IFSCC 2025 full paper (IFSCC2025-1180)

"Establishment and Validation of An Assessment Tool for Evaluating the Global Face Sagging"

Yuchen QIU [1], Yanwen JIANG [1*], Ping Xu [2] and Xiaomin Zheng [3]

[1] Shanghai China-Norm Quality Technical Service Co., Ltd., Shanghai 200072, China; [2] Department of Dermatology, Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, Shanghai, China; [3] Independent Researcher

1. Introduction

Skin aging is a complex biological process, which is regulated by both internal and external factors. From the perspective of the internal mechanism, with the increase of age, the proliferation of basal keratinocytes decreases, the decrease in the amount of sebum on the surface, the self-renewing capability of the epidermis reduce, and the gradual fracture of the collagen extracellular matrix of the dermis, resulting in the overall thinning of the epidermis, the decrease of the skin strength and elasticity and promotes the formation of skin sagging and wrinkles [1-3].

In terms of exogenous factors, photoaging is the most important external cause, prolonged exposure to sunlight leads to sustained degradation of extracellular matrix (ECM) proteins, such as collagen and elastin, and reduces the rate of collagen renewal/synthesis [4]. Smoking increases the level of matrix metalloproteinases, which leads to the degradation of collagen, elastic fibers and proteoglycans [5]. An increase in soot, an increase in particles from traffic, and higher PM10 background concentrations were associated with more pigment spots on the face and more pronounced nasolabial folds. [6]. Dietary factors should not be ignored, and a high-sugar diet can influence the skin aging process through the accumulation of glycosylated end products (AGEs) [7]. The synergistic effect of these internal and external factors eventually leads to the typical aging characteristics of the skin such as pigmentation, decreased elasticity, and wrinkle formation.

To sum up, a comprehensive global aging index encompasses various skin attributes, such as skin tone, elasticity, wrinkles, and sagging. It is evident that sagging is a critical marker in the aging process [8] and there is a strong correlation between wrinkles and sagging [9]. In addition, there are already some grading descriptions about the sagging score, "0-5 Score" for cheek [10], "0-6 score" for low eyelid [11] and "0-9 score" for periorbital and nasolabial folds [12] (Table 1).

Drawing on the aforementioned insights and adapting Griffiths' 10-point grading scale (ranging from 0 to 9) to evaluate [13], this study aims to establish a photographic scale for global sagging indicators. Through this photographic scale, our goal is to offer a tool that is both visual and quantifiable to facilitate comprehensive assessments of skin aging.

Table 1. Grading Descriptions About the Sagging Score

Author	Grading Area	Sagging Score
Tsukahara K, et al. (2007) [10]	Cheek	0=no sagging 1=slight sagging 2=mild sagging 3=moderate sagging 4=severe sagging 5=very severe sagging
Ezure T, et al. (2011) [11]	Lower eyelid	0=no sagging 1=slight sagging 2=mild sagging 3=moderate sagging 4=severe sagging 5=very severe sagging 6=extremely severe sagging
Nguyen T Q, et al. (2021) [12]	Periorbital Nasolabial folds	0=No sagging; tight, lifted appearance 9=Severe, extreme sagging/ drooping appearance

2. Materials and Methods

2.1. Subjects' Population

In this study, a total of 128 female participants, aged between 21 and 64, were enrolled after fulfilling the specified inclusion and exclusion criteria. These participants represented a diverse age range, allowing for a comprehensive analysis of skin aging patterns across different life stages. Throughout the duration of the study, a detailed set of three images was captured for each participant using the VISIA-CR system. These images included frontal, left-side, and right-side views, ensuring a thorough evaluation of facial features from multiple angles. The inclusion of such a diverse population and the use of standardized imaging techniques provided a robust foundation for the subsequent analysis of skin aging characteristics.

