

Using Standardization for Fair Data Evaluation

Kyongho Cho*, Taeseon Yoon**, Sunghee Park*, and Dohyeon Park***

*Education, Gachon University, Seongnam-daero, Sujeong-gu Seongnam-si Gyeonggi-do South Korea

**Computer Science and Engineering, Korea University, Anam-ro, Seongbuk-gu Seoul South Korea

***Natural Science, Hankuk Academy of Foreign Studies, Mohyeon myeon wangsanri 522 Gyeonggi-do South Korea

hispano@hafs.hs.kr, tsyoon@hafs.hs.kr, sungheepark@gachon.ac.kr, 9parkdoh9@naver.com

Abstract— Evaluation on a specific criteria is divided into quantitative evaluation and qualitative evaluation, the former can clearly depict scores in objective manner while the latter is obscure to measure. To reduce the obscurity, we used formula of Z-score and T-score through 'standardization of normal distribution'. To testify the effect, we set a simulation of two groups of which consisted 3 interviewers and 5 interviewees. The result after using the method lead to equal mean and equal standard deviation regardless of different interviewers. Also, by using T-score, we again rearranged scores once again by using T-score, in order to enable comparing between criteria possible. By doing so, we summed up the altered scores, and arranged the ranks of interviewees from the highest to the lowest. This proved to be effective on reducing obscurity compared to raw scores' sum up.

Keywords— Evaluation, standardization, quantitative, qualitative, z-score

I. INTRODUCTION

A. Recent issue

When evaluating a subject, there are quantitative evaluation and qualitative evaluation[1]. In terms of quantitative evaluation, it does not lose fairness, because the evaluation on a subject is expressed in quantitative scores so it is objective[2][3]. However, since qualitative evaluation is universally used these days, it is difficult to satisfy every individual who is being evaluated. An example of qualitative evaluation is official interview[6]. Since it is difficult for interviewers to clearly depict interviewees' scores in objective manner according to evaluation criteria, it occurs obscurity. In addition, due to characteristics of interview, different groups are assigned to have interviews with different interviewers. Thus, it is needed to rearrange the deviation of scores given by each interviewers.

B. Objective of Research

The purpose of this research is to figure out the way to minimize the obscurity of qualitative evaluation. In order to reduce the obscurity, rearrangement of score deviation among interviewers and standardization of various criteria will be used[4][5].

C. Method of Research and Data Collection

The idea of 'standardization of normal distribution' was raised, based on the text book, and looked up on the internet about research about normalized standardization. Based on these research, this research calculates mean value, standard deviation value, and other values which are required for this project through 'spread sheet'.

II. PRIOR RESEARCH

A. Theoretical Review- interview

Because of characteristics of a interview, the fact that numerous interviewees have to do interviews at the same time, it is common to distribute interviewees into different groups rather than one group[7]. Therefore, each groups are evaluated by different interviewers, and each interviewer could make relative evaluation based on their subjective views. Even though it is evaluated by the same criteria, some interviewers could give higher scores, however; the others could give lower scores. Some interviewers could make drastic score gap among interviewees, on the contrary, the others could make little score gap among interviewees. Since mean and dispersion of a score could be different depending on which group you belong to, it is difficult to say that this kind of interview is fairly evaluated. Furthermore, criteria are not one, but several. Suppose that it is the university admission interview and it is going to be evaluated by four criteria: character 25%, impression 25%, leadership 25%, and potential 25%, which make 100% in total. Let's say that student A scores 20% in character, 20% in impression, 15% in leadership, and 10% in potential. Even though both scores of character and impression are the same in aspects of absolute score, the actual value could be different in aspects of relative evaluation because the level of difficulty and assessment function of each criteria are different. Thus, interviewees' mean and standard deviation depending on each criteria could be different, so it is necessary to transform from raw scores to standardized score using 'standardization'.

B. 'standardization of normal distribution' and transformation

While the data of following examples are few, typical interviews have far more data, so it is more close to normal distribution. It is possible to rearrange the score deviation depending on interviewers, by using mean that shows relative

position of given raw score, standard deviation that shows the measure of dispersion, and formula $Z = \frac{x - m}{\sigma}$ that is used for standardization of normal distribution.

Also, in order to compare each criteria which have different standard deviation, we are going to use T-score. By taking 50 as mean and 10 as standard deviation. By transforming the formula that results T-score, and rearranging to compare scores depending on criteria, it is possible to reduce obscurity of qualitative evaluation.

III. RESULT

Suppose that there are two interviews which consist of three interviewers and four interviewees in each room, these are the scores.

TABLE 1. RAW DATA - PARAMETER

Group1	Interview								
	A			B			C		
Criterion	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Interviewee a	25	23	25	28	25	24	20	26	20
Interviewee b	30	29	30	27	30	28	26	28	27
Interviewee c	23	25	23	25	27	24	23	28	26
Interviewee d	22	20	20	25	23	25	21	26	23
Mean	25	24.5	24.5	26.25	26.25	25.15	22.5	27	24
Standard deviation	3.162	3.908	3.873	2.301	2.550	1.673	2.915	1.141	3.162
Group2	D			E			F		
Criterion	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Interviewee e	30	29	30	29	28	30	25	23	27
Interviewee f	26	27	26	22	27	28	23	25	23
Interviewee g	27	30	27	20	29	28	23	29	25
Interviewee h	21	28	22	21	24	27	21	25	22
Mean	26	28.5	26.25	23	27	28.25	23	25.4	24.25
Standard deviation	3.606	2.301	2.915	3.742	1.936	2.828	1.936	2.301	2.301

A. Rearrangement of standardized score among interviewers

The problem with the scores above is that each interviewer have different mean and standard deviation. The fact that sum of the scores that are graded by one interviewer and the deviation are different could be illogical, so we need to make

sum of the scores that are graded by each interviewers or mean of the scores that are graded by each interviewers as same as the standard deviation to solve this problem. In order to do this, three values down below are needed.

- 1) total mean of the scores of each criteria:
- 2) value Z of the scores that are graded by interviewers to interviewees:
- 3) standard deviation of total mean:

Rearrangement value $(\bar{x})$, using the values above, $\bar{x} = m_{total} + z\sigma_{total}$ can be calculated by modifying a formula $z = \frac{\bar{x} - m_{total}}{\sigma_{total}}$

, The reason why we used total mean in number 1) is because is increasingly changing, and showing the data objectively in comparison to other numbers. in number 2) is the value that shows relative evaluation among interviewees by interviewers. Deviation of total mean() in number 3) is the value of in the formula to calculate standard deviation instead of . means if the standard deviation is higher, the scores are graded more distinguishable.

TABLE 2. VALUE Z OF THE SCORES THAT ARE GRADED BY INTERVIEWERS

Group1	Interview								
	A			B			C		
Criterion	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Interviewee a	0	-0.382	0.137	1.347	-0.483	0.737	1.091	-1.009	-1.460
Interviewee b	1.622	1.452	1.521	0.577	1.450	1.677	1.527	1.521	1.095
Interviewee c	-0.648	0.229	-0.488	-0.962	0.290	-0.702	0.218	1.182	0.730
Interviewee d	-0.973	1.300	-1.232	-0.962	1.256	-0.152	0.654	-1.065	-0.365
Standard deviation	0	0	0	0	0	0	0	0	0
Group2	D			E			F		
Criterion	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Interviewee e	1.234	0.447	1.321	1.697	0.534	1.605	1.414	-1.121	1.432
Interviewee f	0	-1.341	-0.087	0.282	-0.299	0.242	0	-0.229	0.650
Interviewee g	0.308	1.341	0.262	-0.848	1.069	-0.204	0	1.605	0.390

	67	1	1	53	5	42			7
Interviewee h	1.5 43 03 35	0.4 47 21	1.4 85 3	0.5 65 69	1.6 03 57	1.1 47 08	1.4 14 21	0.2 29 42	1.1 71 7
Standard deviation	0	0	0	0	0	0	0	0	0

Value that are calculated with modified formula

$$\bar{x} = m_{total} + z\sigma_{total}$$

TABLE 3. FONT SIZES FOR PAPERS

Criterion	1	2	3
Standard deviation of total mean(σ_{total})	3.034238	2.531743	2.878621

TABLE 4. REARRANGED SCORES BY USING RAW SCORES

Group1	Interview								
	A			B			C		
Transformed value x	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Interviewee a	24. 29 16 7	25. 44 86 3	26. 01 94 79 71	28. 37 92 4	25. 19 29	23. 43 00 7	20. 98 10 4	23. 88 49 2	21. 42 05 1
Interviewee b	29. 21 38 5	30. 09 52	29. 96 42 76 8	26. 04 34 8	30. 08 79 6	30. 45 38 4	28. 92 65 4	28. 94 84 1	28. 77 83 7
Interviewee c	22. 32 27 9	26. 99 74 9	24. 44 15 60 87	21. 37 19 7	27. 15 09 3	23. 43 00 7	24. 95 37 9	28. 94 84 1	27. 72 72 5
Interviewee d	21. 33 83 6	23. 12 53 4	22. 07 46 82 62	21. 37 19 7	23. 23 48 8	25. 18 60 1	22. 30 52 9	23. 88 49 2	24. 57 38 8
mean	24. 29 16 7	26. 41 66 7	25. 62 5	24. 29 16 7	26. 41 66 7	25. 62 5	24. 29 16 7	26. 41 66 7	25. 62 5
Group2	D			E			F		

Criterion	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Interviewee e	28. 03 72 1	27. 54 89	29. 39 75 94 15	29. 44 09 4	27. 76 99 4	30. 24 78 1	28. 58 27 3	23. 51 25 6	29. 74 74 1
Interviewee f	24. 29 16 7	23. 01 99 8	25. 37 34 93 72	23. 43 34 5	26. 41 66 7	24. 96 46	24. 29 16 7	25. 83 58 5	23. 75 11 8
Interviewee g	25. 22 80 5	29. 81 33 6	26. 37 95 18 83	21. 71 70 3	29. 12 32 1	24. 96 46	24. 29 16 7	30. 48 24 2	26. 74 92 9
Interviewee h	19. 60 97	25. 28 44 4	21. 34 93 93 3	22. 57 52 4	22. 35 68 5	22. 32 3	20. 00 06 1	25. 83 58 5	22. 25 21 2
Mean	24. 29 17	26. 41 25 66 7	25. 62 5	24. 29 24.25 16 7	26. 41 66 7	25. 24.25 5	24. 29 16 7	26. 41 26.25 66 7	25. 62 5

26.25

Because sum of the value Z_s of interviewees are 0 according to the table 2, mean of $\bar{x}$ from $\bar{x} = m_{total} + z\sigma_{total}$ is same as m_{total} , regardless of interviewers. In addition, from

$$z = \frac{\bar{x} - m_{total}}{\sigma_{total}}$$

the formula σ_{total} , standard deviation agrees with σ_{total} , newly rearranged mean and standard deviation are same regardless of each interviewers, m_{total} and σ_{total} .

B. Standardization among each criteria

By methods shown in number 1, rearranged scores $\bar{x}$ of each criteria are resulted. However, standard deviations of each criteria are different, it is impossible to decide which score is high or low absolutely. This being so, we need to rearrange the scores in order to compare each scores, so we transform T-score which is used in standardized score for university admission test.

By transforming $(50, 10^2)$ of T-score, mean and standard deviation can be replaced each other. Mean and standard deviation of $(30, 5^2)$ are applied in the example. Standard deviation in this formula is the standard deviation for $\bar{x}$ and m_{total} .

$$\sqrt{\frac{\sum_{i=1}^n (\bar{x}_i - m_{total})^2}{n}}$$

TABLE 5. STANDARD DEVIATION FOR $\bar{x}$

Criterion	1	2	3
Standard deviation	3.034238	2.531743	2.877494467

Using $\bar{\sigma}$ and applying in the formula $T = 30 + \frac{x - m_{total} * 5}{\sigma_{total}}$, which is transformed from $T = 50 + z * \sigma$, we result the final score.

TABLE 6. REARRANGED SCORES BY USING RAW SCORES

Group1	Interview								
	A			B			C		
Criterion	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Interviewee a	30	28.0882	30.68545694	36.73575	27.58316	26.18604	24.54455	25	22.69417
Interviewee b	38.11107	37.26483	37.54002632	32.88675	37.25052	38.3907	37.63763	35	35.47937
Interviewee c	26.7557	31.14708	27.94362919	25.18875	31.4501	26.18604	31.09109	35	33.65291
Interviewee d	25.13336	23.49989	23.83088756	25.18875	23.71621	29.23721	26.72673	25	28.17354
Mean	30	30	30	30	30	30	30	30	30
Group2	D			E			F		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Interviewee e	36.17213	32.23607	36.55534562	38.48528	32.67261	38.03269	37.07107	24.461	37.16319
Interviewee f	30	23.2918	29.56297696	28.58579	30	28.85247	30	28.85292	26.744

Interviewee g	31.54303	36.7082	31.31106912	25.75736	35.34522	28.85247	30	38.02955	31.9536
Interviewee h	22.28483	27.76393	22.5706083	27.17157	21.98216	24.26366	22.923	28.85292	24.13921
Mean	30	30	30	30	30	30	30	30	30

Therefore, transforming newly rearranged $\bar{x}$ again in order to compare according to each criteria, we finally have the results which are comparable interview scores.

TABLE 7. REARRANGED SCORES BY USING RAW SCORES

Criterion	(1)	(2)	(3)	Final score X
Interviewee a	30.42677	26.89045	26.52189	83.83911
Interviewee b	36.21182	36.50512	37.1367	109.8536
Interviewee c	27.67847	32.53239	29.26086	89.47173
Interviewee d	25.68295	24.07203	27.08055	76.83553
Interviewee e	37.24283	29.72443	37.25041	104.2177
Interviewee f	29.5286	27.38157	28.38648	85.29665
Interviewee g	29.10013	36.69433	30.70571	96.50017
Interviewee h	24.12845	26.19967	23.65739	73.98551

IV. CONCLUSION

It is impossible to compare with raw scores of interviewees because mean and deviation of scores that are given by interviewers are different. As it is elaborated in the part 3-1, by making mean and standard deviation of the scores by interviewers same, it is possible to compare each others' scores in the same criterion. In addition, as it is explained in the part 3-2, we solve the problem due to mean and deviation's difference from each criteria, and make it possible to compare scores from different criteria.

