Distribution Dependent Equalization of Scores towards Removing Examiners' Bias: A Case Study of an English Paper

S. SAHU[1], K. HARI RAJAN[2], S. MAITI[3] , G. MAHAPATRA[4] S. SAHU[5]

[1]Prof. (Dr.) Somen Sahu, Professor and Head, Department of Fishery Economics and Statistics, West Bengal University of Animal and Fishery Sciences, West Bengal, Kolkata, India, somenndp69@gmail.com

[2]Chairman, West Bengal Police Recruitment Board, Kolkata, West Bengal, India, hari.rajan58@gmail.com

[3]Professor and Head, Department of Statistics, Visva-Bharati University, Shantiniketan, Birbhum, West Bengal, India, dssm1@rediffmail.com

[4]Associate Professor, Department of Computer Science, Asutosh College, Calcutta University, Kolkata, West Bengal, India, gsp2ster@gmail.com

[5]City College, University of Calcutta, Kolkata, West Bengal, India, syammu56@gmail.com

Abstract **Recruitment process now-a-days involves primarily either pen & paper or a computer based examination for the candidates being conducted by a recruiting authority. This method of selection has a basic drawback as the answer scripts are almost always evaluated independently by several examiners who may be either strict or lenient in their evaluation. These results in rating bias for the scores and this can be overcome by application of equi-percentile method. In this case, the examination administrators use the equi-percentile method for normalization. With this equi-percentile method, a toughly evaluated paper and a softly evaluated one are brought to one scale level. This paper presents a case study of a similar situation where the examination scores of English paper of 6194 examinees were evaluated by 26 examiners (after proper coding) during April-May 2019 (for a Graduate level entry post) were statistically converted to a scaled score using equi-percentile method. In this case maximum median is 21 corresponding to Examiner-14 i.e. EN. So, all the other marks given by different examiners are transferred to the same distribution which prevailed with examiner-14. Statistically, a median of medians is the thumb-rule and any of the examiners could be chosen as a reference (examiner). This also satisfies the method since it would be considering the underlying distribution of the reference examiner. However, taking the median of medians as the reference examiner may lead to examinees with higher raw scores being awarded lower scaled scores resulting in grievances for the test takers. In this case maximum median is 21 corresponding to Examiner-14 i.e. EN. So all the other marks given by different examiners are transferred to the same distribution which prevailed in examiner-14 scores and thus all the raw scores converted to scaled scores.**

*Keywords — **Equalization of scores, Equi-percentile method, median, percentile rank, rating bias, Examiners' bias***

 DOI : 10.35291/2454-9150.2020.0176

I. INTRODUCTION

Present day recruitment process involves primarily either pen & paper or computer based examination for the candidates being conducted by a recruiting authority. This method of selection has a basic drawback as the answer sheets are almost always evaluated independently by several examiners who may be either strict or lenient in the evaluation process. The situation may be referred to as dove and hawk examiners wherein an answer script of a particular paper may be assigned to a dove securing an upscale marks. On the other hand a different answer book for the same paper assigned to a hawk may fetch low marks down the scale resulting in rating bias. To overcome such an anomaly, entrance or recruitment exams of several prestigious organizations like IBPS or Bankers' exam, UP Police Recruitment & Promotion Board relied on equi-percentile method. In this case the examination administrators use the equi-percentile method for normalization. With this equi-percentile method, a hawkishly evaluated paper and a leniently evaluated one are brought to at par. The same constraint is also faced when there exists difference in difficulty level for two or more sets of question papers for a single recruitment due to the size of the applicants. This problem can also be nullified by equi-percentile method. This paper presents a case study of a similar situation where the examination scores of English paper of 6194 examinees evaluated by (after proper coding) during April-May 2019 were statistically converted to a scaled score using equi-percentile method.

II. REVIEW OF LITERATURE

In one study Lawton et. al. (2016) demonstrated the method to convert University of Pennsylvania Smell Identification Test (UPSIT) to Brief-SIT (B-SIT) or Sniffin' 16, and Sniffin' 12 to 16 scores in a valid way [1]. This facilitated direct comparison between tests aiding future collaborative analyses and evidence synthesis. Lawton et al (2016) used the equi-percentile and Item Response Theory (IRT) methods to equate the olfaction scales and validated dataset of 128 individuals who took both tests, the Sniffin' 16 (n=1131) or UPSIT (n=980). The equi-percentile conversion suggested some bias between UPSIT and Sniffin' 16 tests across the two groups. The IRT method shows very good characteristics between the true and converted Sniffin' 16 (delta mean = 0.14, median = 0) based on UPSIT. The equi-percentile conversion between the Sniffin' 12 and 16 item worked well (delta mean = 0.01, median = 0). Brossman and Lee (2013) develop observed score and true score equating procedures to be used in conjunction with the multidimensional item response theory (MIRT) framework [2]. Three equating procedures—two observed score procedures and one true score procedure—were created. One observed score procedure was presented as a direct extension of uni-dimensional IRT (UIRT) observed score equating and is referred to as the ''Full MIRT Observed Score Equating Procedure.'' The true score procedure and the second observed score procedure incorporated uni-dimensional approximation procedures to equate exams using UIRT equating principles. These procedures are referred to as the ''Uni-dimensional Approximation of MIRT True Score Equating Procedure'' and the ''Uni-dimensional Approximation of MIRT Observed Score Equating Procedure,'' respectively. Three exams were used to conduct UIRT observed score and true score equating, MIRT observed score and true score equating, and equi-percentile equating. The equi-percentile equating procedure was conducted for the purpose of comparison because this procedure does not explicitly violate the IRT assumption of uni-dimensionality. Results indicated that the MIRT equating procedures performed more similarly to the equi-percentile equating procedure than the UIRT equating procedures, presumably due to the violation of the uni-dimensionality assumption under the UIRT equating procedures. Livingston and Kim (2010) proposed five methods for equating in a random groups design with samples of 50 to 400 Test Takers [3]. The criterion equating was the direct equi-percentile equating in the group of all test takers. Equating accuracy was indicated by the root-mean-squared deviation, over 1,000 replications, of the sample equating from the criterion equating. The methods investigated were equi-percentile equating of smoothed distributions, linear equating, mean equating, symmetric circle-arc equating, and simplified circle-arc equating. The circle-arc methods produced the most accurate results for all sample sizes investigated, particularly in the upper half of the score distribution. The difference in equating accuracy between the two circle-arc methods was negligible Steenoven et. al (2014) applied a simple and reliable algorithm for the conversion of Montreal Cognitive Assessment (MoCA) to Mini-Mental State Examination (MMSE) scores in PD patients [5]. Further, the same algorithm was applied for conversion of Dementia Rating Scale-2 (DRS-2) to both MMSE and MoCA scores. The cognitive performance of a convenience sample of 360 patients with idiopathic PD was assessed by at least two of these cognitive screening instruments. He then developed conversion scores between the MMSE,

