British Journal of Mathematics & Computer Science
13(6): 1-17, 2016, Article no.BJMCS.23213
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

A New Selection Index to Address within Course Competition and between Course Competition for Ranking Examination Scores

S. G. J. Senarathne[1*] and P. Wijekoon[1]

[1]*Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka.*

Authors' contributions

This work was carried out in collaboration between both authors. Author PW designed the study, wrote the protocol and supervised the work. Author SGJS managed the literature searches, carried out all laboratories work and performed the statistical analysis, and wrote the first draft of the manuscript. Author PW edited the manuscript. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJMCS/2016/23213
Editor(s):
(1) Mohamed Rabea Eid Said, Department of Science and Mathematics, Assiut University, Egypt.
Reviewers:
(1) Radoslaw Jedynak, Kazimierz Pulaski University of Technology and Humanities, Poland.
(2) Sidney R. Castle, National University, USA.
(3) Diana Bilkova, University of Economics, Prague, Czech Republic.
Complete Peer review History: http://sciencedomain.org/review-history/12954

Received: 20th November 2015
Accepted: 29th December 2015
Published: 12th January 2016

Original Research Article

Abstract

Different scenarios of examinations have to be handled separately when measuring the true performance of students. If examiners are required to compare the performance of different groups of students who follow different combinations of subjects in an examination, their combined raw scores have to be used. In this case the raw marks can be combined by using a linear equation called a selection index. A proper selection index should correctly address two types of competitions; namely *within course competition* and *between course competition*. Although different selection indices were introduced to address these issues in literature, these methods fail to fulfill the requirements expected from a proper selection method. The main objective of this study is to introduce a new selection index called Skewness based Common Currency Index (SCCI) which addresses both *within course competition* and *between course competitions*. The proposed method considers the relative subject effects, and these effects are identified by introducing a shape parameter to the selection index. The favorability of the proposed SCCI method is compared with three alternative selection indices. According to the statistical analysis it is found that there is a significant difference of ranks between the selected indices at 5% significance level. Further, the

*Corresponding author: E-mail: jagathuop@gmail.com, pushpaw@pdn.ac.lk;

rank differences between the ranks of the SCCI method with the ranks of true student effects show smaller deviations with compared to the rank differences of the ranks of the other three selection methods. Based on the results of Wilcoxon rank sum test, it is revealed that the ranks of SCCI method are much closer to the ranks of the student effects than other three selection methods. Also the ranks of SCCI method have the highest correlation with the ranks assigned to the true student effects. According to the overall results it was confirmed that the new SCCI method can be used as a selection index to compare performance of examinees who follow different courses in an examination.

Keywords: Within course competition; between course competition; subject effect; student effect; selection index.

1 Introduction

1.1 Background of the study

Measuring the true score or true performance of each examinee is the major goal of any examination. The raw marks of an examinee may include not only the student ability but also some impacts from the subject, and other factors which cannot be easily identified. For example, the impacts of a subject may depend on the content of the exam paper [1], and/or the syllabus of the subject.

The term 'relative difficulty of subjects' is used to describe the impact of subjects in the literature. Coe [1] published his investigation on finding the relative difficulty of subjects by using statistical and judgement methods. In his study, subject pairs analysis, common examinee linear models and latent trait models were used as statistical methods while reference tests and 'value-added' methods were used as judgement methods.

If the impact of a subject is called as the subject effect, the following conceptual frame work is valid for examination scores of students.

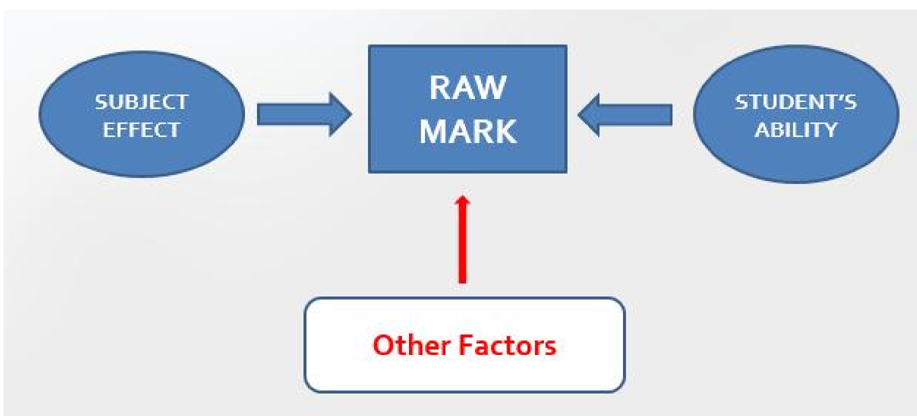

Fig. 1. Conceptual frame work

According to the above frame work, performance of student should be ranked based only on the ability of student by removing subject effect differences. Here, the other factors are considered as negligible. This removal has to be done when identifying a selection index. Therefore, the coefficients of the selection index should be selected in such a way that it minimizes the subject effect differences. If not, some subjects will get an unfavorable advantage over other subjects in the selection index. To identify the subject effects, central tendency measured were used in the literature.

Since the subject effects of raw marks are different from one subject to another subject, different scenarios of examinations have to handle separately when measuring the true performance of students. Some of these situations can be identified as follows:

(i). For the same subject, if one examination cannot be given to all students, then different examination papers have to be used to measure the performance of students.
(ii). The same group of students may sit for test papers of different subjects, and finally a combined score may be required to measure their ability.

Therefore, subject effects can be identified when comparing test scores of different examination papers of the same subject, or when comparing the test scores of different subjects separately.

