JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 8, No. 3, July 2024, pp. 879-895

Equating by Using Circle Equation Approach: Applied Mathematics Formula for Prevent Discrimination

Deni Iriyadi[1], Ahmad Rustam[2], Hevriana Hartati[3]
[1]Program of Postgraduate, Universitas Islam Negeri Sultan Maulana Hasanuddin Banten, Indonesia
[2]Department of Elementary school teacher education, Universitas Sulawesi Tenggara, Indonesia
[3]Department of Mathematics, Universitas Islam Negeri Sultan Maulana Hasanuddin Banten, Indonesia
deni.iriyadi@uinbanten.ac.id

Article History:
Received : 25-02-2024
Revised : 04-06-2024
Accepted : 20-06-2024
Online : 17-07-2024

Keywords:
Circle Equation;
Data Distribution;
Equating;
Simplified Circle Arc;
Prevent Discrimination.

ABSTRACT

This study aims to determine the accuracy of the equating method that uses a circle equation approach in terms of its circular arc (Simplified Circle Arc). This research uses 2015 National Examination data from two questions packages. Using the number of preliminary samples as many as 2135 on the X and 2271 test devices on the Y test. After doing a Rasch analysis using a Mean Square Outfit (MNSQ), the data was acquired and analyzed. Following this, replication was performed up to a maximum of 50 times for each kind of data distribution. For each replication, up to a maximum of 50 respondents were selected from the original data set to be used as data for score equalization. The Root Mean Square Error (RMSE) statistic is then used to analyze the outcomes of the equating score. The results showed that the average RMSE group that has the same distribution will provide a lower RMSE value compared to groups that have different data distributions. The low average RMSE value indicates the accuracy of the equal of the scores performed. Thus, the use of the SCA method is highly recommended to equalize scores, especially in small samples in classes at school to prevent discrimination in grading.

Crossref
https://doi.org/10.31764/jtam.v8i3.22195

——————————— ◆ ———————————

A. INTRODUCTION

One of them is conducting an assessment. This is one of the main components in order to improve the quality of education. Efforts to improve the quality of education can be seen from the improvement of the quality of learning and the quality of assessment (Masino & Niño-Zarazúa, 2016). Assessment is a process carried out in order to monitor the process and progress of learning from students as evaluation material for further improvement of learning. The results of the assessment are presented in the form of numbers or letters as a sign to determine the extent to which students have mastered a subject matter. The results of the assessment are used as a benchmark to see the quality of student education in an educational unit (Antara & Bastari, 2015). The compilation of these items is always based on the grid. Both teachers and the government do this. It is often found that in one school there are parallel classes taught by two or more teachers of the same subject. Each teacher has different teaching characteristics, but in giving tests to students, teachers only base it on the existing grid. This will produce different test devices.

Despite utilizing the same grid when creating questions for the National Examination, the government and teachers rarely use test preparation materials that are actually comparable. A fully similar test is difficult to set up. Since no two test devices are made exactly alike, their results cannot be directly compared (Yu & Osborn-Popp, 2019). It serves as the primary resource for organizing each question in various schools and locations based on the preexisting grid. Test kits made based on the same grid are rare and almost never produce test kits that are truly equivalent (Lissitz & Huynh, 2019). With the different test tools used, automatically the output generated by students who work on the test device is not on the same scale. From these cases, a question arises about a method that can be used to compare the results of a test. From the second section, it is necessary to make a single conversion table from one test device to another. To support this, a statistical method is needed to convert the values of the two test devices. Thus, the proportion of the two test devices can be compared. Measuring instruments, in this case, the questions used to assess the academic abilities of children in urban areas are certainly not fair if they are also applied. There is no way that the 70th set of A test kits and the 70th set in the B test device are the same. This is a result of the different scales that the secrets of the two gadgets have (Souza et al., 2017). So that the scores of the two devices may be compared, a procedure is required. In statistics, the procedure of equating the score is known as equating. Equating is a statistical technique that may be used to translate results from several tests using the same constants (Battauz, 2017; Yu & Osborn-Popp, 2019). This procedure is used to establish the connection between two or more tests (Dimitrov & Atanasov, 2021). Equating is a technique that may be used to equalize test results using statistical and psychometric techniques so that the two test kit scores can be compared (Kolen & Brennan, 2014). To trade or compare equal test energies, equating is utilized to create a scale that is comparable (Haris & Kolen, 2016). From the opinion above, it can be concluded that equating is a statistical procedure that is used to make a transformation from a set of tests to other test devices that measure the same construct.

It's possible that the teaching and learning process will go more smoothly for the teachers when there are fewer students in the classroom (Sumarni, 2015). In this context, instructors, in their function as the implementers of learning in the classroom, obviously require some form of assessment in order to evaluate the successes of their students and determine whether or not they have met the objectives they have set for themselves. As was said before, there may be more than one teacher at one class level at any one time who teaches the same material. When putting together test kits, they will only utilize the grid that has been discussed and agreed upon by all parties involved. When the grades from classes A and B, which are being taught by separate teachers, are compared, it is clear that this is not being done in a fair manner. For this reason, it's important to use a strategy of equalization that's been deemed suitable for the circumstances, especially in regards to the amount of pupils (Himelfarb, 2019). This is because you must use a strategy that is acceptable for using in light of the peculiarities of the class. This is due to the fact that it is essential to use an equalization strategy that is suitable for application in light of the attributes associated with the class. The necessity for the number of students in a classroom to be indicative of a small sample is one of the considerations that goes into the practice of using small samples.

In point of fact, the technique for evaluating the smallest unit, which takes place at the school level, may be discovered. The Indonesian government sets minimum and maximum restrictions on the number of students that are permitted to study together in a single classroom for children who are enrolled in elementary school all the way up to high school. In no one session will there be more than fifty pupils present. There is a cap of 36 students per class, and that's at some of the best universities in the nation. Article 24 of the Republic of Indonesia Minister of Education and Culture Regulation Number 17 of 2017 regarding New Student Admission forms the basis for this determination, which caps class sizes at 36 students. The number of students that can be enrolled in a single class is limited to no more than 36 (Kemendikbud, 2017).

Equating data has been done using a wide variety of different approaches. Some of them are based on the classical technique, which, as everyone knows, is the way that is easier to put into practice in real life. There have been a lot of different approaches to equalization that have been created. The requirements of the contemporary educational landscape inspired the development of these approaches. Some are grounded in traditional theories, while others are grounded on more contemporary ideas. Every one of them comes with a unique set of benefits. Classical techniques of equalization are easier to understand, more logical, and more straightforward to put into practice (Arikan & Gelbal, 2018). The method of scoring by equating points using a circle equation is one that may readily be adapted to grade-level instruction.

