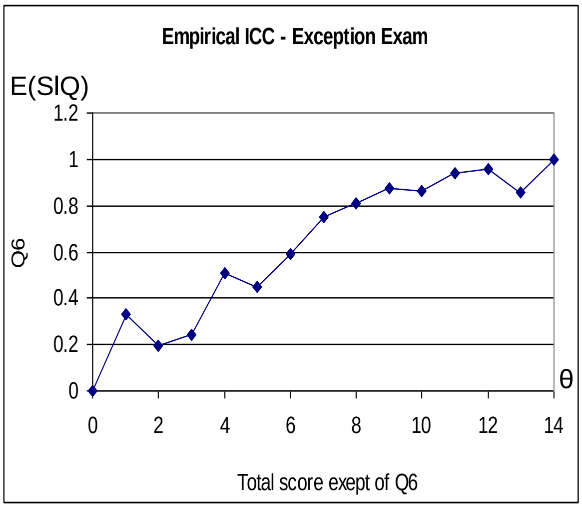
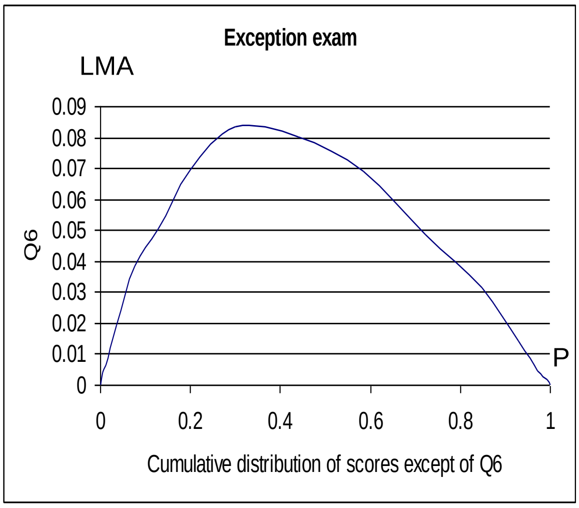


Figure 4b: The LMA curve of question 6 in the exception exam



From figure 4b the following is evident:

(a) The probability of success for 35 percent of the examinees with the lowest ability was lower than the average success in this question. (The curve increases up to this point).

(b) The probability of success in answering the question is a monotonic increasing function of ability (The curve is concave).

(c) The discrimination of the question is at the fourth decile of ability. (The peak is around 35 percent).

The results of running Ordinary Least Squares (OLS) and Gini regressions between response to the question and ability are:

For the OLS: the regression coefficient is 0.072 with standard error[24] of 0.005.

For the Gini: the regression coefficient is 0.071 with a standard error of 0.005.

Conclusion: This example illustrates the properties of a question that fits the trait tested by the exam.

We note that this example is typical of most of the questions that we examined.

(b). An **example of a question that in inappropriate for the subject of the exam**

The example is based on a question that is part of Raven's test conducted in 2006-2008 (the explanation is in appendix B).

Figure 5a represents the "empirical ICC" of the question while figure 5b represents its LMA curve.

Figure 5a: The ICC of question 36 in a Raven's exam

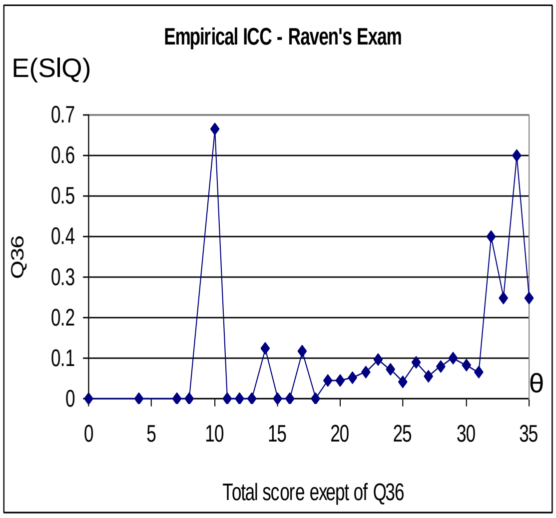
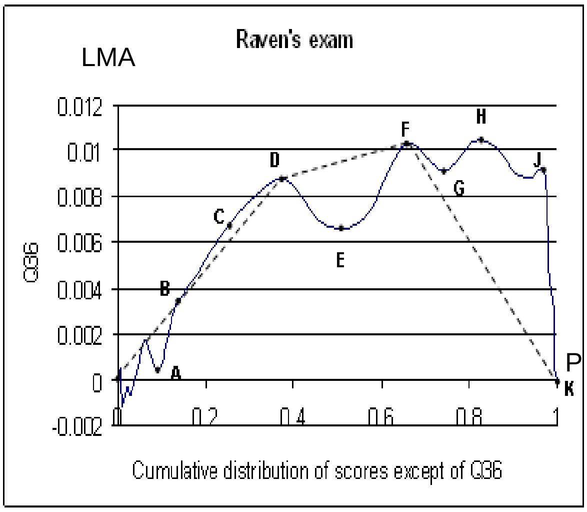