 DOI : 10.35291/2454-9150.2020.0176

MoCA, and DRS-2 using equi-percentile equating and log-linear smoothing. The conversion score tables reported enable direct and easy comparison of three routinely used cognitive screening assessments in PD patients.

The classical test theory for mean equating adjusts the distribution of scores so that the mean scores of one examiner is comparable to the mean scores of another. However, this method lacks flexibility, as there exists the possibility for difference in the standard deviations of the scores. Linear equating resolves this issue and adjusts in a way that the two examiners have a comparable mean and standard deviation. Based on assumptions and mathematics used, linear equating are of several types. Equi-percentile equating determines the equating relationship as one where a score could have an equivalent percentile on either form. This relationship can be nonlinear. Equating is explained as transformation on raw-to-raw basis. It involves estimating a raw score on Form Y equivalent to the raw score in base form X with further application of scaling transformations.

In the Indian context attempt was taken by Staff Selection Commission whose methodology is cited. "Staff Selection Commission has been conducting various examinations in multiple batches because of large number of candidates and difficulties in getting adequate educational institutions for holding the examinations in a single batch. For perhaps the first time in its history, the number of applicants in a single examination exceeded one million when the Combined Higher Secondary Level Examination, 2010 for the recruitment of Lower Division Clerks and Data Entry Operators, elicited response from over 16 lakh candidates, with approx. 21% of them applying online. This would require the Examination, rescheduled on 27 & 28.11.2010 (in view of Common Wealth Games), to be held in at least three batches. The Commission, with the help of experts, has striven to construct question papers of comparable difficulty level. While such an exercise is theoretically possible, in practice it is impossible to have two or more question papers of identical difficulty levels. Even if the difficulty levels of question papers vary slightly, candidates taking more difficult papers may be at a disadvantage viz-a-vis others. Therefore, there is a need for equating of the marks in examinations involving multiple batches and question papers… The Commission had examined the views of an Expert Group, constituted by it with the approval of Government of India in 2009, on this issue. The Commission had placed before the Expert Group that the technique to be followed for equating should be transparent, easily comprehensible to the candidates, acceptable to experts and prove itself in Courts of Law if and when challenged. This was accepted by the Expert Group which further advised the Commission to place a paper on the technique on its website for adequate time, give publicity to such placement through the media, invite comments, observations and suggestions and decide on adopting the technique thereafter ……….. Equating is a statistical process that is used to adjust scores on multiple question papers so that scores on the forms can be used interchangeably. It adjusts for differences in difficulty among Question Papers that are built to be similar in difficulty and content. As per the report the expert committee viewed about four methods of Equating viz. (i) Median/Mean Equating, (ii) Linear Equating (Based on mean and S.D.), (iii) Equipercentile Equating, (iv) Equating using Item Response Theory. Among these methods, SSC proposes to use the Equi-percentile Method in view of its simplicity."[7].

III. MATERIALS AND METHODS

Statistical equating defines a functional relationship between multiple test score distributions and thereby between multiple score scales. When the test forms have been created according to the same specifications and are similar in statistical characteristics, this functional relationship is referred to as an equating function and it serves to translate scores from one scale directly to their equivalent values on another. Whether score distributions are based on samples from a single examinee population or different examinee populations (these are referred to as equating designs), if the appropriate assumptions are met the equating function can be generalized to other examinees (Holland & Dorans, 2006) [6]. Equating methods can be used to adjust for differences in difficulty across alternate forms/ judgments, resulting in comparable score scales and more accurate estimates of ability in most of the cases for different sets of examinees examined by different sets of examiners. Here it is assumed that there exists rating biases in the evaluation of the answer scripts by different examiners. It is further assumed that an examiner is homogenous in respect of his/her rating in respect of his/her examinees but heterogeneous with other examiners. Equating types can be categorized as either linear, including mean or linear equating, or nonlinear, equi-percentile equating. An additional nonlinear type is circle-arc equating, as recently introduced by Livingston and Kim (2009). For the present study the methodology of equi-percentile equating is adopted. The percentile of a

 DOI : 10.35291/2454-9150.2020.0176

candidate will reflect how many candidates have scored below that candidate in that batch.

The procedure of Equi-percentile equating method is discussed briefly:

Informally it is used to equate scores on two tests so that the scores reflect the same percentiles should be based on same set of respondents, but often based on randomly equivalent groups.