Suppose that the number of interviewees as n , mean of scores that are graded by interviewers as m , total mean of the raw scores from each criteria as m_{total} , following is the newly rearranged score m_{total} which can be resulted from part 3-1.

$$\bar{x} = m_{total} + z * \sigma_{total} = \frac{1}{n} \sum_{i=1}^n x_i + \frac{x - m}{\sqrt{\frac{\sum_{i=1}^n (x_i - m)^2}{n}}} * \sqrt{\frac{\sum_{i=1}^n (x_i - m_{total})^2}{n}} = \frac{1}{n} \sum_{i=1}^n x_i + (x - m) * \sqrt{\frac{\sum_{i=1}^n (x_i - m_{total})^2}{\sum_{i=1}^n (x_i - m)^2}}$$

Also, from this result $\bar{x}$, in order to make it possible to compare by each criteria, by using the method of T-score conversion, following is the X , which is the new score that can be resulted from the method (k, l) :

$$X = k + \frac{\bar{x} - m_{total} * l}{\sigma}$$

Through this process, it is possible to compare the final scores X , which are final added scores of all three criteria and evaluate the interviewees. Therefore, the final results will be listed from the highest to lowest; interviewee b, interviewee e, interviewee g, interviewee c, interviewee f, interviewee a, interviewee d, interviewee h.

REFERENCES

- [1] David Steinsaltz, "Introduction to Probability Theory and Statistics for Psychology and Quantitative Methods for Human Sciences", University of Oxford.
- [2] Minayo, Maria Cecília De Souza. "Limits and Possibilities to Combine Quantitative and Qualitative Approaches." *Qualitative versus Quantitative Research*, 2017, doi:10.5772/intechopen.68195.
- [3] LAZAR STAN KOV, ANNE CREGAN, "QUANTITATIVE AND QUALITATIVE PROPERTIES OF AN INTELLIGENCE TEST: SERIES COMPLETION", *LEARNING AND INDIVIDUAL DIFFERENCES* VOLUME 5. NUMBER 2. 1993
- [4] Larry Winner, "Applied Statistical Methods", Department of Statistics, University of Florida, February 23, 2009
- [5] Howard J. Seltman, "Experimental Design and Analysis", September 8, 2015
- [6] Guy Harling, Dumile Gumede, Tinofa Mutevedzi, Nuala McGrath, Janet Seeley, Deenan Pillay, Till W. Bärnighausen and Abraham J. Herbst, Harling et al, "The impact of self-interviews on response patterns for sensitive topics: a randomized trial of electronic delivery methods for a sexual behaviour questionnaire in rural South Africa", *BMC Medical Research Methodology* (2017) 17:125 DOI 10.1186/s12874-017-0403-8
- [7] Paul D. Allison Jeffrey K. Liker, "Analyzing Sequential Categorical Data on Dyadic Interaction: A Comment on Gottman", *Psychological Bulletin* 1982. Vol. 91, No. 2, 393-403



Kyongho Cho was born in Seoul, Korea in 1972. He received the B.A., M.A. and the Ph. D candidate degree in Spanish linguistics from Hankuk University of Foreign Language, Seoul, Korea in 2002. He served on the research staff at BK 21 Cervantes Team until 2004, Since March 2007, he has served as Spanish and Latin language teacher in Hankuk Academy of Foreign Studies, Yongin, Korea. And he has been a visiting professor of the Spanish Department, since March 2013, teaching Spanish Test *DELE* Class to Senior students in Cyber Hankuk University of Foreign Language, Seoul, Korea. Since March 2016, he has been in a doctoral course, gifted education, at Gachon University, Seongnam, Korea.



Sung Hee Park received the B.A. and M.A. degree from Ewha Womans University, Seoul, Korea and Ph. D degree from Purdue University, West Lafayette, USA. She is currently an associate professor of the Graduate School of Education in Gachon University, Korea. Her research interests include e-learning, technology integration in education, and gifted education.



Taeseon Yoon was born in Seoul, Korea, in 1972. He received the Ph.D. candidate degree in computer education from the Korea University, Seoul, Korea, in 2003. From 1998 to 2003, he was with EJB analyst and SCJP. From 2003 to 2004, he joined the Department of Computer Education, University of Korea, as a lecturer and Ansan University, as an adjunct professor. Since December 2004, he has been with the Hankuk Academy of Foreign Studies, where he was a computer science and statistics teacher.