Figure 5b: LMA curve of question 36 in a Raven's exam



The curve in Figure 5b does not increase monotonically with ability as demonstrated in the exam. Small perturbations are ignored because they can be caused by the random elements in responses. However, between the 38th to the 66th percentiles we observe convexity in the curve which is based on about 300 observations. This convexity indicates that the response rate is above the average until the median and then below the average. The apparently smooth change indicates that the ICC is monotonically declining.

[24] The standard errors of both regressions were estimated by the Jackknife approach

We analyzed the relationship with regression coefficients:

The estimate of the overall regression coefficients are: β_{OLS}=0.006, σ_{OLS}=0.002, β_G= 0.005 σ_G=0.002, where σ_{OLS} and σ_G represent the standard errors, respectively. As can be seen, the overall regression coefficients of the Gini and OLS are positive and significant. However, from figure 6b it appears that the overall regression coefficients are composed of negative and positive contributions. To see the contributions of the different ranges of ability, we divided the group of examinees into three sections according to ability: the first group is from zero to the 38th percentile, the second, from the 38th to the 66th, and the remaining composed the third group. The line-segments connecting the points on the curve enable observation of the contribution of each section to the Gini regression coefficient. Table 1 presents the Gini and OLS regression coefficients in each section, and the between-group regression coefficient. As can be seen in the first segment, the curve intersects the line-segment, which means that the sign of the OLS regression coefficient cannot be determined from the figure. On the other hand, the sign of the Gini regression coefficient can be determined by comparing positive and negative areas enclosed between the curve and the line segment. Figure 6b shows that the curve in the lowest section [0,D] intersects the line connecting the extreme points. As can be seen from table 1[25] the OLS intra-group regression coefficient is negative, while the equivalent Gini regression coefficient is positive and negligible. In the second section [D,F] there is no intersection between the line segment and the curve, and the curve is below the line segment, so that both regression coefficients are negative. As expected, the third regression coefficient is positive. The between-group regression coefficient, which is based on the line-segments {0BDFK] is also positive, and since it receives the lion's share of the weight, it determines the sign of the overall regression coefficient.

Table 1: The composition of the OLS and Gini regression coefficients

	OLS		**Gini**	
	β	**weight**	β	**weight**
Section 1	-0.001	0.182	0.000	0.090
Section 2	-0.028	0.009	-0.028	0.013
Section 3	0.025	0.073	0.019	0.054
Between	0.006	0.736	0.006	0.843
Overall	0.006	1.000	0.005	1.000

[25] The decomposition is based on equation (3.6).

(c) Examining whether a question is redundant

To examine redundancy of a question requires at least two exams. Each exam should be designed to evaluate a different type of ability, and the abilities should be statistically independent or negatively correlated. More specifically, while the abilities examined in each exam can be a function of several traits, the traits affecting the abilities in each exam must form two mutually exclusive sets. Using the classification system detailed by Junker and Sijtsma (2001, p-263), the tasks in each exam (a set of "questions", according to our terminology) require different attributes ("abilities", according to our terminology). Further, the attributes required for one exam are not required for success in the other one. Given that the same examinees participated in both exams, the same tool can be used to determine whether the probability of successful responses to a question in one exam monotonically increases with ability as measured in the other exam. The method is identical to the one described in the previous section, only the conclusions are different. If a monotonic relationship is found, then the question is apparently redundant, since the ability that is demonstrated in the other exam covers the same subject matter of the question. Alternatively, if the assumption of mutually exclusive traits in the two exams is violated, then some of the abilities (i.e., sets of traits) required for success in one exam are also required for success in the other exam.