Calculating subject effect is a highly critical task without having a prior knowledge about the examination paper, since subject effects are highly related to the structure and the standard level of the examination paper and a content of the subject. In order to measure the subject effect differences, different types of equating methods and test designs are available in the literature [2-4]. These equating methods are based on statistical techniques which are used to adjust scores on test forms. By carrying out this equating procedure, scores on the test papers can be made interchangeable [3], even though the test papers consist of different items (questions). When applying them in practical situations, an examiner may want to do a pilot test to select the most suitable parameters of his selected equating method, or he may directly apply a method of his choice using operational data after conducting the examination. These two practical approaches are called as pre-equating and post-equating [5], respectively. Each of these two methods has its own benefits and drawbacks.

Other than the above mentioned two situations, there are some special cases where examiners need to compare the performance of different groups of students who follow different combinations of subjects in an examination. In such a situation, not only the subject effect, but also the course effect has to be considered when comparing student ability.

When considering the course effects, a combination of raw marks of subjects plays a major role. The raw marks can be combined by using a linear equation called a selection index. When considering a proper selection index, it should correctly address two types of competitions; namely *within course competition* and *between course competition*. Competition between the students who follow the same course is defined as the *within course competition*, and the competition between students who follow different courses in the same examination is called the *between course competition*. If any selection index can correctly address the *within course competition*, then it is suitable as a selection index for *within course selection*. Similarly, if any selection index can correctly address the *between course competition*, then it is preferred in a situation where the selection is aimed for *between course selection*. Subject effects have to be considered for within course selections, and course effects have to be considered for between course selections.

Average method and Z score method are the two common selection indices used for combination of raw marks. In 2007, Wijerathne [6] has conducted an investigation to identify the anomalies when using average method as a selection index. According to his study, averaging raw marks fails to address within course competition since it does not consider the subjects effects within the course. Further, he suggested to use a scaling for location dispersion of the distribution of marks. Arivalzahan [7] and Rambukwella [8] have discussed the anomalies in the Z-score method when comparing the performance of different groups of examinees.

1.2 Selecting a proper selection index

If a particular course contains m subjects, then a selection index for a student who follow this course can be expressed as a linear combination Y of raw marks as,

$$Y = a_1X_1 + a_2X_2 + a_3X_3 + ... + a_mX_m, \tag{1}$$

where X_j represent the raw marks of a student for the j^{th} subject, and a_j represent the j^{th} subject effect for $j = 1,2,3,.....,m$. The subject effect for a particular subject is the change in Y with respect to a unit change of subject raw mark.

Averaging the raw scores, and Z score method are special cases of (1). For example, when $a_1 = a_2 = ... = a_m = 1/m$ in equation 1, we have the following linear combination to average raw scores

$$Y = \frac{X_1 + X_2 + ... + X_m}{m} = \left(\frac{1}{m}\right)X_1 + \left(\frac{1}{m}\right)X_2 + ... + \left(\frac{1}{m}\right)X_m, \tag{2}$$

for a particular course. In this method, each subject has an equal subject effect ($1/m$) for all subjects on the selection index for a given course. Therefore, it does not correctly measure the performance of students in the *within course selection*. Further, a separate component is not available to measure the course effect, and hence, it is not correctly address the *between course competition* as well.

In the Z score method, selection index of that student is of the following form:

$$Y = \frac{Z_1 + Z_2 + ... + Z_m}{m} = \left(\frac{1}{m}\right)\left\{\left(\frac{X_1 - \overline{X}_1}{s_1}\right) + \left(\frac{X_2 - \overline{X}_2}{s_2}\right) + ... + \left(\frac{X_m - \overline{X}_m}{s_m}\right)\right\} \tag{3}$$

$$= \left(\frac{1}{ms_1}\right)X_1 + \left(\frac{1}{ms_2}\right)X_2 + ... + \left(\frac{1}{ms_m}\right)X_m - K \tag{4}$$

where $\overline{X}_j$ and s_j are the mean and the standard deviation of the j^{th} subject, respectively and $K = \left(\frac{\overline{X}_1}{ms_1}\right) + \left(\frac{\overline{X}_2}{ms_2}\right) + ... + \left(\frac{\overline{X}_m}{ms_m}\right)$ is a constant for a particular course. In this case, $a_1 = 1/ms_1,\ \ a_2 = 1/ms_2,\ \ ...\ \ a_m = 1/ms_m$, and K represents the course effect. In this method, each subject has a different subject effect ($1/ms_j$) on the selection index for a given course. Therefore, this method provides more flexibility to differentiate the performance of students in each subject than the averaging method. However, in the Z score method within a given course the subjects with low variability have higher impact on the final combined Z score, which is Y in equation 4. Since the variability of the raw scores of subjects may play an important role, some anomalies can be identified in the Z score method in the *within course selection*. Further, the constant K can be used to identify the *between course completion.* Since the constant K contains the standard deviation values, similar problem as described in the *within course selection* may occur in the *between course selection*.

Z score method is the currently use method for selection of students to universities in Sri Lanka. The examination is called the General Certificate of Examination (G.C.E.) Advanced Level (A/L) which is a state examination. In this examination, students sit for three subjects depending on the course that they follow. For each student, the Z score for each subject is calculated, and then combined the three Z scores linearly by using equal weights. However, later some anomalies in selection of students were identified [7-8] when selecting students who followed a different combination of subjects to select to the same program offering in the universities.

Therefore, Yatapana and Sooriarachchi [9] proposed an alternative method called Common Currency Index (CCI) to compare the results. In this method, they have used currency conversion technique to convert raw marks in to combine score marks. This method can also be expressed as a linear combination of raw marks.