There have been a lot of different approaches to equating evolved throughout time. The requirements of the contemporary educational landscape inspired the development of these approaches. Some are grounded in traditional theories, while others are grounded on more contemporary ideas. Every one of them comes with a unique set of benefits. Equating using more traditional ways is not only more comfortable, but also more logical and straightforward (Moses, 2022; Pommerich, 2016). Several industry professionals have provided a variety of equalization strategies that are based on traditional strategies. Ozdemir (2017) contrasts the Equipercentil approach with Circle Arc, Aşiret & Sünbül (2016), which compares the techniques of Idenetity, Mean, Linear, Circle Arc, and Presmooted, and so on and so forth. Caglak (2016) provide a comparison between the Linear approach and the Circle Arc methodology. A comparison is made by Babcock and Hodge (2020) between the Chained technique, the Linear method, the Circle Arc method, the Identity method, and the Synthetic method. The Mean Weight Nominal approach is also evaluated. An updated comparison may be carried out on the basis of these methods with the intention of offering the most suited alternative choice for the use of an effective equating approach. In the study that Ozdemir (2017) conducted, the two researchers evaluated the Simetryc method and the Simplified Circle Arc method against a variety of other approaches. However, they did not investigate the degree of precision provided by any of the Circle Arc approaches.

Although both the Circle Arc technique and the Equipercentil approach fall under the category of nonlinear methods that are based on the classical method, Ozdemir (2017) asserts that the Circle Arc method produces better results. This result is possible because to the estimated root mean square error (RMSE). Livingston and Kim modified the prevalent Circle Arc approach, dividing it into two variants: one based on linearity and the other on nonlinearity, but with some linear components. The modifications that Livingston and Kim made to the Circle

Arc method that was already being used can be found here. The Circle Arc technique that was previously in use prior to the improvements that Livingston and Kim made to it may be found here in its modified form (Caglak, 2016; Diao & Keller, 2020). In addition, Some reseacher demonstrated in a separate piece of research that this method generates accurate results based on RMSD values and biases by applying it to a number of different situations, one of which was the number of samples. He did this by applying it to a number of different situations, one of which was the number of samples. He accomplished this by applying the methodology to the research that he carried out (Albano, 2015; Babcock & Hodge, 2020). According to the findings of research that was carried out by Așiret and Sünbül (2016), the Circle Arc approach results in a reduced equalization error when compared to other methods, even when applying them to small sample sizes. This was found to be the case even when applying the other methods to larger sample sizes.

Adjust the score so that it corresponds to the Simplified Circle. The academic level of the class is a straightforward target for use of Arc. Because of this, the teacher, in his or her role as the person responsible fo`r carrying out the assessment in the school, is in a position to compare the worth of students without inadvertently causing discrimination by employing the equating method that is appropriate to the characteristics of the class, in particular when determining whether or not students have successfully graduated or completed a lesson. By analysing the distribution of scores, the purpose of this research is to come up with a decent strategy that can be used for the purpose of equalizing scores on tiny samples. The purpose of this research is to identify a reliable approach that may be used for the purpose of standardizing ratings on limited sample sizes. In addition, the method of sectional distribution is the central topic of investigation at this point. Normal distribution, positive skewness distribution, and negative skewness distribution are all examples of the types of distributions that cannot be isolated from the students' acquisition of responses to the items that are provided. In the course of this research, a comparison of the various modes of dissemination was carried out. Because of this, the teacher, in his or her role as the person responsible for carrying out the assessment in the school, is in a position to compare the worth of students without inadvertently causing discrimination by employing the equating method that is appropriate to the characteristics of the class, in particular when determining whether or not students have successfully graduated or completed a lesson.

B. METHODS

This study is quantitative and uses expo facto research methods. The responder does not undergo any kind of therapy at any point. In this research, there were two different sample groups. Both of the groups were randomly assigned positions on the same grid but received different kinds of test kits. For the purposes of this research, data from two different National Exam packages on mathematical topics were obtained from the Center for Educational Assessment for the regions of DKI Jakarta and Tangerang. The two potential testing sites for the National Examination were chosen because their respective exams shared features with respect to a number of questions (anchor items), as required by the study design that had been established in advance. These characteristics formed the basis for the selection of the two locations as candidates for the National Examination. The research location is not the main focus of this study, considering that the researcher only focuses on the use of two types of tests. Thus, the research location will follow the location where the test is used and the specific information is only owned by the Center for Educational Assessment. The selection of these locations (DKI Jakarta and Tangerang) is because both places use two different forms of questions.The analysis of the Rasch model is used to determine the degree to which the responses are compatible (person fits) with the model according to the accepted requirements for Outfit Mean Square (MNSQ) value, which are 0.5 and 1.5, respectively (Falani et al., 2022; Iriyadi et al., 2024). This range is established by the parameters that are considered acceptable for the MNSQ value. The total number of student replies acquired for the X test device was 2135, whereas the number of responses received for the Y test device was 2271. According to the findings of the Rasch model, 233 respondents did not fulfill the requirements for the Y test device, and 2 respondents did not meet the requirements for the X test apparatus. Accordingly, there were a total of 2133 student replies in the X testing device but only 2048 in the Y testing set.

In this investigation, the average RMSE value of the equalization outcomes obtained by using the preset way of equating scores serves as the study's dependent variable. For the independent variable, namely the results of equalizing the scores of each method used. Regarding the factors that may be considered independent, they include the equalization technique and the kind of data dissemination (normal, positive skewness, and negative skewness). Using the SPSS program to assist in the process, random samples will be drawn from each population using a method called random sampling with replacement. The randomization process was repeated fifty times, with each iteration using a separate set of fifty respondents. It has been established that the RMSE has been equalized based on the findings of the equalization score. Therefore, according to the calculation below, each group will have a total of fifty RMSE formula (Babcock & Hodge, 2020; Karunasingha, 2022):

$$RMSE(x) = \sqrt{\frac{\sum_{j=1}^{N}\left(\hat{x}_j - x_j\right)^2}{N}}$$

where N is the number of respondents, a number of equalization results, and a balanced score. According to Aşiret and Sünbül (2016) and Uysal and Kilmen (2016), RMSE is the metric that is used to evaluate how accurate the equalization approach is. The excellent accuracy of an equalization technique is shown by its mean having a modest RMSE (Dorans & Puhan, 2017; Liemohn et al., 2021). The research conducted by both of them used RMSE as a benchmark in determining the accuracy of the results of a measurement. In addition, according to Epskamp et al. (2018), smaller mean values indicate a higher degree of equality.