The illustration of the methodology is based on question number 8 in a math exam[26] while ability is measured by the "test of exceptions".

Figure 6a: The ICC of question 8 with respect to ability in another exam

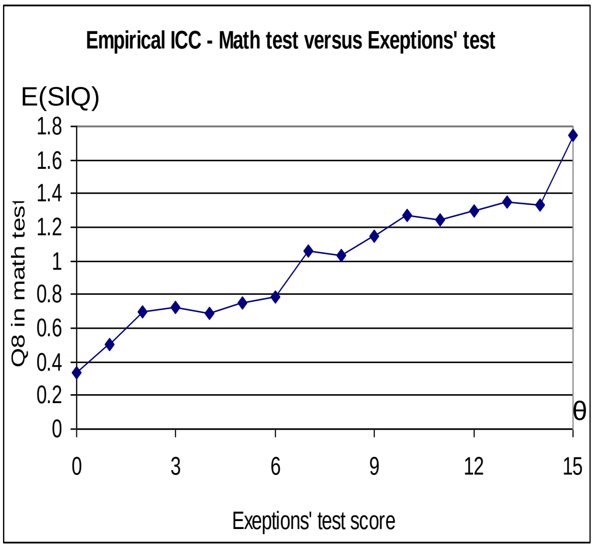
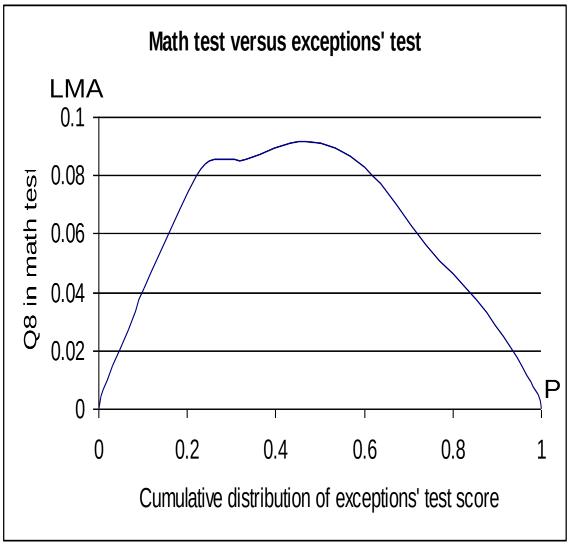


Figure 6b: The LMA curve of question 8 with respect to ability in another exam



[26] The "Mathematical thinking test" was also developed at the Henrietta Szold Institute. It includes questions that require mathematical understanding and logic concluding in math problems that are not taught in school, and are not familiar to the examinees.

Figures 6a and 6b present the "empirical ICC" and the LMA curves. The LMA curve is concave almost everywhere. It thus seems to represent ability as demonstrated in the "test of exceptions". The regression coefficients are positive and significantly different from zero: β_{OLS}=0.074, σ_{OLS}=0.010, β_G= 0.073, σ_G=0.010. The implication is that the question does not test a different aspect of ability from that covered in the test of exceptions.

The last example presents a question and an exam that require almost orthogonal abilities. The question examined is question 5 of the Raven's test and ability is according to the test of exceptions. The sample size is 1052 and the exam was conducted in 2006-2008. Figures 7a and 7b present the "empirical ICC" and the LMA curves.

The LMA curve intersects the horizontal axis twice: around the 65th and 85th percentiles, indicating that the correlation between correctly answering the question is positive for low and medium abilities, and negative for high abilities, and close to zero for extremely high abilities. In addition, the curve is parallel to the horizontal axis between the 33rd and 47th percentiles, which means that the probability of success in this range is equal to average success. The curve changes several times from convex to concave, which implies a non-monotonic relationship with ability. However, this kind of fluctuation may be the result of random perturbations.