2.2. Grading Scale Establishment

Adopting a rigorous and orderly approach, our research team has meticulously crafted clear grade descriptions for the scoring on the evaluation scale. These descriptions are precisely aligned with the validated modified Griffiths' 10-point scale [12], which provides a robust framework for assessing clinical signs, adding an extra layer of detail to our assessment methodology.

The images captured for each participant were assigned random identifiers to ensure unbiased evaluation and were meticulously assessed by an experienced expert. The expert focused on analyzing a comprehensive set of skin aging parameters, including forehead wrinkles, glabella wrinkles, dropping of the upper outer eyelid, underneath eye wrinkles, eye bags, lacrimal grooves, nasolabial folds, crow's feet wrinkles, wrinkles of the corner of the lips, ptosis of the lower face, pores, and plumpness. Each parameter was evaluated based on predefined criteria to ensure consistency and objectivity in the assessment process [12]. The expert's detailed evaluations provided a wealth of data, which was systematically compiled to capture the nuances of skin across different facial regions. This rigorous approach allowed for the identification of key patterns and variations in skin characteristics among the participants [13].

According to existing literature, there is a significant correlation between wrinkles and skin sagging [9]. To comprehensively assess the attributes of global face sagging, a research team invited a panel of experienced expert to conduct detailed assessments on 12 indicators related to skin aging. These indicators include forehead wrinkles, glabella wrinkles, dropping of the upper outer eyelid, underneath eye wrinkles, eye bags, lacrimal grooves, nasolabial folds, crow's feet wrinkles, wrinkles of the corner of the lips, ptosis of the lower face, pores, and plumpness. To reduce data complexity and extract key information, the research team employed Principal Component Analysis (PCA) to perform dimensionality reduction on the evaluation data. Through PCA, several principal components were obtained, each representing a portion of the information from the original indicators. Subsequently, the research team calculated the weight of each principal component based on its variance contribution rate (i.e., the proportion of the original data's variability explained by that principal component) relative to the cumulative variance contribution rate. Using these weights, the team further computed a comprehensive score for each sample. The comprehensive score reflects the overall performance of the sample across multiple aging indicators. Based on the comprehensive scores, the research team ranked all the samples.

To translate the comprehensive scores into a more intuitive evaluation standard, the research team convened experienced expert to further classify the samples. The expert group divided the samples into grades ranging from 0 to 9 based on the comprehensive scores, where 0 represents the best skin condition and 9 represents the most severe degree of global face sagging. Following this, the expert group selected representative images from each grade to establish a standardized evaluation scale. This scale can not only be used for global face sagging assessment in scientific research but also serves as a reference for clinical diagnosis and cosmetic treatments.

2.3. Grading Scale Validation

To validate the newly established photographic scale, three independent experts learned and were tasked with assessing the global face sagging of all participant images using the scale. This multi-expert approach was employed to ensure the reliability and consistency of the grading system. The experts' assessments were compared to evaluate inter-rater agreement, which is a key indicator of the scale's robustness and applicability. The level of agreement among the experts confirmed the validity of the photographic scale as a reliable tool for evaluating global face sagging. This validation process not only reinforced the credibility of the scale but also demonstrated its potential for use in future research and clinical settings to assess skin aging objectively and accurately[14].

2.4. Statistical Analysis

Group differences across varying degrees were meticulously analyzed using one-way ANOVA, a statistical method that allows for the comparison of means among three or more independent groups. This approach is particularly useful for identifying significant differences between the groups being studied. To assess the test-retest reliability, which is a measure of the consistency of the scale, the intraclass correlation coefficient (ICC) was employed. The ICC is a statistical tool that provides an estimate of the degree to which individuals or items can be distinguished from one another relative to the total variation in the data.

Furthermore, the Pearson correlation coefficient was calculated to determine the strength and direction of the linear relationship between image analysis and image evaluation scores.

For all statistical tests, a threshold of statistical significance was set at $p<0.05$. This threshold is conventionally used to determine whether the results of a statistical analysis are likely to have occurred by chance. If the p-value is less than 0.05, it indicates that the observed effects are statistically significant and not due to random variation, thus providing confidence in the validity of the findings.