Suppose an i^{th} student follows the k^{th} course having m number of subjects. Then according to Yatapana and Sooriarachchi [9] the CCI can be written as follows:

$$Y_i = \left(\frac{\tau}{m\tau_k}\right)X_{i1} + \left(\frac{\tau}{m\tau_k}\right)X_{i2} + ... + \left(\frac{\tau}{m\tau_k}\right)X_{im} = \left(\frac{\tau}{\tau_k}\right)\left(\frac{1}{m}\sum_{j=1}^{m} X_{ij}\right) = \left(\frac{\tau}{\tau_k}\right)\overline{X}_i \tag{5}$$

where X_{i1} , X_{i2} …., X_{im} are the 1^{st} , 2^{nd} ,…m^{th} subject marks of an i^{th} student, $\overline{X}_i$ is the row mean marks of all subjects that i^{th} student has taken in a given course, and

$$\tau_k = \left(\frac{1}{n}\sum_{i=1}^{n} \overline{X}_i\right) \quad or \quad \tau_k = median\left(\overline{X}_i \; ; i =,2,..., n\right) \tag{6}$$

for n number of students. This τ_k is considered as the course effect for the k^{th} course. In a given examination if there are p number of courses, then τ is defined as $\tau = max(\tau_1, \tau_2, ..., \tau_p)$, which is the highest course effect in a given examination. According to equation 5, in this method also, the subject effects $\left(\tau / m\tau_k\right)$ are similar for all subjects within a given course, and therefore the CCI method does not correctly measure the performance of students in the *within course selection*. Hence, anomalies exist in the average method also valid for the CCI method in the within course selection.

However, in this method, a relative subject effect is considered, since the coefficient $\left(\tau / m\tau_k\right)$ includes τ , and it depends on course effects of other courses. Therefore this method can be used to compare students who take different courses having different subject combinations, and it is more favorable with compared to average and Z score method in the *between course selection*.

Since some anomalies exist in average, Z score and CCI methods, the main objective of this paper is to introduce a new selection index which will properly address both the *within course competition* and *between course competition*.

The rest of the paper is organized as follows. In section 2 we introduce a new selection index called Skewness based Common Currency Index (SCCI), which is also based on currency conversion technique. The proposed selection index is compared with the average method, Z score method and CCI method using a simulation study in section 3. Finally some concluding remarks are given in section 4.

2 Methodology

It is a known fact that if anyone wants to compare two currencies, both currencies should convert to a common currency, or else it can be converted currency type one to the currency type two directly by using currency indices. Then the currency type two acts as the common currency.

As an example, if anyone wants to compare \$100 with Rs.100, he can convert \$100 into rupees, or he can convert Rs.100 into dollars. Other than that he can convert both currencies to a different currency (let's say Pound). This concept was used in CCI method, and the same method will be used in this proposed method.

As an example, suppose a person have 100 US dollars (\$), 80 Japanese yen (¥), and 60 Sri Lankan rupees (Rs), and he wants to calculate his total wealth in Sri Lankan rupees by considering the common currency as euro (€). Let us take the three currency indices as

1\$ = X €, 1 ¥ = Y € and Rs.1 = Z €.

Then the total wealth of that person is calculated as follows:

$$\text{Total wealth} = \left[\left(\frac{X}{Z}\right)*100\right]+\left[\left(\frac{Y}{Z}\right)*80\right]+\left[\left(\frac{Z}{Z}\right)*60\right]$$

$$=\left[\left(\frac{1/Z}{1/X}\right)*100\right]+\left[\left(\frac{1/Z}{1/Y}\right)*80\right]+\left[\left(\frac{1/Z}{1/Z}\right)*60\right]$$

$$=\left[\left(\frac{Q}{q_1}\right)*100\right]+\left[\left(\frac{Q}{q_2}\right)*80\right]+\left[\left(\frac{Q}{q_3}\right)*60\right]$$

where $Q=\frac{1}{Z}$, $q_1=\frac{1}{X}$, $q_2=\frac{1}{Y}$ and $q_3=\frac{1}{Z}$

Similar type of equation will be used in the SCCI method to combine subject raw marks by considering different subjects as different currencies. Since different subjects are assigned to different currencies, it is clear that marks of one subject can differ from the marks of any other subject. To calculate the combined marks the same coefficient is used (equation 5) in CCI method, and this implies all subjects in a given course have the same currency. Also to find these coefficients mean or median marks of average course mark of each student was used, but distribution patterns were not considered.

The main differences in CCI method and the proposed SCCI method are that the subjects in a given course are considered separately as different currencies, and the distribution patterns of raw marks of each subject is incorporated to the index by adding skewness of the distributional pattern.

2.1 Skewness based common currency index (SCCI) method

Let us assume that ℓ number of subjects are offered for p courses (streams) in an examination, and m_k number of subjects are combined to create the k^{th} course.

In this method skewness of the raw marks of subjects are used to incorporate the distribution patterns (shapes) to the selection index. The shape of raw marks of examinees can be negatively skewed, positively skewed or symmetrical, and the shape parameters are based on the skewness of raw marks.

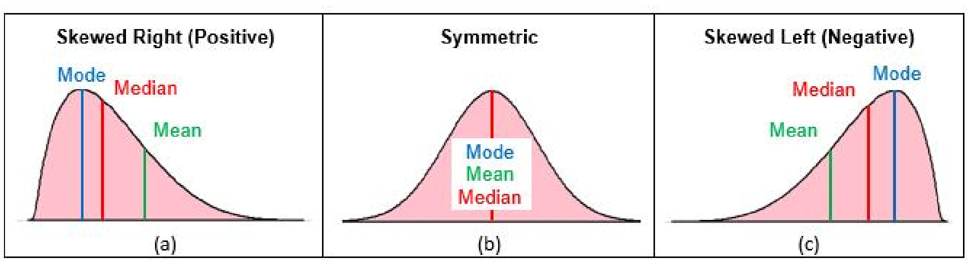

Fig. 2. Distribution shapes of raw marks

Different equations based on different statistics are available to calculate skewness of any random variable [10-12]. However, in generally the Pearson's moment coefficient of skewness [13] is used to find the skewness, which is the third standardized moment of a given random variable, and the same statistic was used to calculate skewness of subject raw marks in this study.