C. RESULT AND DISCUSSION

1. Simplified Circle Arc Method

Circles are a topic that exists in mathematics that is studied starting from the elementary school level to the level of the university. There is nothing different from the material given including the general form of the equation of the circle with the center (0, 0) until the circle equation is centered in (A, B). Likewise, with the equation of the circle radius (r) which from the beginning is taught to have the same form of an equation. Maybe in some literature using different notations but with the same intent. Many methods of equating are developed based on classical theory. Among the methods available, one of them is the circle arc method (Aşiret & Sünbül, 2016; Caglak, 2016; Diao & Keller, 2020; Livingston, 2014; Ozdemir, 2017). There're 2 kinds circle arc method namely symmetric circle arc and simplified circle arc. Both types of methods are based on a linear basis although one of them also has non-linear elements in it (O'Neill et al., 2020; Peabody, 2020).

Equating with the Simplified Circle Arc method uses circle equation approach. The important thing in this method is to determine the two endpoints and midpoints estimated from the data, then limit the approximate equalization curve to pass through these points. The maximum and minimum possible scores on the test are set as endpoints, while the midpoints are determined by the average score of the test scores that will be equalized (Aşiret & Sünbül, 2016; Caglak, 2016; Ozdemir, 2017). Therefore, the *Simplified Circle Arc* method function is identical to the mathematical formula taken from a circle that passes through three predetermined points. The top point of the curve is determined by the maximum score of each test while the lower point is determined by the minimum score of each test. There are 3 coordinate points namely (X_1,Y_1), (X_2,Y_2), and (X_3,Y_3). The middle value (X_2,Y_2) is determined by the average of the two tests. If the equating design is single-group, counterbalanced, or equivalent-groups design, the midpoint (X_2,Y_2) can be determined by equalizing the mean of the new form directly to the mean of the reference form. In the equalization design that uses anchor tests, the midpoint can be determined through chained linear equating (Caglak, 2016).

In the Simplified Circle Arc, the function for estimating the results of equal equations is divided into 2 components, namely linear and curvilinear according to Kim & Livingston (Caglak, 2016; Diao & Keller, 2020). The linear component is only a line that connects two extreme points calculated using the equation (1):

$$L(x) = Y_1 + \frac{Y_3 - Y_1}{X_3 - X_1}(X - X_1) \tag{1}$$

Figure 1 shows the linear component of L(x), which is a straight line connecting the minimum and maximum points. Furthermore, the three points are transformed by subtracting L(x), resulting in the curvilinear component as shown in Figure 2. Because the maximum point and minimum point are in the L(x) line, this step transforms Y_1 to $Y_1^T = 0$ and Y_3 to $Y_3^T = 0$ and transform Y_2 to Y_2^T.

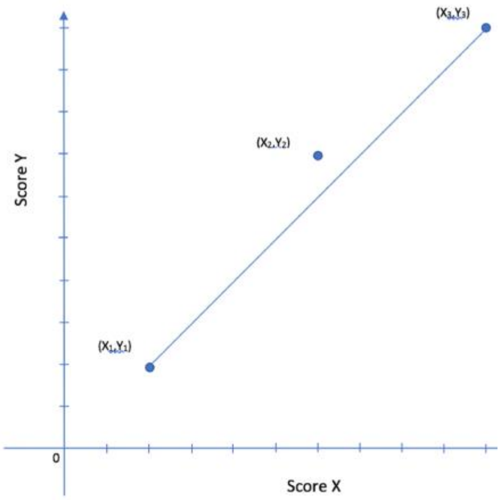

Figure 1. Third Position Point (Maximum, Middle Point, and Minimum)

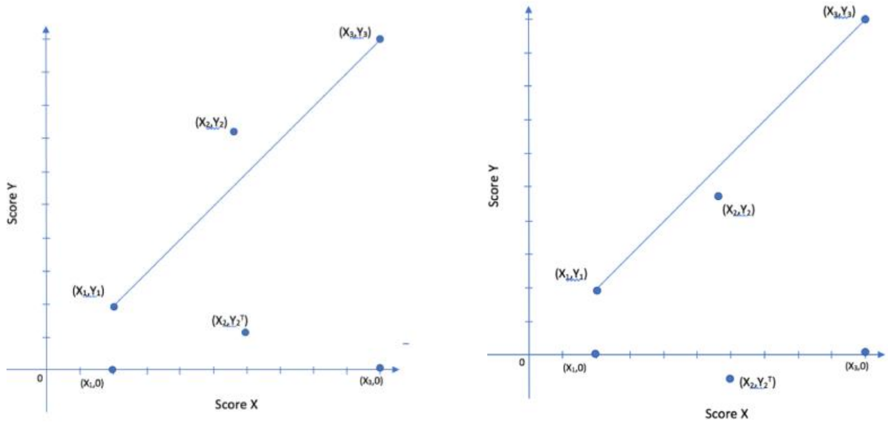

Figure 2. Third Transformation Point

The height of the transformed midpoint (X_2,Y_2^T), has the same distance as the original midpoint (X_2,Y_2) which is above the line L(x). This distance depends on the values of X_2 and Y_2. If the midpoint is above the line connecting the minimum point and the maximum point value, Y_2^T will be positive as shown in Figure 2a. Conversely, if the midpoint is above the line connecting the minimum point and the maximum point value, the value of Y_2^T will be negative as shown in Figure 2b.