3. Results

3.1. PCA for Scale Establish

The Kaiser-Meyer-Olkin test (KMO test) and Bartlett's sphericity test were conducted to assess the suitability of the data for PCA. The KMO value of 0.913 demonstrates that the data are highly suitable for PCA (values above 0.6 are considered acceptable) [17]. Additionally, the Bartlett's sphericity test yielded a significance level of $p<0.001$,indicating a significant correlation among the variables, further confirming that the data are appropriate for PCA analysis (Table 2).

Table 2. KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.913
Bartlett's Test of Sphericity	Approx. Chi-Square	915.426
	df	66
	Significance	0.000

The PCA analysis identified the first two principal components (Component1 and Component2), which cumulatively accounted for 63.758% of the total variance.
Specifically, Principal Component 1 captured 53.912% of the variance, while Principal Component 2 contributed an additional 9.846% (Table 3). Based on the cumulative variance explanation rate and the Scree Plot [15] (Figure 1), the first two principal components were selected to calculate the Component Score Coefficient Matrix (Table 4) and to derive the composite score.

Table 3. Total Variance Explained

Component	Initial Eigenvalues（%）			Extraction Sums of Squared Loadings（%）		
	Total	Variance	Cumulative	Total	Variance	Cumulative
1	6.469	53.912	53.912	6.469	53.912	53.912
2	1.181	9.846	63.758	1.181	9.846	63.758
3	0.878	7.319	71.077			
4	0.713	5.939	77.016			
5	0.598	4.981	81.997			
6	0.433	3.610	85.606			
7	0.392	3.263	88.870			
8	0.358	2.979	91.849			
9	0.324	2.698	94.547			
10	0.266	2.220	96.767			
11	0.214	1.783	98.550			
12	0.174	1.450	100.000			

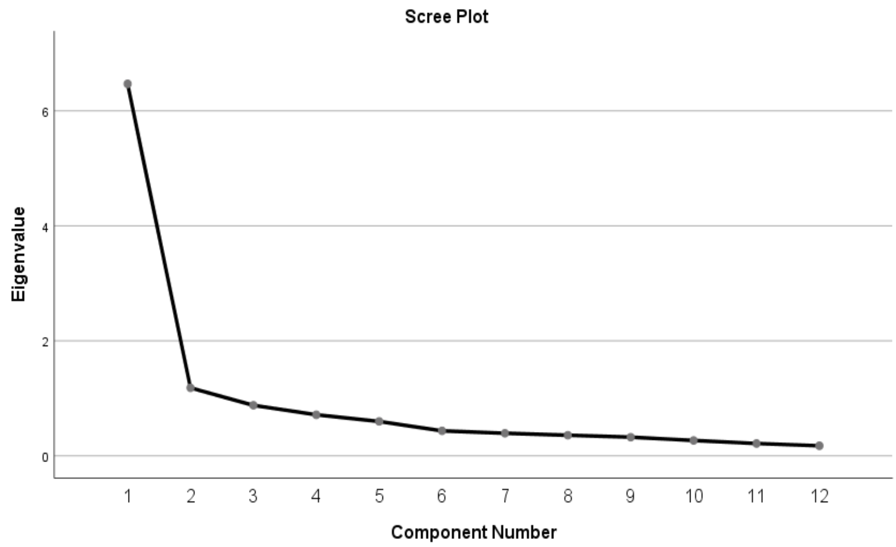

Figure 1. Scree Plot, this figure can be clearly stated that the inflection point is Component 2

Table 4. Component Score Coefficient Matrix

	Component	
	1	2
[1]. Forehead wrinkles	0.117	0.129
[2]. Glabella wrinkles	0.120	0.146
[3]. Dropping of the upper outer eyelid	0.123	0.126
[4]. Underneath eye wrinkles	0.112	0.172
[5]. Eye bag	0.059	-0.646
[6]. Lacrimal groove	0.120	-0.206
[7]. Nasolabial folds	0.127	0.010
[8]. Crow's feet wrinkles	0.134	0.101
[9]. Wrinkles of the corner of the lips	0.127	-0.083
[10]. Ptosis of lower part face	0.123	-0.026
[11]. Pores	0.038	0.495
[12]. Plumpness	0.120	-0.202

The component scores of all subjects' photos were sorted, and the expert group selected a reliable map with a score ranging from 0 to 9 to draw the facial image (Figure 2).