Suppose $\overline{X}_j$ represent the mean of the raw marks of j^{th} subject. The Pearson's moment coefficient of skewness of the raw marks of j^{th} subject can be calculated as

$$CS_j = \frac{\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(X_{ij} - \overline{X}_j\right)^3}{\left[\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(X_{ij} - \overline{X}_j\right)^2\right]^{\frac{3}{2}}} \quad ; i = 1,2...,n \quad and \quad j = 1,2,....,l \tag{7}$$

where X_{ij} is the raw marks of i^{th} student for j^{th} subject. The range of skewness values calculated by using the above expression is on the scale -3 to +3.

Since the skewness is used to identify the subject effects in the selection index, it should be converted in to a positive scale. For this conversion the following method can be used in general. For example, if we want to convert Y measurements having Y_1 and Y_3 are as the minimum and maximum values (Appendix: Fig. 6) respectively, to Z measurements having Z_1 and Z_3 as the minimum and maximum values, the following formula [14] can be used.

$$\left(\frac{Z - Z_1}{Z_3 - Z_1}\right) = \left(\frac{Y - Y_1}{Y_3 - Y_1}\right) \tag{8}$$

This implies,

$$Z = \left(\frac{Y - Y_1}{Y_3 - Y_1}\right)\left(Z_3 - Z_1\right) + Z_1 \tag{9}$$

Note that the subjects which have negatively distributed shape of raw marks (Fig. 2a) may have higher subject effects, and the subjects which have positively distributed shape of raw marks may have lower subject effects. Therefore, negative skewness values may imply higher subject effects, and positive skewness values may imply lower subject effects. Based on the above reasons, negative skewness values are converted to higher positive values than the positive skewness values. In this research we selected the converted range for skewness values as 1 to 2 to represent subject effects q_j, and this can be changed according to the convenience of the examiner.

Now to convert skewness values to range 1 to 2 use $Z_1 = 2, Z_3 = 1, Z = q_j, Y_1 = -3, Y_3 = +3$ and $Y = skewness(X_j)$ in equation 9 to obtain the following formula:

$$q_j = \frac{1}{6}\left(9 - CS_j\right) \quad ; j = 1,2,...,l \tag{10}$$

In the proposed SCCI method, relative subject effect Q is incorporated into the selection index, which is taken as the highest subject effect. Therefore,

$$Q = Max\left(q_j\right) \quad ; j = 1,2,...,l \tag{11}$$

is obtained by considering the q_j of all subjects conducted in a given examination. By incorporating Q, the subject effect of each subject is relatively measured with respect to Q which represents the subject effect of the easiest subject in a given examination. Now, the proposed SCCI index based on the common currency method for the k^{th} course is defined as,

$$Y_k = \left(\frac{Q}{m_k q_1}\right)\left(X_1 - \overline{X}_1\right) + \left(\frac{Q}{m_k q_2}\right)\left(X_2 - \overline{X}_2\right) + \ldots + \left(\frac{Q}{m_k q_{m_k}}\right)\left(X_{m_k} - \overline{X}_{m_k}\right) \tag{12}$$

where m_k is the number of subjects of the k^{th} course.

Now we rewrite equation 12 as

$$Y_k = \left(\frac{1}{m_k}\right)\left\{\left(\frac{X_1 - \overline{X}_1}{q_1 / Q}\right) + \left(\frac{X_2 - \overline{X}_2}{q_2 / Q}\right) + \ldots + \left(\frac{X_{m_k} - \overline{X}_{m_k}}{q_{m_k} / Q}\right)\right\}. \tag{13}$$

Then comparing equations 3 and 13 we can easily identify the similar formats having the Z score method and this method. Therefore, now we define a new score called Shape score or S score as $\dfrac{X_j - \overline{X}_j}{q_j / Q}$; $j = 1,2,\ldots l$ for each subject.

Equation 12 can also be written as

$$Y_k == \frac{1}{m_k}\sum_{j=1}^{m_k}\left(\frac{Q}{q_j}\right)\left(X_j - \overline{X}_j\right) = \frac{1}{m_k}\sum_{j=1}^{m_k}\left(\frac{Q}{q_j}\right)X_j - A_K \tag{14}$$

where

$A_k = \left(\dfrac{Q}{m_k q_1}\right)\overline{X}_1 + \left(\dfrac{Q}{m_k q_2}\right)\overline{X}_2 + \ldots\ldots + \left(\dfrac{Q}{m_k q_{m_k}}\right)\overline{X}_{m_k}$ is a constant for a particular course.

Note that the subject effect of j^{th} subject is measured by the quantity $\left(\dfrac{Q}{m_k q_j}\right)$ which is a relative measure with respect to all subjects conducted in an examination. Since, each subject has a different subject effect $\left(\dfrac{Q}{m_k q_j}\right)$ on the selection index for a given course, this method is more favorable with compared to the average, Z score and CCI methods for *within course selection*. Further, the constant A_K can be used to identify the *between course completion,* and hence it is more favorable for the *between course selection* as well.