Figure 2 shows 3 transformation points used to determine curvilinear on the Simplified Circle Arc method. The three transformation points are used to determine the value of $Y_{1,2}^*$ for each value of the X test device. The index r shows the radius of the circle with the coordinates of the center point (X_c, Y_c) according to equation (1) and (2). The center of the circle on the Simplified Circle Arc method changes according to the new known points (the minimum point becomes (X_1,0) and the maximum point becomes (X_2,0). Based on the center circle equation (Xc) with known 3 points (X_1,Y_1); (X_2,Y_2); and (X_3,Y_3), in the Simplified Circle Arc method, the equation will be,

$$\begin{aligned}
X_c &= \frac{(X_1^2+Y_1^2)(Y_3-Y_2)+(X_2^2+Y_2{}^2)(Y_1-Y_3)+(X_3^2+Y_3^2)(Y_2-Y_3)}{2[X_1(Y_3-Y_2)+X_2(Y_1-Y_3)+X_3(Y_2-Y_1)]} \\
&= \frac{(X_1^2+0)(0-Y_2{}^T)+\left(X_2^2+Y_2{}^{T^2}\right)(0-0)+(X_3^2+0)(Y_2{}^T-0)}{2[X_1(0-Y_2{}^T)+X_2(0-0)+X_3(Y_2{}^T-0)]} \\
&= \frac{(X_1^2)(-Y_2{}^T)+(X_3^2)(Y_2{}^T)}{2[X_1(-Y_2{}^T)+X_3(Y_2{}^T)]} \\
&= \frac{Y_2{}^T(X_3^2-X_1^2)}{2Y_2{}^T[X_3-X_1]} \\
&= \frac{(X_3^2-X_1^2)}{2[X_3-X_1]}
\end{aligned}$$

So that the equation Xc is obtained for the Simplified Circle Arc method as follows:

$$X_{c(SCA)} = \frac{(X_3^2-X_1^2)}{2[X_3-X_1]} \tag{2}$$

while for Yc on the Simplified Circle Arc method the equation becomes,

$$\begin{aligned}
Y_c &= \frac{(X_1^2+Y_1^2)(X_3-X_2)+(X_2^2+Y_2^2)(X_1-X_3)+(X_3^2+Y_3^2)(X_2-X_3)}{2[Y_1(X_3-X_2)+Y_2(X_1-X_3)+Y_3(X_2-X_1)]} \\
&= \frac{(X_1^2+0)(X_3-X_2)+\left(X_2^2+Y_2^{T^2}\right)(X_1-X_3)+(X_3^2+0)(X_2-X_3)}{2\left[0(X_3-X_2)+(Y_2^T)^2(X_1-X_3)+0(X_2-X_1)\right]} \\
&= \frac{(X_1^2)(X_3-X_2)+\left(X_2^2+Y_2^{T^2}\right)(X_1-X_3)+(X_3^2)(X_2-X_3)}{2\left[Y_2^{T^2}(X_1-X_3)\right]}
\end{aligned}$$

so that the Yc equation for the Simplified Circle Arc method is obtained as follows:

$$Y_{c(SCA)} = \frac{(X_1^2)(X_3-X_2)+\left(X_2^2+Y_2^{T^2}\right)(X_1-X_3)+(X_3^2)(X_2-X_3)}{2\left[Y_2^{T^2}(X_1-X_3)\right]} \tag{3}$$

for the equation of the circle radius (r) as follows,

$$r_{SCA} = \sqrt{\left(X_{c(SCA)} - X_1\right)^2 + Y_{c(SCA)}^2} \tag{4}$$

Thus, for the Simplified Circle Arc method in determining the center point, equations (2) and (3) are used using notation $X_{c(SCA)}$ and $Y_{c(SCA)}$ as a differentiator for the center point in the Symmetric Circle Arc method. Likewise, the radius uses equation (4) with r_{SCA} notation. Equation (5) and equation (6) may be used to determine the curve component for all X values in the circular arc produced by the transformation in the Simplified Circle Arc technique (6). Both of these equations are used as formulas to determine the curvilinear component in the Simplified Circle Arc method (Y_1^* or Y_2^*). Things that are different in equations (5) and (6) are used in the Simplified Circle Arc method, which is the center of a circle calculated by a different formula with notation $X_{c(SCA)}$ and $Y_{c(SCA)}$and radius r_{SCA} as follows:

$$Y_1^* = Y_c + \sqrt{r^2 - (X - X_c)^2}$$
$$\text{to } Y_1^* = Y_{c(SCA)} + \sqrt{r_{SCA}^2 - (X - X_{c(SCA)})^2} \quad (5)$$

$$Y_2^* = Y_c - \sqrt{r^2 - (X - X_c)^2}$$
$$\text{to } Y_2^* = Y_{c(SCA)} - \sqrt{r_{SCA}^2 - (X - X_{c(SCA)})^2} \quad (5)$$

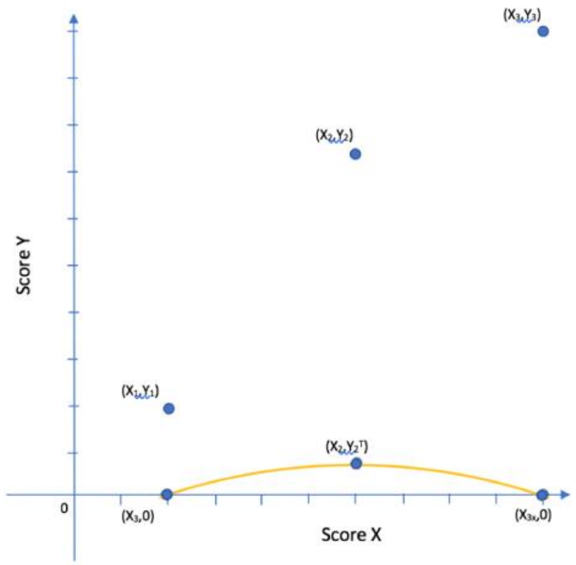

Figure 3. Coordinate of Simplified Circle Arc Transformation Points

Figure 3 shows a circular arc that passes the minimum point (X_1,0) and maximum point (X_3,0) and the midpoint X_2,Y_2^T. The arc determination uses equation (5) or (6) based on the value of Y_2^T obtained. When the arc is continued it will form a circle centered on the point $(X_{c(SCA)}, Y_{c(SCA)})$ and radius r_{SCA}.