Figure 2. Atlas of Global Face Sagging with 10-point Scale, the spectrum reveals the sagging features across the entire face. To safeguard the privacy of the individual depicted to the greatest extent possible, the image has undergone specialized processing rather than being obscured through a covering method.

3.2. Conformity of Grading

The experts conducted repeated grading of all images from the 128 subjects on the same computer screen, which had been calibrated using a calibration tool to verify the reliability of image evaluation. The self-consistency assessments of the three experts were r=0.967 (p<0.001), r=0.852 (p<0.001), and r=0.915 (p<0.001). The consistency between the three expert pairs was r=0.795 (p<0.001), r=0.827 (p<0.001), and r=0.848 (p<0.001).These results indicate that the image evaluation had moderate to good repeatability for the classification and evaluation scale (Table 5, Table 6). According to the criteria, 0.500 ≤ r ≤ 0.750 is considered moderate,and 0.750 ≤ r ≤ 0.900 is considered good.

Table 5. Expert's Self-consistency Results by ICC

Expert	Intraclass Correlation Coefficient	Significance
Expert 1	0.967	p<0.001
Expert 2	0.852	p<0.001
Expert 3	0.915	p<0.001

Table 6. Expert Pairwise Consistency Results by ICC

Expert	Intraclass Correlation	
	Expert 2	Expert 3
Expert 1	*0.795**	*0.827**
Expert 2	/	*0.848**

**Indicates a significant correlation at level 0.05 (two-tailed)*

4. Discussion

During the scale development process, the 12 assessed attributes were broadly categorized into two main dimensions: skin quality and wrinkle characteristics. Notably, both dimensions are closely related to aging. Given that aging factors might introduce bias into the evaluation results, the research team conducted a validation study using new 38 facial photos. Following standardized procedures, a certified expert performed blinded assessments of the 12 individual attributes and the overall degree of sagging using the developed scale. Subsequently, computational analysis was conducted to derive comprehensive scores by applying pre-determined model parameters, including score coefficients and factor weights. Finally, the predictive values generated by the model were compared with the expert assessment results using intraclass correlation coefficient (ICC) analysis to evaluate measurement consistency. The results showed that the consistency between the grading score and the predicted score was 0.843, indicating that the model has good predictive performance (Table 7, Table 8, Figure 3).

Table 7. Grading Sore and Pedicted Score

	Mean	SD	Max	Min
Grading score by expert	4.11	2.44	8	0
Predicted score	4.66	3.25	9	0

Table 8. ICC Aalysis between Gading Sore and Predicted Score

	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	0.843	0.720	0.915	11.752	37	38	0.000
Average Measures	0.915	0.837	0.956	11.752	37	38	0.000

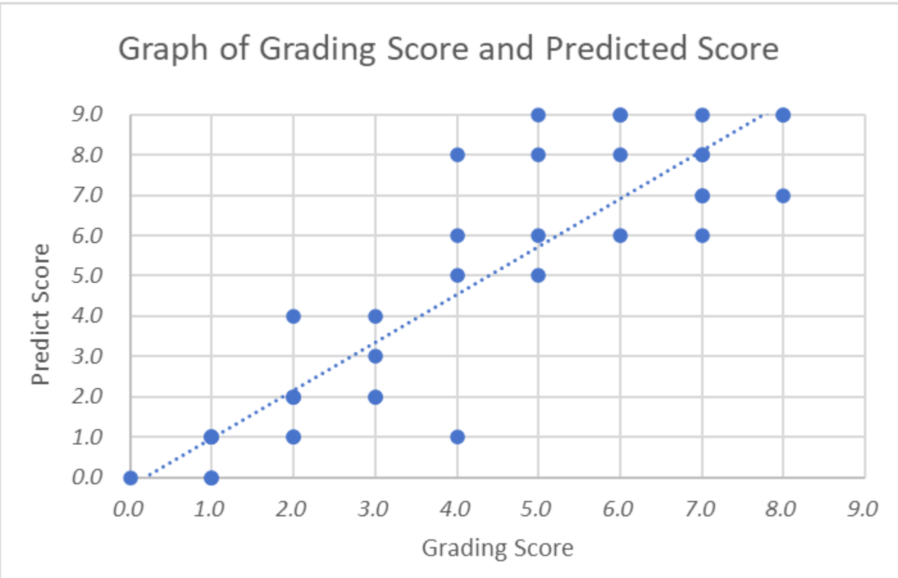

Figure 3. Graph of Grading Score and Predicted Score

Beyond the expert-driven analysis of global sagging and 12 attributes, his study uncovered crucial biomechanical insights. The significant correlations r>0.500) bserved between wrinkles, lumpness, rooping eyelids,and global sagging provide robust support for a tricompartment model of facial aging, ncompassing the lower face, idface, nd periocular regions [1-3]. These results suggest that clinical assessments should prioritize these zones to more effectively address the underlying mechanisms of facial aging.

Notably, the correlation between pore characteristics and global facial sagging is relatively low. This may be attributed to the failure to distinguish between different types of pores. In particular, younger subjects often have sebaceous pores, which can introduce data bias. Moving forward, it would be advisable to focus solely on evaluating aging-related pores. This targeted approach is likely to enhance the accuracy of the model (Table 9, Figure 4).

Table 9. Correlations Analysis between Grading Scores

Area	Attributes	Correlations
The lower part of the face	Ptosis	0.836
	Wrinkles of the corner of the lips	0.824
The middle part of the face	Plumpness	0.802
	Nasolabial folds	0.760
	Pores	0.189
Periocular	Crow's feet wrinkles	0.780
	Dropping of the upper outer eyelid	0.714
	Lacrimal groove	0.679
	Underneath eye wrinkles	0.593
	Eye bag	0.341
Forehead	Glabella wrinkles	0.686
	Forehead wrinkles	0.656