2.2 SCCI method with scaled scores

From the SCCI method, a single score for each student can be obtained by using the subject marks in the examination as described in section 2.1. However, the theoretical range of these SCCI scores can be any value in between minus infinity to plus infinity which are harder to interpret. Therefore, if we can covert these scores to a meaningful range it will be convenient to the examiners. This conversion can be done according to the preference of the examiners. For example, if an examiner wishes to convert SCCI scores to the similar range as Grade Point Average (GPA), which is 0 to +4, then use $Z_1 = 0$ and $Z_3 = +4,$ in equation 9 to obtain the following formula:

$$Z = \left[4 * \left(\frac{Y - Y_1}{Y_3 - Y_1} \right) \right], \tag{15}$$

where Z is the scaled score, Y is the SCCI score, Y_1 and Y_3 are the minimum and maximum values of the SCCI score (After combining all groups).

Similarly, if an examiner prefer to use the range as -4 to +4, he can apply $Z_1 = -4$ and $Z_3 = +4,$ in equation 9 to obtain the following formula

$$Z = \left[8 * \left(\frac{Y - Y_1}{Y_3 - Y_1} \right) \right] - 4 \tag{16}$$

to convert SCCI scores. In this research equation 16 is used as the conversion formula.

2.3 Comparison of selection methods

For example, let us assume that ℓ numbers of subjects are offered for p courses in an examination, where each course is constructed by combining three subjects as the G.C.E. (A/L) examination in Sri Lanka. Then the Skewness based Common Currency Index in this situation can be described as below.

$$Y = \left(\frac{Q}{3q_1} \right) X_1 + \left(\frac{Q}{3q_2} \right) X_2 + \left(\frac{Q}{3q_3} \right) X_3 - A_K \tag{17}$$

where

$$A_k = \left(\frac{Q}{3q_1} \right) \overline{X}_1 + \left(\frac{Q}{3q_2} \right) \overline{X}_2 + \left(\frac{Q}{3q_3} \right) \overline{X}_3 \ .$$

To prove the favorability of SCCI method over average, Z score and CCI methods, simulated data sets were used. By considering a one subject at a time, the raw scores of i^{th} student is simulated using the following mixed effect model:

$$X_i = \left(\beta_j * Ab_i \right) + \varepsilon_i \qquad ; i = 1,2,...,n \ \text{ and } \ j = 1,2,...,l \tag{18}$$

where Ab_i is the student ability of the i^{th} student, β_j is a fixed value for each subject which measures the subject effect, and ε_i is the error term which consists of the negligible other factors as shown in Fig. 1. Here the student ability Ab_i and subject effect β_j are taken as a multiplicative effect. Note that Yatapana and Sooriarachchi [9] also used a mixed effect model having subject effect and student ability as additive effects. Treating a subject effect β_j as a multiplicative effect is the most suitable way to generate subject raw marks, since it is a known fact that the effect of a subject is differently affected for the students with different levels of ability while the value of subject effect is the same for all students.

Since distribution shape of raw marks is considered for the proposed method, distributions of student abilities were generated by using both skewed and symmetrical probability distributions. The parameters of these distributions and the distribution of error terms were selected in a way such that the raw scores are in the range of 0 to 100. As probability distributions (Ab_i), normal, chi-square and beta distributions were used, and random errors (ε_i) were generated by using normal distribution. Subject effects (β_j) were generated from the uniform distribution such that the subjects which have negatively distributed raw marks have high subject effects with compared to the other subjects.

Then raw marks were calculated using equation 18, and these marks were used to apply average, Z score, CCI and SCCI methods. After applying each of these methods, the resulting scores were ranked separately. However, the above generated raw scores include the subject effect, the student abilities and other factors. Note that the true student effect of a subject can be define by adding only the student ability (Ab_i) and other factors (ε_i) for each student.

For a given student, overall student effect of a course having a combination of subjects which he followed can be calculated by adding the individual student effects of each subject. When adding the student effects of each subject it should be mean or median corrected. Otherwise the student effects of some subjects within a course have higher impact on the overall student effect values. Hence, in this study, the median corrected student effects were added to find the overall student effects. Finally, the overall student effects were ranked, and those were compared with ranks of each selection method obtained by using the generated scores.

In the comparison, the method which has ranks closer to the overall student effect ranks can be identified as the best selection method. Therefore, in this study the rank differences (rank given by each selection method – overall student effect rank) were calculated, and these values were used for comparisons.

To make comparisons, the box plots were drawn for the rank differences, and the Spearmen's rank correlations were calculated between the ranks of student effects and the ranks of four scaling methods. To identify whether the absolute rank differences of at least one method is significantly differ from the other methods, Friedman test was used. Then the pair wise comparison was done by using Wilcoxon Rank Sum test. Further, to compare the variability of rank differences of each four methods Ansari Bradley test [15] was used.

3 Results and Discussion

3.1 The simulation study

The distributions of student abilities of six groups of students who follow six different subjects were generated using the given distributions in Table 1 in the usual notations.

In this study three different cases were considered.

- Case 1: For a given course, student effect ranks were compared with the ranks of each selection method. Therefore, a single course with a combination of three subjects was considered, and to generate marks of 40000 students, the first three probability distributions given in Table 1 were used. This is an example for within course selection.
- Case 2: Student effect ranks of two groups of students who follow two different courses were compared with the ranks of each selection method. Here, two subjects are taken as common for the two courses, and only one subject is different. Raw marks of two common subjects were generated by using the first two probability distributions given in Table 1. For the marks of the third subject, the third and fourth probability distributions were used for the first course and second course respectively. The sample sizes for the two courses were selected as 25000 and 15000, respectively. This is an example for both within and between course selections.