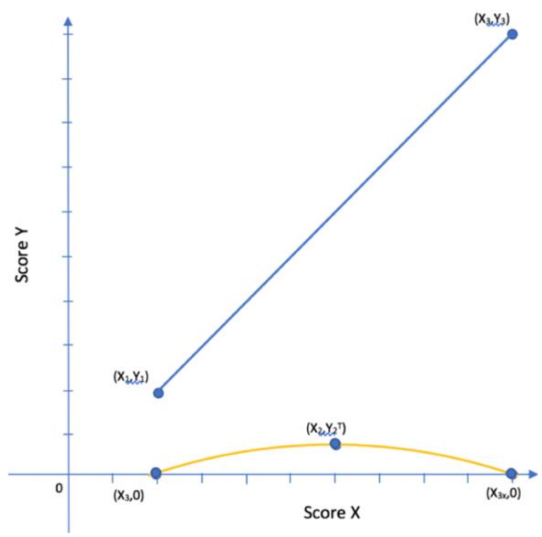

Figure 4. Linear Components and Curvelinear in the Simplified Circle Arc Method

In Figure 4, the curvilinear component (Y_1^* or Y_2^*) is an orange arc and a linear component calculated from L(x). The result of the curve is the estimation of the equalization function for the score which is between the extreme point (minimum point and maximum point). For the Simplified Circle Arc method, the equalization function consists of a linear component (L(x))as described in equation (1) and the curvilinear component (Y_1^* or Y_2^*) as described in equation (5) or equation (6). Both are calculated separately. The sum of both (Y (L(x) dan $Y_1^*{}_{or\ 2}$) is then used to determine the results of the Simplified Circle Arc method.

2. Data Distribution

According to Dziak et al. (2014), the accuracy of each statistical application depends on two main factors, namely sample size and the original form of population distribution. In keeping with this idea, research conducted by Uysal and Kilmen (2016) suggests that the distribution of talents also has an effect on the outcomes of equalization. In order to get an accurate picture of the capabilities of those who participated in the research, a contemporary theoretical methodology was used. When compared to more contemporary ways, the classical method may do the same objective in a far more straightforward manner. Nevertheless, the raw score is the traditional approach that is used, and because of this, the distribution that is evaluated is that of the respondents' data rather than their skill. In addition, according to Uysal and Kilmen, who split the three distributions into Normal, Positive Skewness, and Negative Skewness, the Normal distribution is the most common. The normal distribution is determined by two things, namely average (μ) and standard deviation (σ). In writing it is N (μ,σ). For a normal distribution, it has a value of N (0.1) which is called a normal standard distribution. Between the mean and median in a population is symmetrical by looking at the ratio equal to 1 (Cain et al., 2017; Singh & Masuku, 2014).

Some ordinary researchers use formal tests for normality tests such as Shapiro-Wilk or Kolmogorov-Smirnov. They arrange a hypothesis that states the normality of data. By looking at the probability values that are generated which are then used to determine the normality of a data (Sainani, 2012). Furthermore, according to Nornadiah, Mohd Razali and Yap Bee Wah used 4 types of formal tests to determine normality, namely: Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, and Anderson-Darling (Razali & Wah, 2011).

As with normal distribution, positive skewness (right) and negative skewness (left) also have several provisions. If positive skewness (right) has an average and median ratio of more than 1 because in this distribution the average value will usually be greater than the median (Cain et al., 2017). Similarly, skewness to the right means the right tail is relatively long with the left tail. If the long tail is located on the right side of the slash, then the slash is positive or accurate. The slope is considered to be left or negative if the long tail is located on the left (Hippel, 2010). According to the findings of previously conducted research, ability distribution is another factor that influences how closely outcomes align with expectations (Uysal & Kilmen, 2016). The distribution of data in this research contrasts with the distribution of abilities seen in a prior study because of the difference in the method of distribution utilized.

3. Data Description

The degree to which items are suitable for use with the Rasch model is measured by analysis of the Rasch model. The term "item fit" describes this phenomenon. Ee & Yeo (2018), Falani et al. (2022), Muller (2020), and Rahayu et al. (2020) provide criteria for determining whether or not an item is properly fitted for use in taking measurements. However, there are experts who warn against using the ZSTD criteria if the sample size is more than 500 persons. According to the results of the fit item analysis, the item 30 for the X test equipment (POC1101 code) is the only one that does not fit because it has an MNSQ of 1.70. On the other hand, the Y test devices (POC5530 question code) obtained a total of 4 items that do not fit, specifically points 3, 10, 37, and 26 with respective MNSQ values of 2.73, 1.76, 1.58, and 1.51.

The computation of the Rasch model and the design of the earlier study led to the selection of 30 items to be utilized as research instruments. Within these 30 things, there were 6 anchor items, which accounted for 20% of the overall number of items. For X test equipment (POC1101 question code) anchor items are items 2, 6, 17, 22, 34, and 38 (as many as 6 items), while for nonanchor items are items 1, 3, 4, 5, 7, 8, 11, 12, 13, 15, 19, 20, 21, 23, 25, 28, 29, 31, 32, 33, 35, 36, 37, and 40 (as many as 24 items). For Y test devices (POC5530 question code) anchor items are items 4, 8, 17, 29, 32, and 39 (as many as 6 items), while for nonanchor items are items 1, 2, 5, 6, 7, 9, 11, 13, 14, 15, 16, 19, 22, 23, 25, 28, 30, 31, 33, 34, 35, 36, 38, and 40 (as many as 24 items). Thus, for the two test devices, there are 30 questions with details of each of the 24 nonanchor items and 6 anchor items.

The purpose of the research is to determine if the instrument that was used assesses the unidimensional build, as well as how well each item fits in the underlying design. In accordance with the findings of the Rasch model analysis, the raw variance value for the X test device (identified by the code POC1101) is 30.4%, whereas the raw variance value for the Y test device (identified by the code POC5530) is 33.1%. These findings are presented in comparison to one another. Both of these numbers are higher than the 20% threshold that constitutes the minimum value for the unidimensional criterion (Hsiao et al., 2015; Sinnema et al., 2016). The following are the outcomes of RMSE calculations performed on each set of data distribution pairings using 50 sample sizes, 50 replications, a total of 30 items, and an anchor item that accounts for 20% of the total number of items, as shown in Table 1, Figure 5 and Figure 6.