Figure 7a: The empirical ICC of question 5 from Raven's test versus ability in the exceptions' test

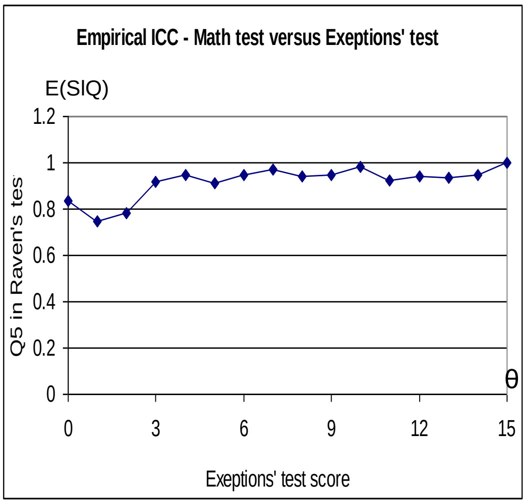
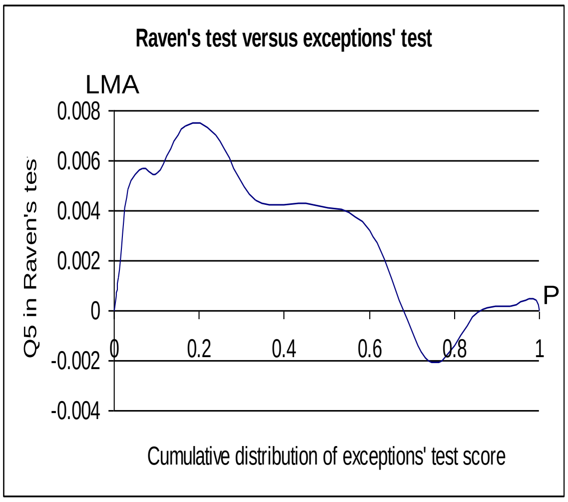


Figure 7b: The LMA curve of question 5 from Raven's test versus ability in the exceptions' test



The Gini and OLS regression coefficients yield the same sign but only the OLS differs significantly from zero: β_{OLS}=0.005, σ_{OLS}=0.002, β_G= 0.004, σ_G=0.003.

The analysis shows that the question is not related to the subject matter of the test of exceptions, thus justifying two different exams that complement each other.

7. Conclusions and further research.

The purpose of this paper was to present and demonstrate a nonparametric representation of the ICC. The methodology is based on the Gini's Mean Difference and the associated concentration curves and parameters. We showed two alternative uses of the methodology: one was to examine whether a question belongs to the subject matter of a given exam, and the other whether a question belongs to the subject matter of a different exam. The advantage of the method is that it reveals more than the parametric methods on the discrimination and congruence of questions to the subject matter of the exam. Further theoretical and empirical research is needed to determine how the methodology can be used to create exams that test one type of ability. For example, we expect a good exam to include only questions that form monotonic relationships with grades in that exam. Once this is achieved, questions can be classified according to different abilities. A question requires a different ability if its responses do not form a monotonic relationship with the scores of other exams. A question, for which its responses form a monotonic relationship with another exam is redundant and should be discarded, unless there are abilities that are needed for success in both exams. In this way utilization of the proposed methodology can contribute to the classification of abilities.

Another direction for further research is the application of other properties of the Gini methodology to the IRT model, in particular, decomposing the success of answering a given question into components of traits and background of the examinees, via the Gini multiple regression (Schechtman et al., 2011). In other words, the Gini IRT model should be extended to multiple regression. A third direction is the development of non-parametric tests for intersections of LMA curves with the horizontal axis.

Acknowledgement: We are grateful to three anonymous referees for helpful comments.

References

Bolt, D. M. (2001). Conditional covariance-based presentation of multidimensional test structure. *Applied Pschological Measurement*, 25, 244-257.

DeGroot, M. H. (1975). *Probability and Statistics*, London: Addison-Wesley Publishing Company.