Case 3: Student effect ranks of two groups of students who follow two completely different courses were compared with the ranks of each selection method. The first three probability distributions were used to generate raw marks of three subjects of the first course, and the last three probability distributions were used to generate raw marks of the three subjects of the second course. The sample sizes for the two courses were selected as 25000 and 15000, respectively. This is also an example for both within and between course selections.

Table 1. Distributions of student ability, subject effect, and random error

Subject no.	Student ability	Subject effect	Error
1	N(μ_0, σ_0^2)	$\beta_1 \in$ U($0,1$)	N(μ_1, σ_1^2)
2	$\chi^2(p_1)$	$\beta_2 \in$ U($0,1$)	N(μ_2, σ_2^2)
3	$100-\chi^2(p_2)$	$\beta_3 \in$ U($0,1$)	N(μ_3, σ_3^2)
4	U(a,b)	$\beta_4 \in$ U($0,1$)	N(μ_4, σ_4^2)
5	A*Beta(α_7, β_7) ; $\alpha_7 < \beta_7$	$\beta_5 \in$ U($0,1$)	N(μ_5, σ_5^2)
6	C*Beta(α_8, β_8) ; $\alpha_8 > \beta_8$	$\beta_6 \in$ U($0,1$)	N(μ_6, σ_6^2)

The comparison of selection indices was done by conducting 1000 different simulations. For each simulation different values for the parameters was taken to identify whether the results depend on the selected parameter values. Model parameters of the distributions of student ability were selected in such a way that 90% of raw marks represent student ability, and the rest (10%) represents the distribution of errors. Since it was noted that the same results were given for all comparisons, only one simulation study having the following parameter values were used in the below discussion.

$$\mu_0 = 45.00, \mu_1 = 4.89, \mu_2 = 4.10, \mu_3 = 4.83, \mu_4 = 6.90, \mu_5 = 5.91, \mu_6 = 5.67, \sigma_0^2 = 136.11,$$
$$\sigma_1^2 = 0.33, \sigma_2^2 = 0.34, \sigma_3^2 = 0.43, \sigma_4^2 = 1.86, \sigma_5^2 = 2.69, \sigma_6^2 = 2.42, \beta_1 = 0.95, \beta_2 = 0.88,$$
$$\beta_3 = 0.98, \beta_4 = 0.96, \beta_5 = 0.90, \beta_6 = 0.99, \beta_7 = 6.89, \beta_8 = 3.52, p_1 = 40,$$
$$p_2 = 42, a = 9, b = 82, \alpha_7 = 3.80, \alpha_8 = 7.32, A = 90, B = 70.$$

For statistical analysis R and R-Studio were used.

3.2 Comparison of selection methods by using graphical techniques

To compare the favorability of each selection method, boxplots were drawn for the rank differences of each case stated above.

According to Fig. 3, it is clear that the rank differences related to the SCCI method have a smaller deviation with compared to the rank differences of the other selection methods for case 1. Therefore, the ranks of the SCCI method are very close to the actual ranks of student effects with compared to the other methods. It is also clear that the rank differences of average and CCI methods have a similar ranking distribution in this case. However, the ranks of Z score method have a higher deviation from the ranks of student effects with compared to the other three methods.

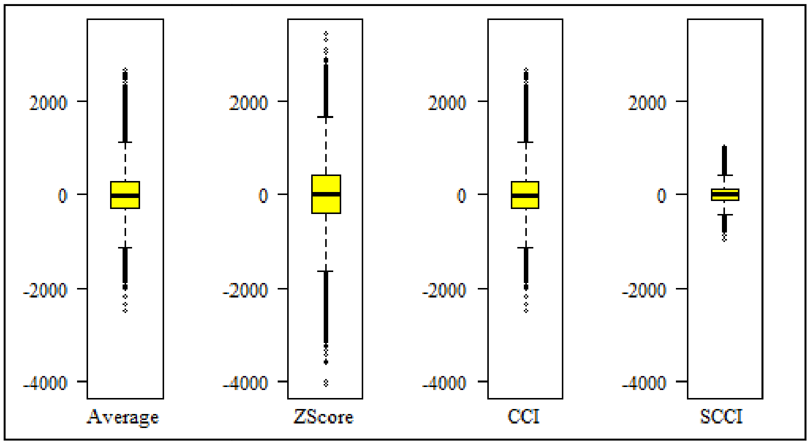

Fig. 3. Comparison of the rank differences of four selection methods for case 1

Therefore, SCCI method was identified as the best method while Z score method was the least efficient method, for *within course selection*. Further, both CCI and average methods act as the similar way in the *within course selection.*

Box plots drawn for case 2 (Fig. 4) also imply that the rank differences related to the SCCI method have smaller deviation with compared to the rank differences of the other selection methods. Further, the rank differences of CCI methods have a smaller deviation with compared to the rank differences of Z score and Average methods, and the ranks of average method have the highest deviation from the ranks of student effects with compared to the other three methods.

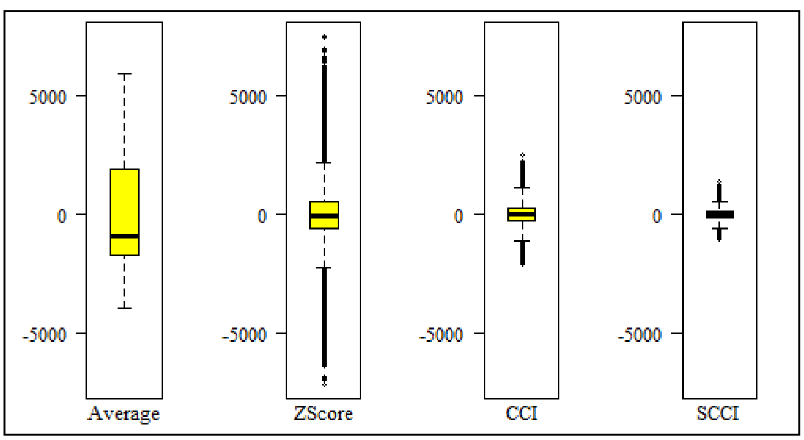

Fig. 4. Comparison of the rank differences of four selection methods for case 2

The box plots drawn for case 3 (Fig. 5) also indicate that the rank differences of the SCCI method have smaller deviation with compared to the rank differences of the other three selection methods. Also, CCI method performed well with compared to the Z score and Average methods as in the previous two cases.