Table 1. Data Description

Group	Mean	Std. Deviation	Minimum	Maximum
N-N	.3424	.27602	.05	1.10
SP-SP	.4405	.31629	.05	1.37
SN-SN	.2852	.18766	.06	.70
N-SP	.6245	.32490	.07	1.18
N-SN	.7230	.45101	.06	2.11
SP-SN	.6179	.45462	.04	1.91

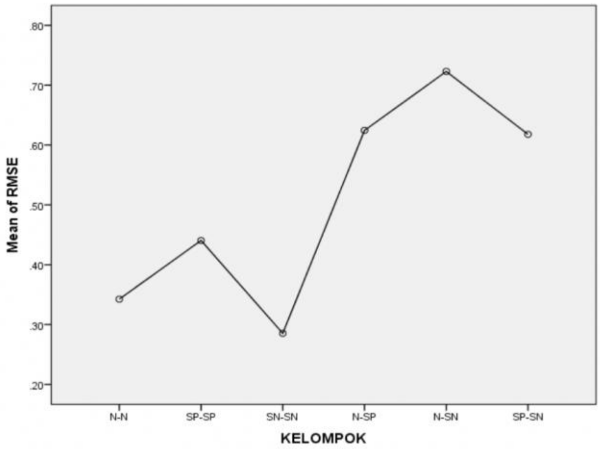

Figure 5. Graph of Differences in Average RMSE Paired Data Groups

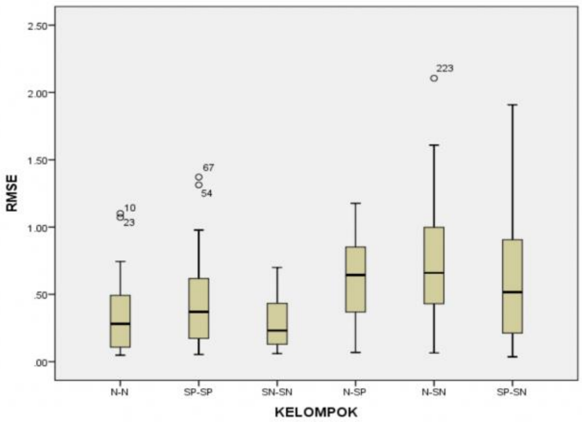

Figure 6. Boxplot of RMSE Value Pair of Data Groups that are Equated

As can be seen in Table 1, above, the average RMSE value for data pairs that have the same distribution, such as a normal distribution with a normal distribution, skewness that is positively distributed with a positive skewness distribution, and distribution of negative skewness with a negative skewness distribution, is lower than the average RMSE value for data pairs that have different data distributions. This is the case regardless of whether the distributions are normal, positive, or negative. This is supported by the views of a small but vocal group of experts in the field; for example, one of these experts argues that the distributions of the two test devices must have the same form is responsible for the reliable comparisons that were made between the two sets (Liemohn et al., 2021; Wang & Lu, 2018). The distributions of the two test devices must be identical lends credence to this view. This point of view might be responsible for the truth of the previously stated assertion.

In their research on the distribution of capabilities, Uysal and Kilmen (2016) distinguished between three distinct distributional patterns: the normal distribution, the distribution with positive skewness, and the distribution with negative skewness. According to the findings, the Equating Error was minimized when groups had the same distribution of abilities (normal distribution with normal distribution, positive skewness distribution with positive skewness distribution, and negative skewness distribution with negative skewness distribution). This was the case for normal distributions, positive skewness distributions, and negative skewness distributions. In addition, proper equalization results are offered by the similarities in the form of the beginning distribution of the two testing devices that are being compared to one another. This is in agreement with what Souza et al (2017) has said, which is that the two test devices' modes of distribution must be identical. The distributions of the two sets of data must be the same in order for the equalization findings to be reliable; this is in addition to the fact that the score may be equalized by using the classical technique (Uto, 2021). This is something that may help ensure the equalization findings are reliable. This is so even if the mean and standard deviation are two different quantities. Muraki, Hombo, & Lee The same thing has been done by Tong and Kolen, which demonstrates that when the raw score distribution is the same, the outcomes of the equalization will be positive (Schalet et al., 2021). It is possible to compare or match the results of two or more testing devices provided that they have the same distribution (Moses, 2022). According to von Davier et al. (2019), one of the assumptions that must be made in order to get a satisfactory outcome with equalization is the existence of secondary distributions between the two scales or test devices.

According to Aşiret and Sünbül (2016), the Circle Arc technique produces reliable results when the sample value is at least 50 for any number of samples in the range from 25 to 200. A low RMSE value is used as the basis for measuring the accuracy value. It was also clarified that even when taking into account the varying degrees of difficulty posed by the various test kits, the Circle Arc approach produces satisfactory results. Because of this, the Circle Arc technique has a wider range of applications. It would seem that the top whisker line in Figure 1 is longer than the lower whisker line for pairings of data groups that have the same distribution. This can be seen by comparing the lengths of the two lines. Assuming the same starting data distribution, the root-mean-squared error (RMSE) for equalization using the SCA technique has the same properties whether the data are normally distributed, positively skewed, or negatively skewed. These findings assume that the underlying data set was normally distributed to begin with. Results from equalization using the SCA method are those whose distributions are both normal and uniform; that is, the skewness of the distribution is uniformly positive. The longest portion of the whisker line is located at the top, while the shortest is at the bottom. Beyond the obvious alcohol's positive effects, all three have other characteristics. On the basis of this information, one may get the conclusion that the RMSE value for all three typically resulted in rather minor findings. The range that was produced as a consequence of this is fairly vast, and the value of drinking is comparatively higher than the value produced by pairs of data groups that have uneven data distribution.

When it comes to equalization, the most common model is one based on linear equations. However, this approach assumes (without checking the data) that scores on test device X and test device Y are distributed normally (mean standard deviation) within the population of interest (Altintas & Wallin, 2021). Given this, it's tough to buy into the idea that a uniform set of guidelines governs the development of the test kit. However, the normalization links between the different test formats may not always follow a linear pattern in circumstances when the test formats give diverse degrees of difficulty. Since it is likely that X and Y testing devices would provide various levels of difficulty, nonlinear methods are a useful option in these cases (Albano, 2015). Unless you use the Circle Arc approach, which specifies nonlinear interactions between scales, this is not the case. When using the SCA technique, the component of the scale that represents the center of the distribution is the one that provides the most accurate estimation of the equalized score. This is particularly true for samples that are relatively small. This is due to the fact that the approach in question utilizes a curve as its equalization distribution form. In accordance with this, Kim and Livingston claimed that the SCA approach exhibited high accuracy in the region of the 25 to 75 percentiles, particularly in small samples (LaFlair et al., 2017). In addition, Kim and Livingston state that the findings of equalization accuracy will have excellent accuracy when the two groups that are being compared have the same distribution. If this is the case, then the SCA technique will have good accuracy.