Formally, X-score x and Y-score y are linked in T if FT(x) = GT(y), When these two CDF's are continuous and strictly increasing, then this equation can always be satisfied. This is a very effective method for equating.

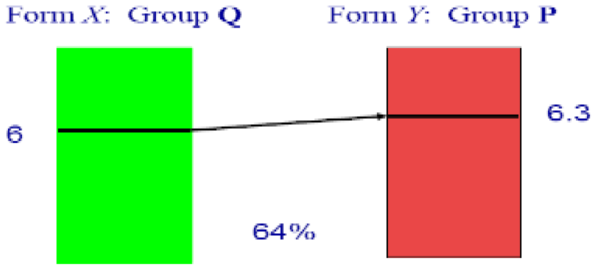


Figure 1: Pictorial representation of equating method

Equi-percentile equating defines a nonlinear relationship between score scales by setting equal the percentile ranks for each score point. Specifically, the equi-percentile equivalent of a form- X score on the Y scale is calculated by finding the percentile rank in X of score i, and then the form-Y score associated with that form-Y percentile rank:

$$eY (xi) = Q-1[P (xi)].$$

Here, P (x) is the percentile rank function in X and Q−1(x) is the inverse percentile rank function in Y. According to the Kolen & Brennan the process is complicated by the fact that scores are discrete, and must be made continuous [4]. Because it involves estimation at each score point, equi-percentile equating is especially susceptible to random sampling error. Smoothing methods are typically used to reduce irregularities in either the score distributions or the equating function itself.

Case study:

1) The case under study involves the test of English (subjective in nature) which has no specified and distinct guideline for awarding marks as it is subjective. As a consequence a difference in evaluation is obvious among the examiners resulting in variation of the marks for the same style, hand and content of writing.

2) The marks scoring pattern for the answer sheets depends on the difficulty level of checking by the examiners and varies among the different examiners entrusted for the purpose.

3) Such variation in scores necessitates normalization. Equi-percentile Method takes care of the differences in difficulty of checking level and resultant rating bias of the examinees.

Therefore, using equi-percentile method, all the raw scores were first converted to a scaled score for each examiner followed by clubbed ranking. This ensures smoothing out the hidden / underlying distribution corresponding to each examiner converting it to a standard scale i.e., percentile scale. This methodology is appropriate for selection procedure where there is no a further scored test or interview and the selection solely depends upon the written exam scores of only one paper or a subject. This methodology is being followed in the following examinations as seen recently viz. RRB, NTPC, CAT, MAT, IBPS, UPPR & PB.

Drawback of the above procedure:

However, the above procedure has a drawback.

In this method it is not possible to incorporate the underlying distribution pattern to the scores and as the ranks are not additive in nature, it cannot be used for more than one subject. To rectify the problem, one examiner is considered to be the standard and chosen as reference. Then the distribution equation for that reference examiner is determined by the method of multivariate analysis. In that equation, percentile rank is considered as independent parameter and raw scores is considered as dependent parameter. Then percentile rank corresponding to each raw scores of each examiner is fitted to the mentioned distribution of the reference examiner and by this way every raw marks awarded by each examiner will be scaled to this particular distribution generating the scaled scores for each individual examinee. Then by clubbing all the scaled scores of the all the examinees it is possible to select the candidates for the next stage of recruitment or say a Interview / Personality Test or Viva -voice, as the case may be.

Moreover, in some selection procedures, a written examination is followed by an interview, the written percentile ranks and interview percentile rank can be clubbed assigning some weightage to these two parameters. These weights may be the ratio of the maximum marks assigned to each test or paper. But as the scores are converted to ranks the weighted method will not give the desired level of efficiency to the selection procedure. The only rectification method is that, after completing the interview by all the interviewers, scores will again be converted to scaled scores are brought by applying the previous procedure. As all the scores where there is a possibility of evaluators' bias thus removed by the above procedure of equi-percentile equating method fitted to some reference distribution generating the scaled scores on an absolute scale. These scaled scores can be taken for selection purpose compatible to other raw scores which are free from human bias.

The collected data was statistically analyzed through SPSS 21.0 and Microsoft Excel Work sheet.

 DOI : 10.35291/2454-9150.2020.0176

ISSN : 2454-9150 Vol-05, Issue-12, Mar 2020

IV. RESULTS AND DISCUSSION

The present study involves the examination scores of 'English' language of a sample size of 6194 examinees. The answer scripts were randomized and distributed among 26 examiners for evaluation during April-May-2019. Although the randomized distribution satisfies the normality for each individual examiner but the inherent bias of the examiners commonly called rating bias is a major drawback. Therefore, the equi-percentile method has to be applied to smoothen out the rating bias. To judge about the central tendency of each examiner the following table depicts the descriptive statistics for the selected sample.

From Table 1 it is evident that Maximum median is 21 corresponding to Examiner-14 i.e. EN. Statistically, a median of medians is the thumb-rule and any of the examiners could be chosen as reference (examiner). This also satisfies the method since it would be considering the underlying distribution of the reference examiner. However, taking the median of medians as the reference examiner may lead to examinees with higher raw scores being awarded lower scaled scores resulting in

Table 1: Distribution of Marks (English) and their descriptive Statistics (Arranged on the basis of median)