Therefore, according to the above box plots (Figs. 4, 5) it was clear that SCCI method was the best method while CCI method was the second best method for both *within course* and *between course selections.* Further, both average and Z score methods were identified as the least efficient methods in both *within course* and *between course selections.*

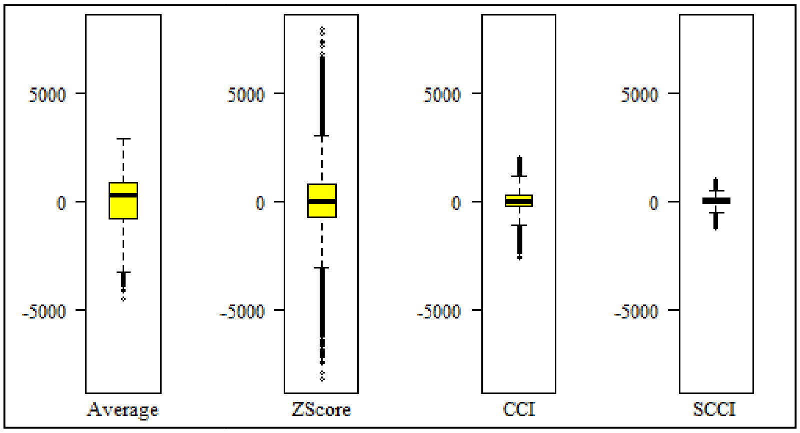

Fig. 5. Comparison of the rank differences of four selection methods for case 3

3.3 Comparison of selection methods by using statistical techniques

Standard statistical techniques were used to compare the four selection methods, and to find the favorability of SCCI method over the other three selection methods. All three cases discussed above have been considered in this section by using the same set of simulated data.

First, ranks of each method were compared with the student effect ranks by using the Spearman Rank correlation test at 5% significance level based on the following hypothesis test.

$$H_0: \rho = 0 \qquad \text{vs} \qquad H_1: \rho > 0$$

where ρ is the population correlation coefficient.

Table 2. The Spearmen's rank correlation test for comparison of rank differences

	Method	Sample correlation	*P*-value
Case 1	Average	0.9990	.000
	Z score	0.9978	.000
	CCI	0.9990	.000
	SCCI	**0.9998**	.000
Case 2	Average	0.9827	.000
	Z score	0.9935	.000
	CCI	0.9991	.000
	SCCI	**0.9997**	.000
Case 3	Average	0.9951	.000
	Z score	0.9921	.000
	CCI	0.9990	.000
	SCCI	**0.9998**	.000

In all three cases, ranks obtained by using average, Z score, CCI and SCCI have a high positive correlation (>0.9) with the ranks assigned to student effect. However, the ranks obtained by using SCCI method show the highest sample correlation with the ranks of student effect.

The absolute rank differences of four selection methods were compared using the Friedman test at 5% significance level based on the following hypothesis test.

H_0: All selection methods are equal vs
H_1: At least one selection method is different from the other selection methods

Table 3. Test statistics for Friedman test

Case	Test statistics	Degrees of freedom	*P*-value
1	68058.90	3	.000
2	82234.65	3	.000
3	67650.84	3	.000

According to the Friedman test (Table 3), it could be identified that the absolute rank difference of at least one selection method is significantly different from the other three methods at 5% significance level for all three cases.

Therefore, the pairwise comparisons were done using the Wilcoxon Rank Sum test, and the statistical hypothesis can be stated as follows:

H_0: Treatment 1 is greater than or equal to the treatment 2 vs
H_1: Treatment 1 is less than the treatment 2

Table 4. Test statistics for Wilcoxon rank sum test

Case	Treatment 1	Treatment 2	Test statistic	*P*-value
Case 1	SCCI	Average	440465575	.000
	SCCI	Z score	327889471	.000
	SCCI	CCI	440465575	.000
	CCI	Z score	961122201	.000
	CCI	Average	8e+08	.500
	Average	Z score	961122201	.000
Case 2	SCCI	Average	42277707	.000
	SCCI	Z score	274698497	.000
	SCCI	CCI	513023322	.000
	CCI	Z score	496197581	.000
	CCI	Average	101897757	.000
	Z score	Average	321781254	.000
Case 3	SCCI	Average	151979993	.000
	SCCI	Z score	175958137	.000
	SCCI	CCI	477826436	.000
	CCI	Z score	382331870	.000
	CCI	Average	345527580	.000
	Z score	Average	786177157	.000

According to the Wilcoxon rank sum test (Table 4), it is clear that the absolute rank differences of SCCI method are significantly less with compared to the other three methods for all three cases. This implies that the median rank of the SCCI method is closer to the median rank of the student effects with compared to the other selection methods. Further, compared to the Average and Z score methods, the CCI method performs well in all three cases. However, in case1 there is no significant difference between the CCI method and Average method. When comparing Average and Z score methods, Average method performs well in case1 while Z score method performs well in the other two cases. These results were also clearly shown in Fig. 3.