Livingston & Kim conducted a study comparing the Chained, Levine, Tucker, Mean, Circle Arc, and Identity methods (O'Neill et al., 2020). The study uses small samples and uses the Root Mean Square Error (RMSE) and bias as a tool to evaluate the equalization results. The results of the research conducted indicate that the Circle Arc method provides a low RMSE value and bias compared to other methods. Ozdemir (2017) conducts research by comparing the methods included in the nonlinear category namely Equipersentil and Circle Arc. This study aims to equalize TIMSS 2011 data with TIMSS 2007 by using anchor items. The results obtained based on the results of Root Mean Square Error (RMSE) and Mean of Bootstrap Standard Error (MBSE) indicate that the Circle Arc method is better than the Equipercentile and Presmoothing method in terms of decreasing Error Standards and bias. Aşiret & Sünbül (2016) conducted a study comparing several methods of equalization and the number of samples. The study used the Root Mean Square Error (RMSE) as a criterion in evaluating the results of the equivalence. The results of the study show that the Circle Arc and Mean methods in the number of samples are 50 or more (in the study using samples 10, 25, 75, 100, 150, and 200) with different levels of difficulty giving Equating Errors lower than other equalization methods.

D. CONCLUSION AND SUGGESTIONS

In light of the findings of the study and the conversations that have taken place so far, one may reach the following conclusion: The Simplified Circle Arc equalization technique can be employed as an alternate equalization method when tiny samples are involved. This is supported by the results of the analysis using RMSE, where this method always produces smaller errors than other methods. A small RMSE value indicates high accuracy. With the existence of an accurate score equalization method with fairly simple use because it uses a classical test theory approach, it is able to answer the problem of assigning grades at the school level. This method of equalization may be used at the class level, provided that the criterion of the number of students belonging to a small sample is satisfied, and that attention is paid to the similarities in the distribution of data values among students who are comparable.

ACKNOWLEDGEMENT

The Center for Educational Assessment has released the scores from the National Exams for mathematics courses taken at DKI Jakarta and Tangerang. Many thanks are due to the institutions and the Center for Educational Assessment for their generous distribution of resources.

REFERENCES

Albano, A. D. (2015). A General Linear Method for EquatingWith Small Samples. *Journal of Educational Measurement*, *52*(1), 55–69. https://doi.org/10.21831/pep.v19i1.4551

Altintas, Ö., & Wallin, G. (2021). Equality of admission tests using kernel equating under the non-equivalent groups with covariates design. *International Journal of Assessment Tools in Education*, *8*(4), 729–743. https://doi.org/10.21449/ijate.976660

Antara, A., & Bastari, B. (2015). Penyetaraan Vertikal Dengan Pendekatan Klasik Dan Item Response Theory Pada Siswa Sekolah Dasar. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *19*, 13–24. https://doi.org/10.21831/pep.v19i1.4551

Arikan, Ç. A., & Gelbal, S. (2018). A comparison of traditional and kernel equating methods. *International Journal of Assessment Tools in Education*, *5*(3), 417–427. https://doi.org/10.21449/ijate.409826

Aşiret, S., & Sünbül, S. Ö. (2016). Investigating test equating methods in small samples through various factors. *Educational Sciences: Theory & Practice*, *16*(2). https://doi.org/https://doi.org/10.12738/estp.2016.2.2762

Babcock, B., & Hodge, K. J. (2020). Rasch versus classical equating in the context of small sample sizes. *Educational and Psychological Measurement*, *80*(3), 499–521. https://doi.org/10.1177/0013164419878483

Battauz, M. (2017). Multiple equating of separate IRT calibrations. *Psychometrika*, *82*(3), 610–636. https://doi.org/10.1007/s11336-016-9517-x

Caglak, S. (2016). Comparison of several small sample equating methods under the NEAT design. *Turkish Journal of Education*, *5*(3), 96–118. https://doi.org/10.19128/turje.16916

Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, *49*, 1716–1735. https://doi.org/10.3758/s13428-016-0814-1

Diao, H., & Keller, L. (2020). Investigating repeater effects on small sample equating: Include or exclude? *Applied Measurement in Education*, *33*(1), 54–66. https://doi.org/10.1080/08957347.2019.1674302

Dimitrov, D., & Atanasov, D. (2021). An Approach to Test Equating under the Latent D -scoring Method. *Measurement: Interdisciplinary Research and Perspectives*, *19*, 153–162. https://doi.org/10.1080/15366367.2020.1843107

Dorans, N. J., & Puhan, G. (2017). Contributions to score linking theory and practice. *Advancing Human Assessment: The Methodological, Psychological and Policy Contributions of ETS*, 79–132. https://library.oapen.org/bitstream/handle/20.500.12657/28140/1001854.pdf?sequence=1

Dziak, J. J., Lanza, S. T., & Tan, X. (2014). Effect size, statistical power, and sample size requirements for the bootstrap likelihood ratio test in latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 534–552. https://doi.org/10.1080/10705511.2014.919819

Ee, N. S., & Yeo, K. J. (2018). Item Analysis for the Adapted Motivation Scale Using Rasch Model. *International Journal of Evaluation and Research in Education*, *7*(4), 264–269. https://eric.ed.gov/?id=EJ1198607

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*, 195–212. https://doi.org/10.3758/s13428-017-0862-1

Falani, I., Iriyadi, D., Ice, Y. W., Susanti, H., & Nasuition, R. A. (2022). A Rasch analysis of perceived stigma of covid-19 among nurses in Indonesia questionnaire. *Psychological Thought*, *15*(1), 12. https://doi.org/10.37708/psyct.v15i1.530

Haris, D. J., & Kolen, M. J. (2016). A Comparison of Two Equipercentile Equating Methods for Common Item Equating. *Educational and Psychological Measurement*, *50*(1). https://doi.org/https://doi.org/10.1177/0013164490501

Himelfarb, I. (2019). A primer on standardized testing: History, measurement, classical test theory, item response theory, and equating. *Journal of Chiropractic Education*, *33*(2), 151–163.

Hippel, P. V. (2010). Skewness. *International Encyclopedia of Statistical Science*, *100*, 1–4. https://doi.org/10.4135/9781412952644

Hsiao, Y. Y., Shih, C. L., Yu, W. H., Hsieh, C. H., & Hsieh, C. L. (2015). Examining unidimensionality and improving reliability for the eight subscales of the SF-36 in opioid-dependent patients using Rasch analysis. *Quality of Life Research*, *24*(2), 279–285. https://doi.org/10.1007/s11136-014-0771-z

Iriyadi, D., Asdar, A. K., Afriadi, B., Samad, M. A., & Syaputra, Y. D. (2024). Analysis of the Indonesian Version of the Statistical Anxiety Scale (SAS) Instrument with a Psychometric Approach. *International Journal of Evaluation and Research in Education*, *13*(2). http://doi.org/10.11591/ijere.v13i2.26517

Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, *585*, 609–629. https://doi.org/10.1016/j.ins.2021.11.036

Kemendikbud. (2017). Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 17 Tahun 2017: Penerimaan peserta didik baru pada taman kanak-kanak, sekolah dasar, sekolah menengah pertama, sekolah menengah atas, sekolah menengah kejuruan, atau bentuk lain. In *Retrieved from* https://psma.kemdikbud.go.id/index/home/lib/files/SALINAN%20PPDB.pdf.