Code	Examiner Code	Freq.	Mean	Median	SD	Min	Max	Code	Examiner Code	Freq.	Mean	Median	SD	Min	Max
EC	EX-3	80	6.15	5.00	5.25	0.00	22.00	EU	EX-21	150	14.47	14.00	7.82	0.00	33.00
EE	EX-5	60	5.80	5.50	4.26	0.00	16.00	EQ	EX-17	300	14.89	15.00	5.82	0.00	33.00
ED	EX-4	80	7.65	6.00	5.60	1.00	28.00	EA	EX-1	80	15.71	16.00	6.09	1.00	29.00
ES	EX-19	300	8.07	6.00	6.18	0.00	36.00	EO	EX-15	300	17.20	16.00	9.26	0.00	42.00
EB	EX-2	99	8.61	8.00	4.80	0.00	21.00	ET	EX-20	300	16.66	16.00	6.57	0.00	35.00
EL	EX-12	300	10.44	9.00	6.06	0.00	29.00	EZ	EX-26	345	16.18	16.00	8.79	0.00	36.00
EM	EX-13	300	9.87	9.00	5.29	0.00	35.00	EF	EX-6	300	17.30	17.00	7.70	0.00	37.00
EY	EX-25	300	10.47	9.00	7.69	0.00	33.00	EW	EX-23	300	17.94	18.00	6.76	0.00	35.00
EJ	EX-10	200	10.22	9.50	6.65	0.00	31.00	EG	EX-7	300	19.93	20.00	8.18	0.00	39.50
EI	EX-9	300	10.85	10.00	5.94	0.00	32.00	ER	EX-18	300	19.71	20.00	4.83	1.00	31.00
EH	EX-8	300	12.45	12.00	6.75	0.00	35.00	EV	EX-22	150	18.63	20.00	6.89	2.00	31.00
EX	EX-24	300	13.56	12.00	7.39	0.00	39.00	EK	EX-11	150	19.86	21.00	6.81	0.00	33.00
EP	EX-16	300	13.90	13.00	6.56	1.00	30.00	EN	**EX-14**	300	21.02	**21.00**	7.48	2.00	40.00
Total number of Examinees = 6194, Total number of Examiners = 26, Evaluated Paper: English (Subjective in nature)															

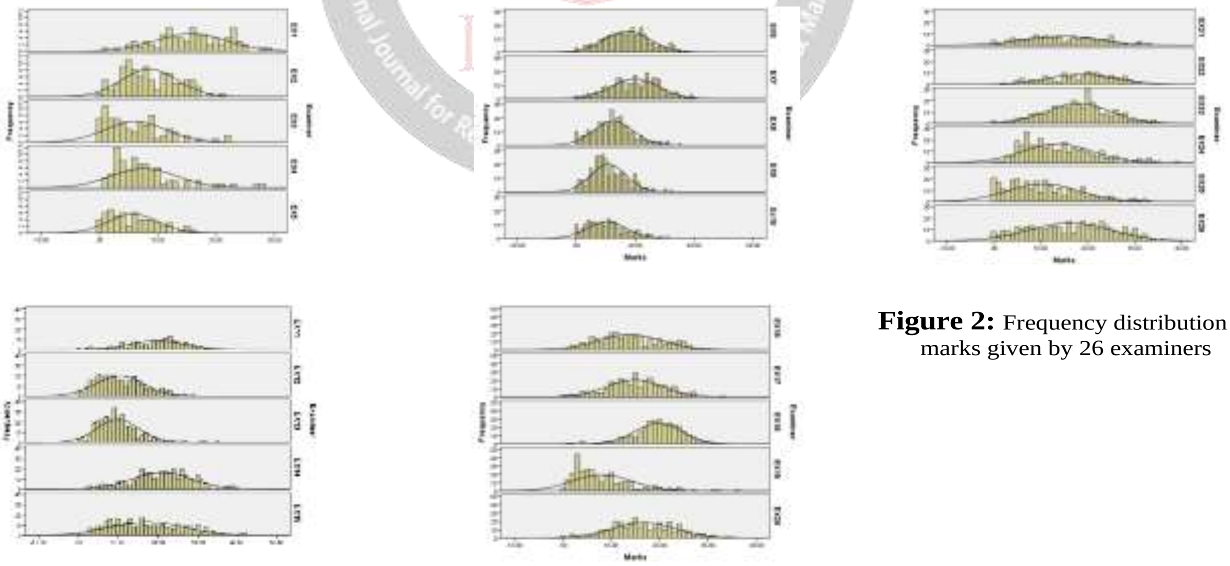


Figure 2: Frequency distribution of marks given by 26 examiners

Table 2: Model Summary and Parameter Estimates

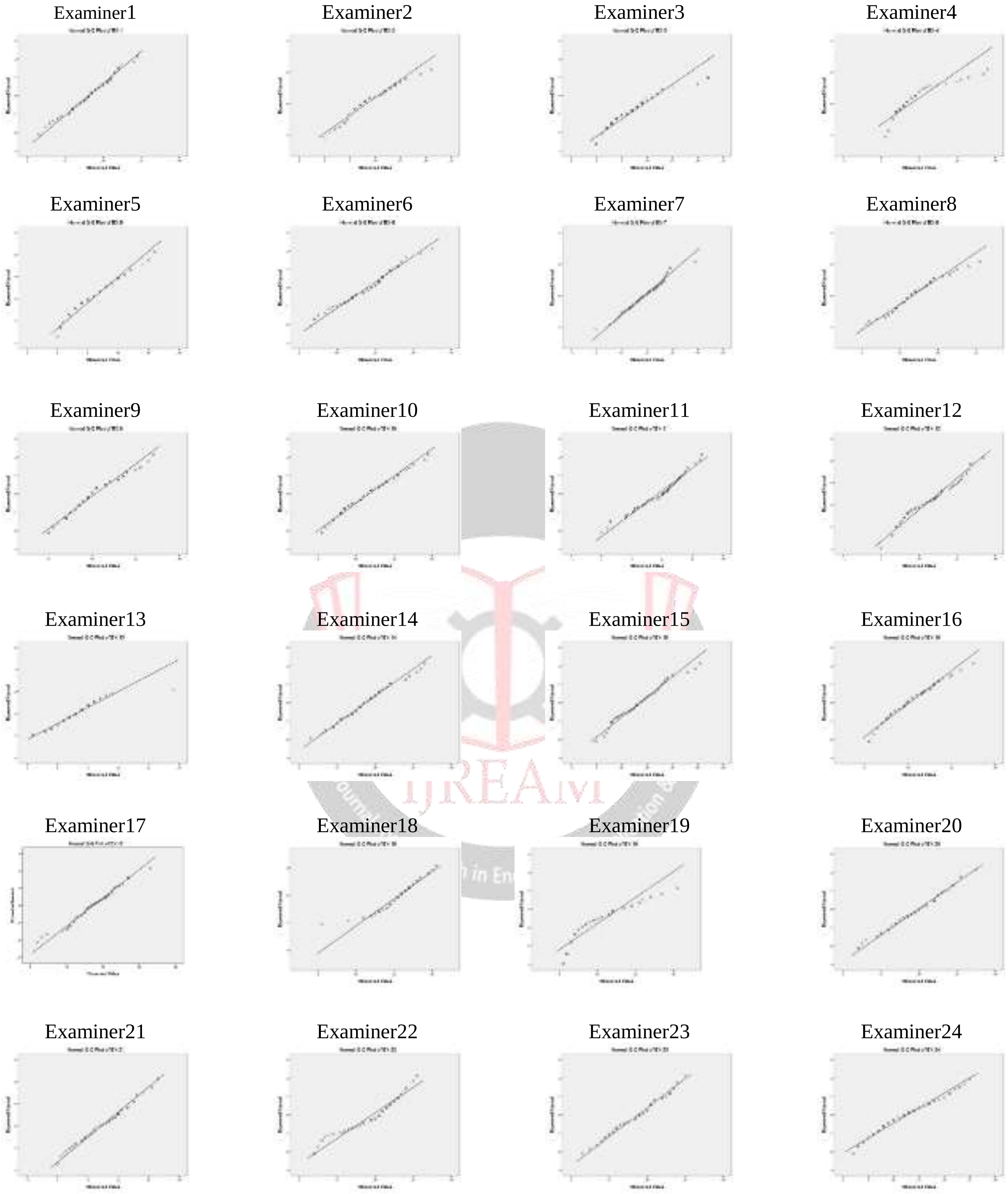
Equation	R^2	Sig
Linear	.955	.000
Quadratic	.956	.000
Cubic	**.988**	**.000**
Compound	.801	.000
Growth	.801	.000
Exponential	.801	.000

Note:
1. Dependent Variable: marks
2. The independent variable is percentile.
3. The independent variable (percentile) contains non-positive values. The minimum value is .00. The Logarithmic and Power models cannot be calculated.
4. The independent variable (percentile) contains values of zero. The Inverse and S models cannot be calculated.

DOI : 10.35291/2454-9150.2020.0176

Table 3: Fitting of cubic equation for Examiner-14 to get the fitted marks

Y (Scaled Scores) = 3.792+ 0.713 x percentile rank -0.011 x percentile rank^2 + 0.0000704 x percentile rank^3



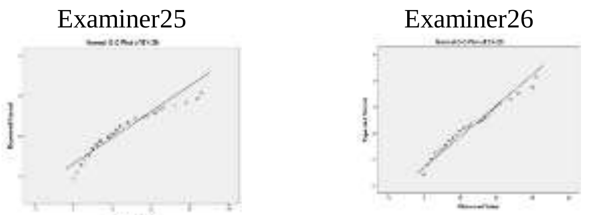


Figure 3 : Normal Q-Q plots of marks given by 26 examiners

 DOI : 10.35291/2454-9150.2020.0176

Table 4: Model of raw marks and final scale score of English scores (Only the first 10 examinees of Examiner 14 are depicted)

Code of Examinees	**Raw marks**	Individual Rank	Percentile Rank	**Scaled score**	Code of Examinees	**Raw marks**	Individual Rank	Percentile Rank	**Scaled score**
EN1	**10**	273	9	**9**	EN6	**11**	265	12	**11**
EN2	**23**	118	61	**22**	EN7	**9**	280	7	**8**
EN3	**19**	178	41	**19**	EN8	**15**	239	20	**14**
EN4	**10**	273	9	**9**	EN9	**16**	220	27	**16**
EN5	**16**	220	27	**16**	EN10	**3**	297	1	**4**

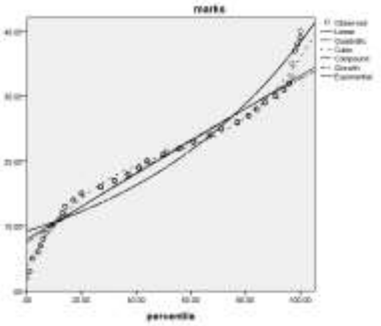


Figure 4: Checking the best fitted curve

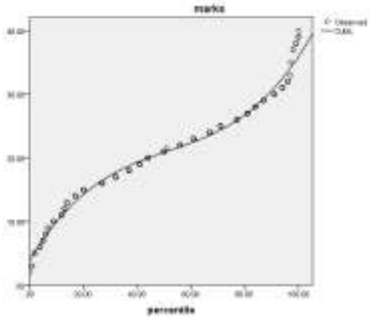


Figure 5: Cubic curve between independent and dependent variables

Table 5: Comparison of distribution based conversion of raw score to scaled score for final merit list

Sl No	Examiners Code	Candidates Code	Raw score	Scaled score	Original Rank as per Raw	Rank as per Scaled score	Sl No	Examiners Code	Candidates Code	Raw score	Scaled score	Original Rank as per Raw	Rank as per Scaled score	Sl No	Examiners Code	Candidates Code	Raw score	Scaled score	Original Rank as per Raw	Rank as per Scaled score
1	EX5	EEY27	16	36	1	1	26	EX8	EHY100	31	36	31	1	51	EX6	EFY201	35	36	31	1
2	EX2	EBY42	21	36	1	1	27	EX8	EHY278	31	36	31	1	52	EX6	EFY49	35	36	31	1
3	EX3	ECY 22	22	36	1	1	28	EX8	EHY7	31	36	31	1	53	EX8	EHY121	35	36	1	1
4	EX3	ECY45	22	36	1	1	29	EX10	EJY77	31	36	1	1	54	EX13	EMY209	35	36	1	1
5	EX12	ELY217	27	36	64	1	30	EX13	EMY239	31	36	64	1	55	EX20	ETY31	35	36	20	1
6	EX12	ELY51	27	36	64	1	31	EX18	ERY4	31	36	20	1	56	EX23	EWY164	35	36	20	1
7	EX17	EQY13	27	36	45	1	32	EX22	EVY13	31	36	20	1	57	EX24	EXY278	35	36	45	1
8	EX17	EQY145	27	36	45	1	33	EX22	EVY148	31	36	20	1	58	EX19	ESY291	36	36	20	1
9	EX17	EQY175	27	36	45	1	34	EX9	EIY154	32	36	1	1	59	EX26	EZY278	36	36	20	1
10	EX17	EQY19	27	36	45	1	35	EX13	EMY291	32	36	31	1	60	EX6	EFY164	37	36	1	1
11	EX17	EQY24	27	36	45	1	36	EX19	ESY 67	32	36	45	1	61	EX7	EGY140	39	36	64	1
12	EX17	EQY26	27	36	45	1	37	EX25	EYY156	32	36	45	1	62	EX7	EGY178	39	36	64	1
13	EX4	EDY1	28	36	1	1	38	EX25	EYY54	32	36	45	1	63	EX7	EGY53	39	36	31	1
14	EX16	EPY234	28	36	64	1	39	EX11	EKY29	32.5	36	73	1	64	EX14	ENY105	39	36	31	1
15	EX16	EPY278	28	36	64	1	40	EX11	EKY52	33	36	1	1	65	EX14	ENY141	39	36	31	1
16	EX1	EAY9	29	36	1	1	41	EX17	EQY22	33	36	20	1	66	EX14	ENY151	39	36	31	1
17	EX10	EJY50	29	36	61	1	42	EX21	EUY54	33	36	20	1	67	EX24	EXY285	39	36	20	1
18	EX12	ELY153	29	36	1	1	43	EX25	EYY14	33	36	20	1	68	EX7	EGY233	40	36	1	1
19	EX12	ELY277	29	36	1	1	44	EX26	EZY301	33	36	62	1	69	EX14	ENY246	40	36	1	1
20	EX16	EPY266	29	36	31	1	45	EX26	EZY310	33	36	62	1	70	EX15	EOY104	41	36	64	1
21	EX9	EIY151	30	36	31	1	46	EX20	ETY145	34	36	45	1	71	EX15	EOY50	41	36	64	1
22	EX9	EIY237	30	36	31	1	47	EX23	EWY144	34	36	45	1	72	EX15	EOY266	42	36	1	1
23	EX16	EPY255	30	36	1	1	48	EX23	EWY290	34	36	45	1	73	EX15	EOY281	42	36	1	1
24	EX18	ERY226	30	36	45	1	49	EX23	EWY292	34	36	45	1	TOTAL NO. OF RANK 1 IN SCALED SCORE = 73 CANDIDATES						
25	EX18	ERY36	30	36	45	1	50	EX26	EZY253	34	36	44	1							

grievances for the test takers. In this case Maximum median is 21 corresponding to Examiner-14 i.e. EN. So, all the other marks given by different examiners are transferred to the same distribution which prevailed in examiner-14. The Histograms and Normal Q-Q Plots shown in the figures Fig. 2-3 reveal the nature of the data for further analysis. From Table 2 the R^2 value is highest for the case of Cubic equation. So, Cubic equation will explain more or less 98.8 % of the variability at 1% significant level. So, it is evidently clear that for cubic equation the data fitted best. This is also supported by the following different fitted curve shown in Fig. 4 - 5. Final fitted cubic equation with respective coefficients is shown in Table 3 and this is used as the working formula for the model. Similarly all the 6194 examinees corresponding to all the 26 examiners are being

DOI : 10.35291/2454-9150.2020.0176

calculated. The comparison towards Raw scores and Scaled scores (thus obtained) are being represented by the following Box plot shown in Fig. 6 & 7 and sample tabular representation shown in Table 4. Table 5 is showing all seventy three Rank 1 candidates from different examiners with respective raw and scaled score.

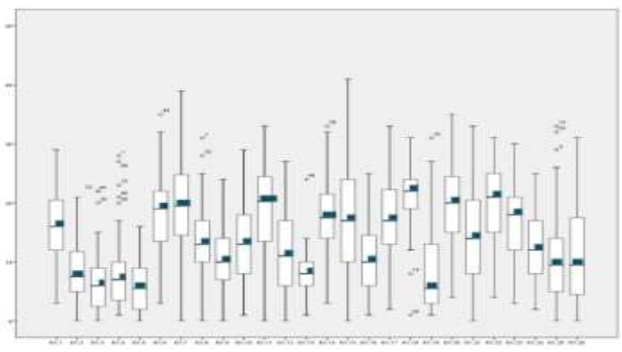


Figure 6: Box-plots of Raw scores given by 26 examiners

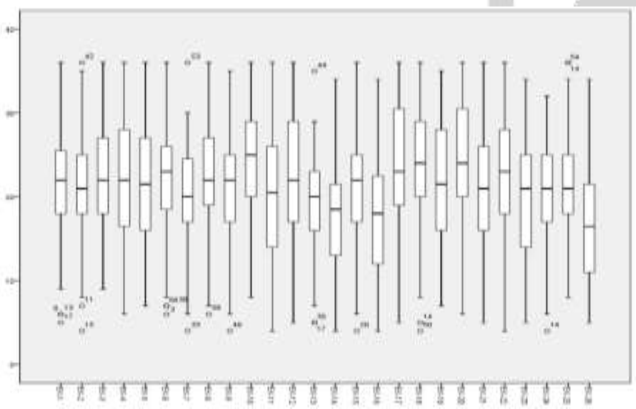


Figure 7: Box-plots of Scaled scores given by 26 examiners

V. Conclusion

For a single subject paper of a descriptive type, which is judged by several examiners, the equi-percentile method can be used for removing the examiners' bias. This is also applicable to several test papers (mainly objective in nature) of different difficulty levels. But, in case of different test papers carrying different maximum marks required for admission, recruitment or academic tests where marks are awarded for test papers on different subjects, case studies, group discussion, interview or personality tests a simple equi-percentile method would not be able to remove the examiners' bias. In all such cases, to solve the problem, the underlying distribution of marks awarded by different examiners is transferred to the distribution of the reference examiner through the process of converting raw scores to percentile scores could be adopted towards removing examiners' bias. The beauty is that these scaled scores are additive in nature, which enables one to prepare the final merit list.

Acknowledgment

West Bengal Police Recruitment Board and West Bengal University of Animal and Fishery Sciences

References

[1] Lawton M, Hu MTM, Baig F, Ruffmann C, Barron E, Swallow DMA, Malek N, Grosset KA, Bajaj NP, Barker RA, Williams N, Burn DJ, Foltynie T, Morris HR, Wood NW, May MT, Grosset DG, Ben-Shlomo Y, Equating scores of the University of Pennsylvania Smell Identification Test and Sniffin' Sticks test in patients with Parkinson's disease, Parkinsonism and Related Disorders (2016), doi: 10.1016/j.parkreldis.2016.09.023.

[2] Lee, Eunjung. "Equating multidimensional tests under a random groups design: A comparison of various equating procedures." (2013).

[3] Livingston, S. A., & Kim, S. (2010). Random‐ groups equating with samples of 50 to 400 test takers. Journal of Educational Measurement, 47(2), 175-185.

[4] Kolen, Michel J. Brennan, Roert L, Test Equating, Scaling and Linking, Methods and Practices; Statistics for Social and Behavioral Sciences, 2004, ISBN978-1-4757-4310-4.

[5] Van Steenoven, I., Aarsland, D., Hurtig, H., Chen Plotkin, A., Duda, J. E., Rick, J., & Moberg, P. J. (2014). Conversion between mini mental state examination, montreal cognitive assessment, and dementia rating scale 2 scores in Parkinson's disease. Movement Disorders, 29(14), 1809-1815.

[6] Holland, P.W. & Dorans, N.J.(2006). Linking and Equating. In R.L. Brennan (Ed.), *Educational measurement* (4th. Ed.,pp 187-220). Westport, CT: American Council for Education/ Prager.

[7] Citation from : http://www.examscomp.com/wp-content / uploads / 2014 / 12 / Equalisation-of-scores-in-SSC-Examinations.pdf

BIOGRAPHY

Professor Somen Sahu, a Statistician and Fishery Economist by profession, is a Professor and Head of Fishery Economics & Statistics, Faculty of Fishery Sciences, West Bengal University of Animal and Fishery Sciences, Kolkata. Prof. Sahu completed B.Sc. Honours in Statistics from Ramkrishna Mission Residential College, Narendrapur (Calcutta University) in 1991, Post-graduated in Statistics from Burdwan University in 1993

(1st. class 2nd.), M.B.A. (P.G.D.P.M.) from National Institute of Personnel Management in 1995 with **Gold Medal**. He completed his Ph.D. from Jadavpur University in 2006. He was a National Scholar. His areas of interest are Bio-Statistics, Statistical Software Handling, Biomonitoring, Management Information System and Extension Education in different Agricultural fields. He introduced a new Model viz. **"Dr. Sahu's Networking Model"** which is an econometric model for optimization of production-technology (enhancement

 DOI : 10.35291/2454-9150.2020.0176

of Productivity Index) in inland fisheries which was adopted by Department of Fisheries, Government of West Bengal. He is a field based fishery research scientist and academician of repute. Another pioneer work in his credit that he has successfully introduced the concept of deep sea marine cage culture which is acclaimed by all concerned relating to fishery sciences. He got published more than 40 (Forty) research publications in different peer reviewed National and International journals , **guided and guiding more than 61 (Sixty One) [Master's (42) and Ph.D.(19)] scholars**. He had also bagged the Intellectual Property Right accreditation applicable over 177 countries on the globe, on the **"PROF. SAHU'S METHODOLOGY OF DISTRIBUTION DEPENDENT EQUALIZATION OF SCORES TO REMOVE EXAMINER'S BIAS AND/OR DIFFICULTY BIAS"** (Registration No: L-89634/2020 Dated17-02-2020) as a co-author. He is the founder Secretary of International Organization of Biological Data Handlers. He has life membership with various scientific & professional societies &organizations. At present, he is associated with different organizations in State and National level. For his active contributions in the field, he has been awarded **"LIFE TIME ACHIEVEMENT AWARD-2017"** , **"MATSYA SATHI SAMMAN-2018", "BEST SCIENTIST AWARD-2020"** during Bengal Aqua Expo (over different years) by Sri Chandranath Sinha, Hon'ble MIC, Dept. of Fisheries, Govt. of W.B. . He received the **"BEST TEACHER AWARD-2019", F. F. Sc.,** from his University. He is also associated with the following organizations in different capacity viz. **"MARINE ADVISOR", Department of Fisheries, Govt. of West Bengal, Expert Member of ICAR-CMFRI** (Central Marine Fisheries Research Institute), Govt. of India, **Expert Member of ESSO-INCOIS** (Indian National Center for Ocean Information Services), Govt. of India, **"Honorary Statistical Adviser", West Bengal Police Recruitment Board**, Govt. of West Bengal, former Editor In-Chief, The Indian Journal of Agriculture Business, Co-Principal investigator, All India Network Project on Mariculture, ICAR.

Mr. K. Hari Rajan served in the Indian Police Service for 33 years and has been holding the post of Chairperson, West Bengal Police Recruitment Board since late 2012, having retired in mid - 2018. He is a recipient of the Indian Police Medal for meritorious service and President Police Medal for distinguish service. He has had the privilege of conducting 21 recruitment drives and one departmental promotion examination. The board has been conducting recruitments for Secondary school level and Graduate level entry posts for the subordinate uniformed services under the Govt. of West Bengal. He has been taking an avid interest in fine tuning the written examination system of the police recruitment board and has also developed processes to curb malpractices in examinations. The co-author has held several posts within the police department including about seven years in the traffic department in the Kolkata Police & West Bengal Police and still retains a keen interest in traffic, transportation and urban issues.

Professor Sudhansu Sekhar Maiti, is a Professor and Head Department of Statistics, Visva Bharati University. Prof. Maity completed B.Sc. Honours in Statistics from Ramkrishna Mission Residential College, Narendrapur (Calcutta University) in 1988, Post-graduated in Statistics from Calcutta University in 1990. He completed his Ph.D. from Calcutta University in 1999. He was accredited with Young Scientists Award in Statistics, Indian Science Congress Association 1994 and CSIR-NET Fellowship in 1991. He got published more than 33 (Thirty Three) research publications in different peer reviewed National and International journals. He has life membership with various scientific & professional societies &organizations.

Professor Gautam Mahapatra well known academician in the field of Computer Science and Engineering, serving for 24 Years, since 1995-96, Founder Head of the Department of Computer Science, Asutosh College, University of Calcutta, under his leadership Undergraduate Level B.Sc. with Honours, Postgraduate Level M.Sc., UGC-Approved and Financed B. Voc. Degree and Advanced Diploma under Community

College Scheme has been started, On-line Admission, Digital Library and Student Management Systems at the College and State Level, 20KV Rooftop Solar Plant, under State Govt. Hot-Line Counseling for AIDS, Haj-Pilgrims Management System, Lab Instruments Integration System under West Bengal Pollution Control Board etc. have been started. He has also served Birla Institute of Technology, Mesra and Regional Engineering College (National Institute of Technology) Durgapur as Assistant Professor and several institutes like Jadavpur University, Kalyani University, and Vidyasagar University as Guest / Visiting Faculty. Prof. Mahapatra completed B.Sc. Honours in Physics with **Top Score and Gold Medal** from Ramkrishna Mission Residential College, Narendrapur (Calcutta University) in 1989, Post-Graduated B. Tech. and M. Phil. equivalent M. Tech. in Computer Science and Engineering from University of Calcutta respectively in 1992 and 1994, joined as JRF of Indian Statistical Institute (ISI) Kolkata. He has cleared UGC-CSIR JRF (7^{th} Rank), ISI JRF, BIT Mesra JRF in CSE (1^{st} Rank) & ECE (2^{nd} Rank) and GATE-93 (96 Percentile). He has enjoyed National Scholarship in School, Undergraduate and Post-Graduate levels. His research interests are Algorithms, Artificial Intelligence, Soft Computing, Nature Inspired Algorithms, Data Science and Big-Data Analysis, Machine Learning, Wireless Sensor Network and Internet of Things (IoT). He has published more than ten research papers in International Journals – IEEE, Springer, Elsevier and others. He is also Associate Editor of Centurion Teachers' Council Journal (ISSN: 2456-4206, Multi-lingual and Multi-disciplinary Journal) and members of different Committees, Communities and Societies.

Mr. Syamantaka Sahu, has completed his schooling from Don Bosco School, Park Circus, Kolkata and is currently pursuing his B.Sc. (H) final year in Zoology Honours from the University of Calcutta . From a very young age he has been homeschooled in Statistics and Statistical software handling. He has actively been involved in ground level fishery research and data collection. He has received his training on Time-Series-Forecasting from Earth Space Science Organisation (ESSO) - Indian National Center for Ocean Information Services (INCOIS). He has published 6 international papers in peer reviewed journals to his credit. He had also bagged the Intellectual Property Right accreditation applicable over 177 countries on the globe, on the "PROF. SAHU'S METHODOLOGY OF DISTRIBUTION DEPENDENT EQUALIZATION OF SCORES TO REMOVE EXAMINER'S BIAS AND/OR DIFFICULTY BIAS" (Registration No: L-89634/2020 Dated17-02-2020) as a co-author.

DOI : 10.35291/2454-9150.2020.0176