Douglas, J. (1997). Joint consistency of nonparametric item characteristic curve and ability estimation. *Psychometrika*, 62, 1, 7-28.

Holland, P. W. & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics*, 14, 4, 1523-1543.

Junker, B. W. & Sijtsma K. (2000). Latent and manifest monotonicity in item response models, *Applied Psychological Measurement*, 24, 1, 65-81.

Junker, B. W. & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 3, 258–272.

Lee, Y. S, Wollack, J. A., & Douglas J. (2009). On the use of nonparametric item characteristic curve estimation techniques for checking parametric model fit. *Educational and Psychological Measurement*, 69, 2, 181-197.

Linacre, J. M. (1995). Prioritizing misfit indicators, *Rasch Measurement Transactions*, 9, 422-423.

Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of outcome measurement*, http://www.winsteps.com/a/Linacre-multidimensionality-residuals.pdf

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*, Readind, Massachusetts: Adison-Wesley.

Lord, F. M. (1970). Item characteristics curves estimated without knowledge of their mathematical form – a confrontation of Birnbaum's logistic model. *Psychometrika*, 35, 43-50.

Olkin, I. & S. Yitzhaki (1992). Gini Regression Analysis. *International Statistical Review*, 60, 2, 185-196.

Ramsay, J. O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika*, 56, 4, 611-630.

Ramsay, J. O. (1995). A similarity-based smoothing approach to nondimensional item analysis. *Psychometrika*, 60, 323-339.

Ramsay, J. O. (2000). TestGraf A Program for the Graphical Analysis of Multiple Choice Test and Questionnaire Data. http:// www.psych.mcgill.ca/misc/fda/downloads/testgraf/TestGraf.

Raven, J., Court J., & Raven J.C. (1995). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford: Oxford Psychologists Press.

Schechtman, E. & Yitzhaki, S. (1987). A measure of association based on Gini's Mean Difference. *Communications in Statistics*: *Theory and Methods*, A16, 1, 207-231.

Schechtman, E., Yitzhaki, S., & Artsev, Y. (2008a). Who does not respond in the household expenditure survey: An exercise in extended Gini regressions. *Journal of Business & Economic Statistics*, 26, 3, 329-344.

Schechtman, E., Sofer, E., Yitzhaki, S. (2008b). The robustness of conclusions based on TIMSS mean grades, Israeli Central Bureau of Statistics working paper series no. 37, http://ssrn.com.

Schechtman, E., Shelef, A., Yitzhaki, S., Zitikis, R. (2008c). Testing hypothesis about absolute concentration curve and marginal conditional stochastic dominance, *Econometric Theory*, 24, 4, 1044-1062.

Schechtman, E., & Yitzhaki, S. (2009). Ranking groups' abilities – Is it always reliable? *International Journal of Testing*, 9, 3, 195-214.

Schechtman, E., Yitzhaki, S., & Pudalov, T. (2011). Gini's multiple regressions: two approaches and their interaction. *Metron*, LXIX, 1, 67-99.

Sijtsma, K. & Molenaar, I. W. (1987). Reliability of test scores in nonparametric item response theory. *Psychmetrika*, 52, 79-97.

Stout, W. (2001) Nonparametric item response theory: A maturing and applicable measurement modeling approach. *Applied Psychological Measurement*, 25,3, 300-306.

Wainer, H., Gessaroli, M., & Verdi, M. (2006). Finding what is not there through the unfortunate binning of results: The Mendel effect. *Chance*, 19, 1, 49-52.

Xu, X. & Douglas J. (2006). Computerized adaptive testing under nonparametric IRT models. *Psychometrika*, 71, 1, 121-137.

Yang, S. (2007). *A comparison of unidimensional and multidimensional Rasch models using parameter estimates and fit indices when assumption of unidimensionality are violated*. Ph. D. dissertation, The Ohio State University.

Yitzhaki, S. (1990). On the sensitivity of a regression coefficient to monotonic transformations. *Econometric Theory*, 6, 2, 165-169.