Kolen, M. J., & Brennan, R. L. (2014). *Test Equating, Scaling, and Linking (Third Edit)*. Springer. https://doi.org/10.1007/978-1-4939-0317-7

LaFlair, G. T., Isbell, D., May, L. D. N., Gutierrez Arvizu, M. N., & Jamieson, J. (2017). Equating in small-scale language testing programs. *Language Testing*, *34*(1), 127–144. https://doi.org/10.1177/026553221562082

Liemohn, M. W., Shane, A. D., Azari, A. R., Petersen, A. K., Swiger, B. M., & Mukhopadhyay, A. (2021). RMSE is not enough: Guidelines to robust data-model comparisons for magnetospheric physics. *Journal of Atmospheric and Solar-Terrestrial Physics*, *218*, 105624.

Lissitz, R. W., & Huynh, H. (2019). Vertical equating for state assessments: Issues and solutions in determination of adequate yearly progress and school accountability. *Practical Assessment, Research, and Evaluation*, *8*(1), 10. https://doi.org/10.7275/npzw-wd59

Livingston, S. A. (2014). Equating test scores (without IRT). *Educational Testing Service*. https://eric.ed.gov/?id=ED560972

Masino, S., & Niño-Zarazúa, M. (2016). What works to improve the quality of student learning in developing countries? *International Journal of Educational Development*, *48*, 53–65. https://doi.org/10.1016/j.ijedudev.2015.11.012

Moses, T. (2022). Linking and comparability across conditions of measurement: Established frameworks and proposed updates. *Journal of Educational Measurement*, *59*(2), 231–250. https://doi.org/10.1111/jedm.12322

Müller, M. (2020). Item fit statistics for Rasch analysis: can we trust them? *Journal of Statistical Distributions and Applications*, *7*, 1–12. https://link.springer.com/article/10.1186/s40488-020-00108-7

O'Neill, T. R., Gregg, J. L., & Peabody, M. R. (2020). Effect of sample size on common item equating using the dichotomous Rasch model. *Applied Measurement in Education*, *33*(1), 10–23. https://doi.org/10.1080/08957347.2019.1674309

Ozdemir, B. (2017). Equating TIMSS Mathematics Subtests with Nonlinear Equating Methods Using NEAT Design: Circle-Arc Equating Approaches. *International Journal of Progressive Education*, *13*(2), 116–132. https://eric.ed.gov/?id=EJ1145605

Peabody, M. R. (2020). Some methods and evaluation for linking and equating with small samples. *Applied Measurement in Education*, *33*(1), 3–9. https://doi.org/10.1080/08957347.2019.1674304

Pommerich, M. (2016). The fairness of comparing test scores across different tests or modes of administration. *Fairness in Educational Assessment and Measurement*, 111–134. https://doi.org/10.4324/9781315774527-9

Rahayu, W., Putra, M. D. K., Iriyadi, D., Rahmawati, Y., & Koul, R. B. (2020). A Rasch and factor analysis of an Indonesian version of the Student Perception of Opportunity Competence Development (SPOCD) questionnaire. *Cogent Education*, *7*(1), 1721633. https://doi.org/https://doi.org/10.1080/2331186X.2020.1721633

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33. https://doi.org/https://doi.org/doi:10.1515/bile-2015-0008

Sainani, K. L. (2012). Dealing With Non-normal Data. *PM and R*, *4*(12), 1001–1005. https://doi.org/10.1016/j.pmrj.2012.10.013

Schalet, B. D., Lim, S., Cella, D., & Choi, S. W. (2021). Linking scores with patient-reported health outcome instruments: A validation study and comparison of three linking methods. *Psychometrika*, *86*(3), 717–746. https://doi.org/10.1007/s11336-021-09776-z

Singh, A. S., & Masuku, M. B. (2014). Sampling techniques & determination of sample size in applied statistics research: An overview. *International Journal of Economics, Commerce and Management*, *2*(11), 1–22. https://www.researchgate.net/publication/341552596_Sampling_Techniques_and_Determination_of_Sample_Size_in_Applied_Statistics_Research_An_Overview

Sinnema, C., Ludlow, L., & Obinson, V. (2016). Journal of Educational Administration and History. *Journal of Educational Administration and History*, *35*(2). https://doi.org/10.1080/713676155

Souza, A. C. de, Alexandre, N. M. C., & Guirardello, E. de B. (2017). Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e Servicos de Saude*, *26*, 649–659. https://doi.org/10.5123/S1679-49742017000300022

Sumarni, W. (2015). The strengths and weaknesses of the implementation of project based learning: A review. *International Journal of Science and Research*, *4*(3), 478–484. https://www.ijsr.net/archive/v4i3/SUB152023.pdf

Uto, M. (2021). Accuracy of performance-test linking based on a many-facet Rasch model. *Behavior Research Methods*, *53*(4), 1440–1454. https://doi.org/10.3758/s13428-020-01498-x

Uysal, İ., & Kilmen, S. (2016). Comparison of Item Response Theory Test Equating Methods for Mixed Format Tests. *International Online Journal of Educational Sciences*, *8*(2). https://doi.org/https://doi.org/10.15345/iojes.2016.02.001

von Davier, M., Yamamoto, K., Shin, H. J., Chen, H., Khorramdel, L., Weeks, J., Davis, S., Kong, N., & Kandathil, M. (2019). Evaluating item response theory linking and model fit for data from PISA 2000–2012. *Assessment in Education: Principles, Policy & Practice*, *26*(4), 466–488. https://doi.org/10.1080/0969594X.2019.1586642

Wang, W., & Lu, Y. (2018). Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model. *IOP Conference Series: Materials Science and Engineering*, *324*, 12049. https://doi.org/10.1088/1757-899X/324/1/012049

Yu, C. H., & Osborn-Popp, S. E. (2019). Test equating by common items and common subjects: Concepts and applications. *Practical Assessment, Research, and Evaluation*, *10*(1), 4. https://www.researchgate.net/publication/253274721_Test_Equating_by_Common_Items_and_Common_Subjects_Concepts_and_Applications