If the absolute rank difference of the proposed SCCI method has the lower dispersion than the other methods which indicates that the SCCI method is more suitable to use as a selection method. Therefore, the variances of the absolute rank differences of each selection method were compared by using Ansari Bradley test by using the following hypothesis test.

H_0: The variance of treatment 1 is greater than or equal to the variance of treatment 2 vs
H_1: The variance of treatment 1 is less than the variance of treatment 2

Table 5. Ansari Bradley test

Case	Treatment 1	Treatment 2	Test statistic	*P*-value
Case 1	SCCI	Average	877931229	.000
	SCCI	Z score	890250832	.000
	SCCI	CCI	877931229	.000
	CCI	Z score	840565066	.000
	CCI	Average	800020722	.499
	Average	Z score	759476090	.000
Case 2	SCCI	Average	826930392	.000
	SCCI	Z score	865615430	.000
	SCCI	CCI	857279427	.000
	CCI	Z score	846796939	.000
	CCI	Average	839016551	.000
	Average	Z score	756460918	.000
Case 3	SCCI	Average	861771592	.000
	SCCI	Z score	864259951	.000
	SCCI	CCI	869634655	.000
	CCI	Z score	853175891	.000
	CCI	Average	848389253	.000
	Average	Z score	763224665	.000

According to the results in Table 5, the absolute rank differences of the SCCI method have a lower dispersion compared to other methods for all three cases. This indicates that the ranks assigned by using the SCCI method are very close to the ranks of the student effects. Further, with compared to the Average and Z score methods, the absolute rank differences of the CCI method have a lower dispersion in the last two cases. When comparing average and Z score method, the dispersion of absolute rank differences of average method has a lower dispersion with compared to the Z score in all three cases.

4 Conclusion

In this study the sample raw scores of students were generated using different probability distributions to represent different distributional shapes. Student effects were also obtained, and used them as actual student performances. The proposed new SCCI method was compared with the other three methods; average method, Z score method and CCI method; by considering three different cases of examinations. It was noted that the proposed SCCI method is the best method among the other methods which were considered in the study to represent the student performances, since the rank differences between the ranks of the SCCI method with the ranks of true student effects show small deviations than the rank differences of the ranks of the other three methods. Also the ranks of SCCI method have the highest correlation with the ranks assigned to student effects. Further, it was noted that there is a significance difference between the rank differences of the three methods.

According to the overall results, it was confirmed that the new SCCI method can be used as a selection index to compare performance of examinees who takes combination of different test papers in an examination. Moreover, this index can be used to compare the performance of two groups of students who follow two different courses having different number of subject combinations.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Coe R, Searle J, Barmby P, Jones K, Higgins S. Relative difficulty of examinations in different subjects. Report for SCORE (Science Community Supporting Education). 2008;1-132.

[2] Song T. Investigating different item response models in equating the examination for the certificate of proficiency in English (ECPE). Spaan Fellow Working Papers in Second or Foreign Language Assessment. 2009;7:86-98.

[3] Kolen JM, Brennan LR. Test equating scaling and linking 2nd edition. Springer. 2010;1-153.

[4] Sayaka Arai, Shin-ichi Mayekawa. A comparison of equating methods and linking design for developing an item pool under item response theory. Behaviormetrika. 2011;38(1):1-16.

[5] Ye-Tong, Sz-Shyan, Ming-Xu. A comparison of pre-equating and post equating using large-scale assessment data. American Educational and Research Association Annual Conference. 2008;1-20.

[6] Wijeratne AW. Distribution of subjects Grades and Relationships among subjects of the G.C.E. (A/L) Examination. Sabaragamuwa University Journal. 2003;3(1):119-136.

[7] Arivalzahan S. A/L results and the Z score method. Newspaper Article, The Island 19th January 2011. Accessed 25 November 2014.
Available:http://www.island.lk/index.php?page_cat=article-details&page=article-details&code_title=43493#

[8] Rambukwelle WY. GCE A/L - 2011: Urgent Review of Results a Must. Newspaper Article Sunday Times 22th January 2012. Accessed 27 November 2014.
Available: http://www.sundaytimes.lk/120122/Plus/plus_03.html

[9] Yatapana PS, Sooriarachchi MR. University selection among students from different streams: A comparison method. Sri Lankan Journal of Educational Research. 2006;10(1):1-22.

[10] MacGillivray HL. Skewness and asymmetry: Measures and orderings. The Annals of Statistics. 1986;14 (3):994-1011.

[11] Tabor J. Investigating the investigative task: Testing for skewness. Journal of Statistics Education. 2010;18:1-13.

[12] Doane DP, Seward LE. Measuring skewness: A forgotten statistic. Journal of Statistics Education. 2011;19(2):1-17.

[13] Pearson K. Contributions to the mathematical theory of evolution II: Skew variation in homogeneous material. Transactions of the Royal Philosophical Society Series A. 1985;186:343-414.

[14] Ye-Tong, Pearson, Kolen MJ. Scaling: An items module. Educational Mesurement: Issues and Practice. 2010;29(4):39-48.

[15] Ansari AR, Bradely RA. Rank-sum test for dispersion. The Annals of Mathematical Statistics. 1960;31:1174-1199.

Appendix

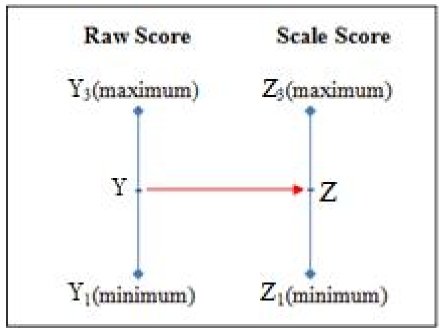

Fig. 6. Scaling raw scores

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/12954