Yitzhaki, S. & Olkin I. (1991). Concentration indices and concentration curves, in Karl Mosler and Marco Scarsini (eds.) *Stochastic Orders and Decisions under Risk*, Institute of Mathematical Statistics: Lecture-Notes Monograph Series, 19, 380-392.

Yitzhaki, S. (1998). More than a dozen alternative ways of spelling Gini. *Research on Economic Inequality*, 8, 13-30.

Yitzhaki, S. (2003). Gini's mean difference: A superior measure of variability for non-normal distributions, *Metron*, LXI, 2, 285-316.

Yitzhaki, S. (2013). The efficient exam. Preliminary draft. http://ssrn.com

Yitzhaki, S. & Eisenstaedt, M. (2003). Ranking groups' versus individuals' ranking. In Amiel, Yoram and John A. Bishop (eds.) *Fiscal Policy, Inequality, and Welfare, Research on Economic Inequality*, 10, Amsterdam: JAI, 101-123.

Yitzhaki, S. & Schechtman, E. (2012). Identifying monotonic and non-monotonic relationships, *Economics Letters*, 116, 23–25.

Yitzhaki, S. & Schechtman, E. (2013). *The Gini Methodology: a primer on a statistical methodology*, Springer Series in Statistics, NY: Springer.

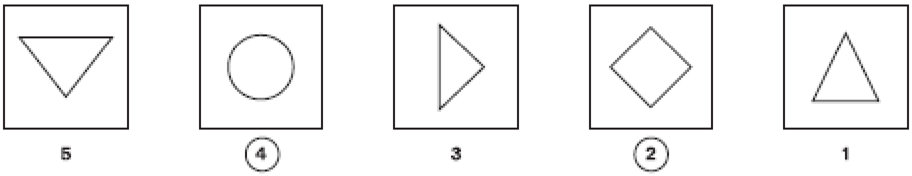
Appendix A:

The "Test of Exceptions" requires inductive ability and visual perception. The 15 items of the test display series of shapes. The examinee is requested to identify the two shapes that do not fit in the series. To do so s/he has to identify the common feature or principle in the three shapes (from the 5 given) and mark the two that do not share this common feature or principle. This test was developed at the Henrietta Szold Institute.

The explanation given to the examinee is:

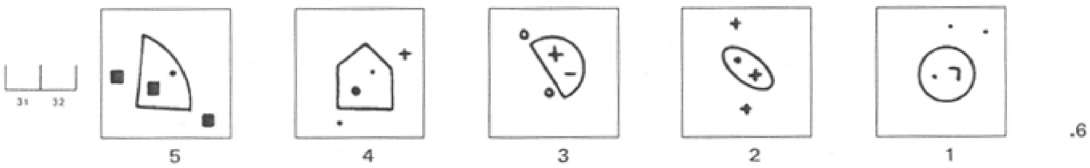
"In this test you will see rows of shapes/ figures. In every row, three *shapes* share certain features or principles, and two do not.

<u>Example A:</u>



In example A three shapes are triangles. We circled the numbers (2 and 4) below the two shapes that are not triangles."

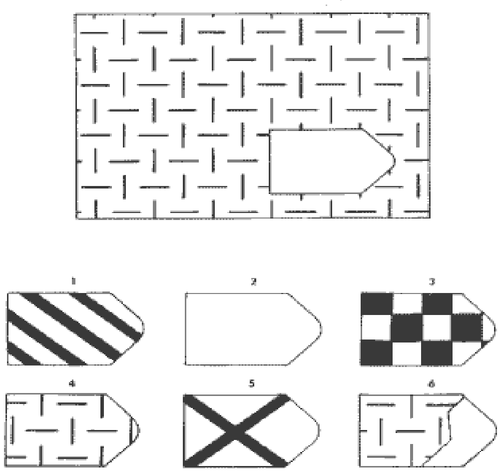
Below are the shapes for question 6 of the exam conducted in the years 2006-2008.



Appendix B:

Below is an illustration of an example of an exam intended to measure the ability required for detecting relationships among abstract objects.

The example is: the examinee is asked to select the piece that best fits the missing part of the carpet:



Question 7 of Raven's test is as follows: The examinee is asked to select the piece that best fits the missing part of the carpet: