Research on the Fairness of MPACC Selection Based on Examiner Heterogeneity

Yan Wang^{1,a}, Zhuqian He^{*2,b} (Corresponding Author), Jingjie Zheng^{3,c}

¹ School of Accounting, Guangdong University of Foreign Studies, Guangdong, China

² School of Accounting, Guangdong University of Finance and Economics, Guangdong, China

³ CNOOC deep burning energy Co., Ltd, Guangdong, China

Abstract. The selection of MPACC (Master of Professional Accountant) is a key step in the training of senior accounting personnel. This paper examines the relationship between examiner heterogeneity and MPACC second test scores. We try to clarify the reason for the unfair phenomenon because of the heterogeneity of examiners in MPACC second test results and seek ways to solve this problem. The study found that the MPACC second test results are unfair. This unfairness is caused by the heterogeneity of the examiner. However, standardized algorithms balance the differences in MPACC examiner heterogeneity. The regression model was constructed by using the MPACC second test scores before and after standardization, which verified the existence of examiner heterogeneity and the effect of the standardized algorithm on the examiner heterogeneity. This article is based on the differences of MPACC second test scores due to examiner's heterogeneity. We propose the application of standardized algorithm, which will play an important role in improving the quality of MPACC enrollment and promoting the training of senior accounting personnel.

1 Introduction

Accounting is an important part of business management, and it is a comprehensive management work for organizing capital movements and handling financial relations. Accounting is also considered to be a system that provides all economic and non-economic information on businesses, goods, services, markets, customers, etc. [1]. In recent years, the relationship between the quality of accounting information and the usefulness of decision-making has been recognized by the theoretical and practical circles [2-4]. The high-quality accounting talents with strong combination of theory and practice have gradually become the core competitiveness of enterprises. Institutions of tertiary education are the main source for delivering high-quality accounting talents to the society. The researchers and practitioners of accounting education are more willing to believe that the ability and quality should be the first in personnel training [5-7]. Since recruiting two-year accounting masters in 2004, after more than ten years of exploration and practice, China has accumulated rich research results in MPACC (Master of Professional Accountant) education. He and Li [8] compared the accounting training model of colleges and universities in China and the United States, and believed that the training of social demand-oriented accounting talents should be based on the acquisition and improvement of students' professional ability. Sheng and Zhai [9] believe that in the education of MPACC, the enrollment needs to focus on practice, teaching should

emphasize interaction, and meanwhile local education should be strengthened.

In the "new normal" economic environment, there is a large gap in application-oriented talents, and the MPACC enrollment scale continues to rise [10]. Nikolaou and Judge [11] argue that interview is one of the best-rated and most favorably appraised methods across students, and students also have a positive attitude towards psychometric test (i.e., ability, personality, honesty). The MPACC enrollment has two selection modes. One is mainly for the promotion mode for the students from their schools, and the other is for the "written test + interview" selection mode for the students from other schools. In the "written test + interview" selection mode, based on the influence of factors such as the size of the enrollment and the interview, the interview session of the MPACC selection is divided into 10-20 groups. Interview is an effective way to evaluate students' ability [12], face validity and opportunity to perform are the most important bases for considering personnel techniques favorably [13]. Previous studies have shown that examination scores are closely related to candidates' own performance and the examiner heterogeneity [14-16]. This heterogeneity is an objective phenomenon. People's gender, age, ethnicity, education level, cognition, values, preferences, attitudes, etc. are all causes of heterogeneity [17-19]. For example, Goh and Moore [14] examine the impact of "personality fitness" on academic achievement, and find that the personality dimension of introversion has the highest correlation with academic performance for the university sample. Furnham and Chamorro-Premuzic [19] explores the relationship

^ayan.wang@gdufs.edu.cn ^{*b}zhuqian_he@126.com ^czhengjj10@cnooc.com.cn

between personality, IQ, gender, beliefs about intelligence, and preference for assessment methods at university. They find that preferences are associated with individual differences, rather than academic performance.

Existing research literature deeply discusses the cultivation of accounting talents from the aspects of training mode, training scheme and teaching characteristics [20-23]. In the practice of education, the academic literature on the balance of interview scores for examiner heterogeneity is also abundant [24-27]. However, the research literature on examiner heterogeneity in the MPACC selection mechanism is still rare. In practice, the fairness of the selection mechanism of MPACC has profoundly affected the quality of accounting masters. In the large-scale differential interview process, there is a need for a method that eliminates examiner heterogeneity and promotes fairness of selection, which also has a strong operational consideration [28-29]. It provides a new research dimension for the application of high-end talents in accounting applications, in order to promote the goal of talents training in universities and the accounting profession. This paper takes the interview results of colleges and universities in Guangdong Province as the research object, and explores the connotation of examiner heterogeneity. By constructing a difference balance model of interview score, we tend to examine the relationship between examiner heterogeneity and the difference in interview scores through the OLS linear regression model. This paper seeks the theoretical basis and practical approach for the MPACC selection fairness mechanism. The main contribution of this paper is to make up for the lack of documentation on the MPACC examiner heterogeneity and the MPACC selection fairness mechanism. At the same time, in the educational practice, the existing MPACC second test selection process is provided with a scientific algorithm of score standardization to eliminate the unfairness of MPACC enrollment and selection due to examiner heterogeneity. In addition, this paper has certain significance of reference for other aspects of fairness research.

2 Theoretical analysis and research hypothesis

Human individual heterogeneity is an objective phenomenon. It is closely related to education level, cognition, values, preferences, attitudes and so on. Ramasubbu et al. [30] studied customer satisfaction with the enterprise service system. He observes customer heterogeneity based on personal value and finds that customers with different values show different satisfaction with the same enterprise service system. Pennings and Garcia [31] studied the hedging behavior of small and medium-sized enterprises. It is found that in different corporate ownership structures, members' attitudes and beliefs are heterogeneous, and the generalized mixed regression model GLM is used to test the impact of heterogeneity on hedging behavior. Li and Li [32] found that in the process of selection, due to the individual heterogeneity of the examiner or the reviewer, the selection mechanism is unfair. And that in the college entrance

examination and postgraduate entrance examination, the differences in preferences between different reviewers may make the assessment of performance in a certain subjective manner. It is of great practical significance to use the appropriate method to standardize the original absolute scores and eliminate the heterogeneity characteristics of the reviewers. Wang et al. [33] believe that the actual scoring standards are not identical because each person sees different papers. (Although it has been met before the review to discuss the unified marking criteria) Every candidate's scores are not directly comparable, so they cannot be simply synthesized and require pre-processing.

Based on the above analysis, we believe that in the current MPACC second test process, there is a problem of unfairness in the selection mechanism brought by the examiner heterogeneity. The MPACC second test method is to group the candidates in parallel according to the initial test results. Each group of candidates is interviewed by different examiners, and each group is interviewed at the same time. After the candidate's second test scores and the initial test scores are weighted averaged, the candidates are selected in meritocracy. In the interview process, due to different examiner values and personal preferences, the final scores of different groups of candidates may not be comparable. This will directly affect the admission of candidates. Furthermore, we believe that from the perspective of examiner heterogeneity, exploring how to eliminate examiner heterogeneity will promote the fairness of the MPACC selection mechanism. Therefore, this paper draws on the existing individual heterogeneity processing model to construct a standardized algorithm for MPACC second test results conversion algorithm [32-34]. This paper eliminates the influence of examiner heterogeneity characteristics in the MPACC selection mechanism. The details as follows:

Firstly, assume that when candidates enter the same school for MPACC second test, the students' abilities are at the same level. On this assumption, the candidates and examiners are randomly divided into m groups. The number of candidates in the i -th group is n_i , using i to indicate the group number in y_{ij} , and j represents the candidate number in the group, and the y_{ij} represents the original score of the j -th candidate in the i -th group. y_{ij} is a random variable and obeys the $N(\mu_i, \rho_i^2)$ distribution. $\bar{y}_i$ is the mean of the i -group MPACC score, calculated as follow:

$$s_i^2 = \frac{1}{n_i - 1} + \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (1)$$

Then we use the formula:

$$y'_{ij} = \frac{y_{ij} - \bar{y}_i}{s_i} \sim t(n_i - 1) \quad (2)$$

to convert the random variable y_{ij} to the y'_{ij} obeying the t distribution. Different groups of candidates have different degrees of freedom in the y'_{ij} obeying the t distribution, which is related to the number of people in the group.

We approximate the conversion process by applying the distribution from t to the standard normal distribution

$$x_{ij} \approx y'_{ij} + \sum_{i=1}^7 \frac{1}{n_i} B_i(t) \quad (3)$$

We convert y'_{ij} to approximate X_{ij} that follows the standard normal distribution. Finally, according to X_{ij} , the candidates' second test scores are sorted to achieve fair selection.

Table1. The variables definition of Mpacc second test score standardization algorithm conversion formula

Variables	Definition
$\bar{y}_i$	Mean of the i -th group
n_i	Total number of group i
y_{ij}	The raw score of the j -th candidate of group i
y'_{ij}	The score of the j -th candidate in the i -th group matches the score of the t -distribution
S_i^2	Sample variance of the i -th group
X_{ij}	The score of the j -th candidate of the i -th group is in accordance with the standard normal distribution

By using the above model, we will get a standardized MPACC score. MPACC original scores and standardized scores were tested for differences between groups. Based on the above theory, it is predicted that the difference between the original scores of MPACC is significant and the difference between the scores after standardization is no longer significant. This result indicates that there is an inequality in the MPACC selection mechanism, and the standardized algorithm can eliminate this inequality to some extent. Further analysis shows that this unfairness is caused by the heterogeneity of the examiner, which will significantly affect the original scores of the candidates. Based on the above analysis, this paper proposes the following assumptions:

H1: There is a significant difference in the MPACC second test scores between different groups before standardization, and there is no significant difference after standardization.

H2: After controlling other influencing factors before and after standardization of MPACC results, examiner heterogeneity was significantly correlated with MPACC original scores.

3 Research design

3.1 Sample selection and data source

Since the results of the second round of the national college MPACC will not directly announce which student's group is scored by which examiner, the complete real data is not easy to obtain. We conducted a parallel grouping of candidates according to the initial test results in a university of Guangdong Province. After that, the second test scores of each group were standardized, and the original data before standardization and the second test scores after standardization were obtained. In view of the objective existence of the differences caused by the examiner heterogeneity, we use Bootstrap to randomly generate some data that meet the requirements for large-scale empirical research based on small sample data. Based on the intangibility of heterogeneity and the

influence of examiner heterogeneity on the MPACC second test score, we select the raw data before the standardization of the MPACC second test and the statistic F of the difference test between the standardized scores. By examining the trend of F value, the effect of standardization of MPACC second test results on examiner heterogeneity was examined. Finally, based on the real small sample data, we obtained five sets of random data with a sample size of 55, for a total of 275 research samples. Selecting 55 data per group is a simplified treatment of research questions based on statistical principles. When the sample size exceeds 30, the t distribution approximates the standard normal distribution [35]. Although in the statistical principle, there is a process in which the t distribution is converted to the standard normal distribution. However, the use of statistical software such as SAS to achieve complete mathematical transformation is very cumbersome and complicated. In order to simplify this step and not interfere with the empirical results, we select 55 samples per group. Since 55 is greater than 30, after converting the original data into a t distribution, it can be directly regarded as a standard normal distribution. This paper uses SAS9.3 to empirically analyze the data.

3.2 Variable definition and model selection

In the verification study hypothesis 1, only the difference between the mean scores of the groups should be tested, and no model verification is needed. In order to gain insight into the reasons for the change in the difference between the original score and the standardized score group, also proved H2, we have the following description of variable definition and model selection:

(1) Interpreted variable: original score of MPACC second test

Based on the screening process of the same difficulty factor when the MPACC candidates who apply the same school, we believe that the level of candidates entering the second test is not much different. After the parallel grouping according to the test results of the candidates, the results of the second tests of each group are not expected to be significantly different. Therefore, when the MPACC second test scores between different groups have significant differences, we can think that the first half of H1 is verified.

(2) Explanatory variables: standardized results of MPACC second test

We standardize the original scores by using the following equation:

$$y'_{ij} = \frac{y_{ij} - \bar{y}_i}{s_i} \sim t(n_i - 1) \quad (4)$$

$$x_{ij} \approx y'_{ij} + \sum_{i=1}^7 \frac{1}{n_i} B_i(t) \quad (5)$$

We expect that the standardized scores will no longer have significant differences, at which point H1 is verified.

The specific standardized conversion algorithm is as follows:

Step 1: Import raw data.

Step 2: Verify the normality of the original data. If it does not conform to the normal distribution, it needs to be

converted to a normal distribution. If it is met, go directly to the next step.

Step 3: Test the difference between different sets of data. If there is a difference, it is necessary to carry out standardization and proceed to the fourth step; otherwise, it is considered that there is no difference between the scores of the different groups, and the hypothesis H1 is rejected.

Step 4: By using equation (4) and (5), the data is converted by the SAS program, and the standardized results are obtained.

Step 5: After the standardization, the difference test is carried out. The expected test result is not significant, and H1 is supported. The examiner heterogeneity had a significant impact on the MPACC second test scores. After eliminating the examiner heterogeneity, the MPACC second test scores tend to be fair. If the standardized scores between the groups are different, it is expected that in addition to the examiner heterogeneity, there are other factors that affect the MPACC second test score and reject H1.

Step 6: Fit the linear regression model to convert the standardized scores into 200-point system and use the scores directly, so that the research conclusions are more operable and practical.

OLS regression model selection

To validate hypothesis 2, the OLS regression model (6) was used to examine the effect of examiner heterogeneity on the fairness of MPACC selection. y_{ij} represents the original score of the MPACC second test, x_{ij} represents the standardized score of the MPACC second test, β represents the effect of the MPACC second test score normalization algorithm on the original score, and α represents the examiner heterogeneity, which is a fixed effect. We believe that after controlling the impact of the MPACC second test performance standardization algorithm on the original scores, the difference between the original score and the standardized score is brought about by the examiner heterogeneity, which is reflected by the intercept term. We expect that there will be significant in the OLS regression, that is, the examiner heterogeneity has a significant impact on the fairness of the MPACC selection mechanism.

The OLS regression model in this paper is as follows:

$$y_{ij} = \alpha + \beta x_{ij} \quad (6)$$

Table2. Variable definition of examiner heterogeneity test model

Variables	Definitions
y_{ij}	The raw score of the jth candidate of group i
x_{ij}	The score of the jth candidate of the i-th group is in accordance with the standard normal distribution

When using model (6), two aspects should be considered. First, the model usage conditions should be considered. If the conditions are not met, the application model (6) may get inaccurate results. Second, it is assumed that y_i follows a normal distribution. We do the following calculation as in equation (7):

$$y_i = \frac{x_i - \bar{x}}{s_i} \sim t(n_i - 1) \quad (7)$$

After the process, the data of the group is consistent with the t distribution, rather than the standard normal distribution. It is unreasonable to compare a set of data that conforms to the t distribution directly. The statistical principle shows that the t distribution can be approximated to the standard normal distribution only when the sample data exceeds 30. Therefore, the basic condition of applying model (6) is to ensure that the sample size exceeds 30. $x_i, \bar{x}, S_i$ are temporary variables used for problem description, x_i is a selected set of sample data, $\bar{x}$ is the mean of the set of sample data, S_i is the standard deviation of the set of sample data, y_i is converted after t distribution). On the other hand, after standardizing the data with equation (7), it is necessary to convert the MPACC second test standardized scores into two-percentage scores by equation $z_i = \theta + \lambda y_i$. Different schools can choose the λ and θ values that suit their situation. Since the selection of MPACC is independent from other schools, it only needs to be comparable within the school. Therefore, the use of standardized results is sufficient to distinguish different levels of students, to achieve the purpose of selecting outstanding students. $z_i = \theta + \lambda y_i$ is a key step in converting standardized scores to two-percentage score when publishing MPACC second test scores. y_i is the standardized score of the second test scores of the MPACC candidates, and z_i is the two-percentage score of the MPACC candidates' second test scores. Since the original score in the model (6) is 200 points, the λ and θ values can be taken as the integer part of the α and β in the model (6) OLS linear regression equation. The basic purpose of MPACC enrollment is achieved by fitting z_i to achieve reasonable use of standardized scores.

4 Empirical analysis

4.1 Descriptive statistical analysis and differential test before and after standardization of MPACC results Empirical analysis

Before the statistical analysis, the five groups of data have been tested for normality, and the five groups of data all conform to the normal distribution. The purpose of this test is to use OLS regression as follows, and the data conforms to the normal distribution as a prerequisite for OLS regression. In Table 3, TeamiF indicates the MPACC second test score before the i -th group standardization, and TeamiL indicates the MPACC second test score after the i -group standardization. From the mean and standard deviation of the data of each group of TeamiF in Table 3, the difference exists. The mean value is between 160 and 170, and the standard deviation is between 10 and 12. It can be judged that there are differences between the five groups of data. In further study of the differences between the data, the F statistic is used in this paper. The F statistic is the output of the TeamiF variance homogeneity test of the five groups of results. The P value can measure whether the mean difference between the groups is significant. It can be seen from Table 4 that the P value of

the test result of the F statistic before the standardization of the score is less than 0.01, indicating that there is a significant difference between the five sets of data. The statistical analysis of TeamIL of each group after standardization in Table 3 shows that the mean and variance of the five groups of data are all consistent. The mean value is 168, and the standard deviation is 11. The P value of the F statistic after normalization in Table 4 is greater than 0.1, which is not statistically significant, indicating that the mean values of the five groups of data are no longer different. The comparison of the results of the differential test before and after the standardization of the MPACC results supports the research hypothesis H1.

Table3. Descriptive statistics before and after standardization of results

Grouping/score	Max.	Min.	Mean	Med.	SD.
Team1F	193.00	145.00	172.26	173.00	10.79
Team1L	189.14	140.20	168.00	168.75	11.00
Team2F	190.00	136.00	167.58	169.00	12.06
Team2L	188.45	139.20	168.00	169.29	11.00
Team3F	188.00	143.00	166.09	167.00	11.23
Team3L	189.45	145.39	168.00	168.89	11.00
Team4F	195.00	143.00	172.02	173.00	11.71
Team4L	189.59	140.74	168.00	168.93	11.00
Team5F	185.00	140.00	162.96	163.00	11.25
Team5L	189.55	145.55	168.00	168.04	11.00

Note: Team*F indicates the results before standardization, and Team*L indicates the results after standardization.

Table4. F statistical test before and after standardization of results

	F Statistics
Before standardization	6.44***
After standardization	0.00

Note: p<0.01 is indicated by ***.

4.2 Correlation test between MPACC original score and standardized score

Correlation analysis showed that the normalized MPACC score was highly correlated with the original MPACC score, as R was 0.95, which was significant at the 1% significance level, consistent with theoretical analysis assumptions. The standardized results are transformed from the original results through mathematical formulas, which are the performance values after eliminating the examiner's heterogeneity. There is an inherent logical connection between them. The theoretical logical connection and the correlation analysis results in the empirical analysis.

Table5. Correlation analysis before and after standardization of achievements

Variable	X_{ij}	y_{ij}
X_{ij}	-	0.95***
y_{ij}	0.95***	-

Note: p<0.01 is indicated by ***. 4.3 Linear regression results of MPACC standardized scores and original scores.

The regression results of model (1) show that 168.18, t value is 760.47, and p value is less than 0.001, which indicates that there is an inherent factor between the

MPACC original score and the standardized post score. This factor leads to significant differences between the two, and the intrinsic factors are presented by the linear model intercept term, which is the MPACC examiner heterogeneity. The empirical results verify hypothesis H2. After controlling the other influencing factors before and after the standardization of MPACC, the examiner heterogeneity is significantly correlated with the MPACC original scores. The examiner heterogeneity makes the scores between the MPACC groups unfair. In the regression model, β is the regression coefficient of the MPACC standardized score to the MPACC original score, which is 11.41, the t value is 51.10, and the p value is less than 0.01. The regression results are consistent with the above correlation analysis and theoretical analysis. It shows that there is a significant positive correlation between the original MPACC scores and the standardized scores.

Table6. Linear regression results of MPACC standardized score sand original scores

Variables	Regression coefficient value
α	168.18*** (760.47)
β	11.41*** (51.10)

Note: p<0.01 is represented by ***. () indicates the t value.

5 Conclusion

This paper is based on the unfairness of the MPACC selection method. Through the real data construction model of a college in Guangdong Province and the empirical data, the research shows that the MPACC second test results are unfair. This unfairness is caused by the heterogeneity of the MPACC examiner. Based on the objective support of the heterogeneous objective, this paper also proves this research hypothesis with the model. This article provides an algorithm for how to eliminate the heterogeneity of examiners. This method of standardization of results will greatly benefit the improvement of enrollment quality, and it will also promote the development of senior accountants. By extension, the selection process of other disciplines may also have heterogeneity and cause unfairness problems. It is especially important to find such differences in time and adopt corresponding methods to eliminate differences. Therefore, the research results of this paper can be used in other subjects. The limitation of this paper is that there are too few research samples. If more sample data can be obtained, the research conclusions will be more scalable.

Acknowledgment

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions, and gratefully acknowledge the financial support of the Guangdong Provincial Natural Science Fund (2018A030313103); Guangdong University Key Research Projects (2019WZDXM015); Graduate Education Innovation Plan of Guangdong Province

(2019SFKC23); and the 2019 Teaching Quality and Teaching Reform Project for Higher Education of Guangdong Province: International Teaching Team for Accounting Major.

References

1. H. Duman, İ. Apak, M. Y. İcerli, and M. Yücenurşen, "The entrepreneurs' expectations and satisfaction level in today's accounting education: a survey of aksaray organized industrial zone," *Procedia-Social and Behavioral Sciences*, vol.103, pp. 1059-1066, 2013.
2. B. Lev, and P. Zarowin, "The boundaries of financial reporting and how to extend them," *Journal of Accounting Research*, vol.37, pp.353-385, 1999.
3. L. Shivakumar, "The role of financial reporting in debt contracting and in stewardship," *Accounting and Business Research*, vol.43, pp.362-383, 2013.
4. E. Leung, and D. Veenman, "Non-GAAP earnings disclosure in loss firms," *Journal of Accounting Research*, vol.56, pp.1083-1137, 2018.
5. W. S. Albrecht, and R. J. Sack, "Accounting education: charting the course through a perilous," *Accounting Education Series*, vol.16, pp.1-72, 2000.
6. B. Howieson, "Accounting practice in the new millennium: is accounting education ready to meet the challenge?," *British Accounting Review*, vol.35, pp.69-104, 2003.
7. S. C. Barrie, "A research - based approach to generic graduate attributes policy," *Higher Education Research & Development*, vol.2, pp.261-275, 2004.
8. Y. R. He, and X. H. Li, "A study on the training mode of accounting talents in China's colleges and universities - based on accounting education in ten American colleges and universities field research," *Accounting Research (in Chinese)*, vol.4, pp. 26-31, 2013.
9. M. Q. Sheng, and S. B. Zhai, "Research on innovation of master of accounting (MPACC) training program," *Friends of Accounting (Chinese)*, vol.9, pp. 120-128, 2011.
10. "Medium and long term talent development plan of accounting industry (2010-2020)," *Accounting Research (in Chinese)*, vol.11, pp. 12-15, 2010.
11. I. Nikolaou, and T. A. Judge, "Fairness reactions to personnel selection techniques in Greece: the role of core self-evaluations," *International Journal of Selection and Assessment*, vol.2, pp.206-219, 2007.
12. M. Bertolino, and D. D. Steiner, "Fairness reactions to selection methods: an Italian study," *International Journal of Selection and Assessment*, vol.2, pp.197-205, 2007.
13. S. Moscoso, and J. F. Salgado, "Fairness reactions to personnel selection techniques in Spain and Portugal," *International Journal of Selection and Assessment*, vol.1, pp.187-196, 2004.
14. D. S. Goh, and C. M. Moore, "Personality and academic achievement in three educational levels", *Psychological Reports*, vol.1, pp.71-79, 1978.
15. M. R. Barrick, J. A. Shaffer, and S. W. DeGrassi, "What you see may not be what you get: relationships among self-presentation tactics and ratings of interview and job performance", *Journal of Applied Psychology*, vol.94, 1394-1411, 2009.
16. T. Macan, "The employment interview: a review of current studies and directions for future research", *Human Resource Management Review*, vol.19, pp.203-218, 2009.
17. D. D. Steiner, and S. W. Gilliland, "Fairness reactions to personnel selection techniques in france and the united states", *Journal of Applied Psychology*, vol.2, pp.134-141, 1996.
18. M. A. Campion, D. K. Palmer, and J. E. Campion, "A review of structure in the selection interview", *Personnel Psychology*, vol.3, pp. 655-702, 1997.
19. A. Furnham, and T. Chamorro-Premuzic, "Individual differences and beliefs concerning preference for university assessment methods", *Journal of Applied Social Psychology*, vol.9, pp.1968-1994, 2005.
20. V. Grupta, S. Kumar, and V. Sriranga, "A framework for dissemination of accounting education", *Academy of Business Research Journal*, vol.3, pp.110-123, 2013.
21. L. Crawford, C. Helliar, E. Monk, and M. Veneziani, "International accounting education standards board: organizational legitimacy within the field of professional accountancy education", *Accounting Forum*, vol.38, pp.67-89, 2014.
22. B. Jackling, and R. Natoli, "Employability skills of international accounting graduates: internship providers' perspectives", *Education & Training*, vol.57, pp.757-773, 2015.
23. N. Dong, M. Bai, H. Zhang, and J. Zhang, "Approaches to learning IFRS by chinese accounting students", *Journal of Accounting Education*, vol.48, pp.1-11, 2019.
24. S. Zedeck, A. Tziner, and S. E. Middlestadt, "Interviewer validity and reliability: An individual analysis approach", *Personnel Psychology*, vol.36, pp.355-370, 1983.
25. D. S. Kiker, and S. J. Motowidlo, "Effects of rating strategy on interdimensional variance, reliability, and validity of interview ratings", *Journal of Applied Psychology*, vol.83, pp. 763-768, 1998.
26. C. H. Van Iddekinge, C. E. Sager, J. L. Burnfield, and T. S. Heffner, "The variability of criterion - related validity estimates among interviewers and interview panels ", *International Journal of Selection and Assessment*, vol.14, pp.193-205, 2006 .
27. R. D. Arvey, and J. E. Campion, "The employment interview: a summary and review of recent research", *Personnel Psychology*, vol.35, pp.281-322, 2010.
28. C. E. Lance, D. J. Woehr, and S. A. Fisicaro, "Cognitive categorization processes in performance

- evaluation: confirmatory tests of two models. *Journal of Organizational Behavior*, vol.12, pp.1-20, 1991.
29. S. K. Simola, S. Taggar, and G. W. Smith, "The employment selection interview: disparity among research-based recommendations, current practices and what matters to human rights tribunals", *Canadian Journal of Administrative Sciences*, vol.24, pp.30-44, 2007.
 30. N. Ramasubbu, S. Mithas, and M. S..Krishnan, "High tech, high touch: the effect of employee skills and customer heterogeneity on customer satisfaction with enterprise system support services", *Decision Support Systems*, vol.44, pp. 509-523, 2008.
 31. J. M. E. Pennings, and P. Garcia, "Hedging behavior in small and medium-sized enterprises: the role of unobserved heterogeneity", *Journal of Banking and Finance*, vol.5, pp. 951-978, 2004.
 32. W. B. Li, and H. B. Li, "A preliminary study on the standardization of test scores, Silicon Valley (in Chinese)", vol.4, pp. 82-96, 2011.
 33. X. R. Wang, F. L. Yuan, and Q. B. Mou, "A standardized model and algorithm for test scores", *Science and Technology Plaza (in Chinese)*, vol.10, pp. 35-37, 2012.
 34. D. Xu, H. J. Lu, and W. Qi, "Research on standardized evaluation method of public computer course performance", *Computer Education (in Chinese)*, vol.6, pp. 82-85, 2015.
 35. R. A. Johnson, *Probability and Statistics for Engineers*, Ninth Edition, Publishing House of Electronics Industry, 2017.