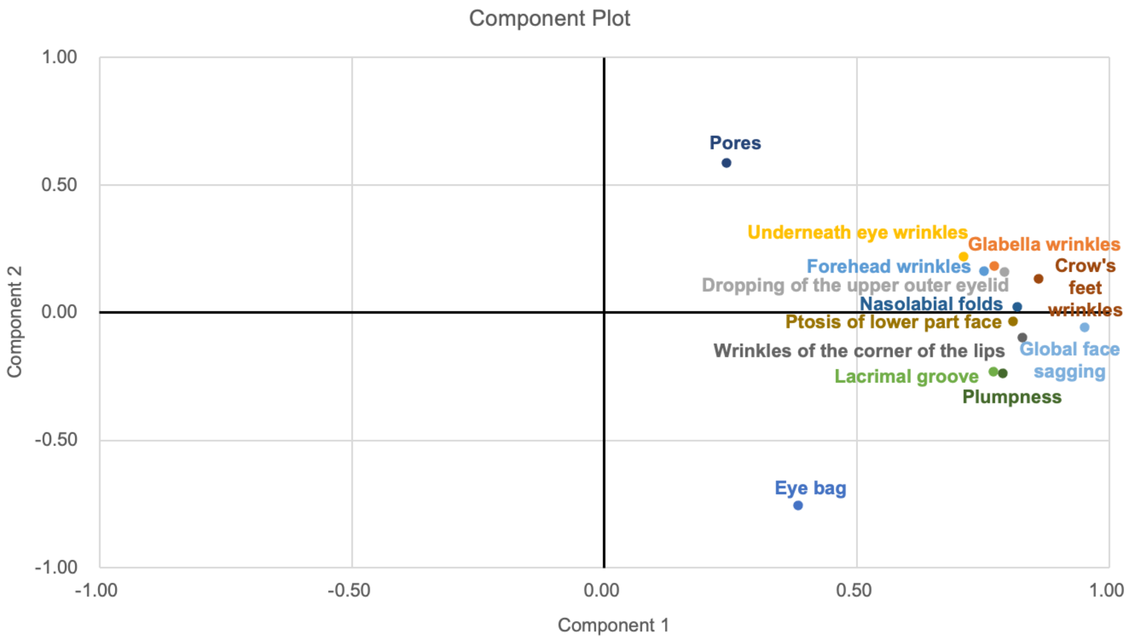

Figure 4. Component Plot

Methodologically, while expert scores are adept at capturing clinically perceptible changes, future work could integrate 3D imaging technology to objectively quantify and incorporate relevant image analyses, such as wrinkles and pores. This advancement would enhance the precision of model establishment. Additionally,other auxiliary data could be incorporated to further validate the scale.

5. Conclusion

This study has successfully developed a standardized atlas system for assessing global sagging by integrating principal component analysis (PCA) with expert experience. Dual evaluations conducted by three experts revealed excellent intra-rater reliability (ICC=0.852–0.967) and inter-rater reliability (ICC=0.795–0.827), highlighting superior consistency in both self-assessment and cross-evaluator comparisons. By combining dimensionality reduction technology with clinical expertise, the modeling approach ensures both objective scientific evaluation and practical clinical applicability. This atlas serves as a standardized tool for the quantitative assessment of skin laxity, offering significant value for clinical treatment planning, efficacy evaluation, and multicenter research. Future work will focus on expanding the sample size and incorporating multi-ethnic data to enhance the universality of this atlas system.

Reference

[1]. Rittie L , Fisher G J .Natural and Sun-Induced Aging of Human Skin [J]. Cold Spring Harb Perspect Med, 2015, 5(1): a015370.DOI:10.1101/cshperspect.a015370.

[2]. Farage M A , Miller K W , Elsner P .Intrinsic and extrinsic factors in skin ageing:a review. [J]. International Journal of Cosmetic Science, 2008(2):30.

[3]. Ganceviciene R , Liakou A I , Theodoridis A ,et al.Skin anti-aging strategies [J].Dermato-Endocrinology, 2012, 4(3):308-319.DOI:10.4161/derm.22804.

[4]. Pavida P , Jitlada M , Ornicha P ,et al.Role of Matrix Metalloproteinases in Photoaging and Photocarcinogenesis [J].International Journal of Molecular Sciences, 2016, 17(6).DOI:10.3390/ijms17060868.

[5]. Morita,Akimichi.Tobacco smoke causes premature skin aging [J].Journal of Dermatological Science, 2007, 48(3):169-175.DOI:10.1016/j.jdermsci.2007.06.015.

[6]. Vierk?Tter A , Schikowski T , Ranft U ,et al.Airborne Particle Exposure and Extrinsic Skin Aging [J].Journal of Investigative Dermatology, 2010, 130 (12):2719-2726.DOI:10.1038/jid.2010.204.

[7]. Luevano-Contreras C , Chapman-Novakofski K .Dietary Advanced Glycation End Products and Aging[J].Nutrients, 2010, 2(12).DOI:10.3390/nu2121247.

[8]. Robic, Julie, Alex Nkengne, Bee Leng Lua, Aurore Bellanger, Rachel Hu, and Katell Vie. 'Modeling of Global Skin Aging Indices among Caucasian and Asian Women'. Skin Research and Technology 29, no. 2 (February 2023): e13222. https://doi.org/10.1111/srt.13222.

[9]. Carina T , Gabor D , Andrea L ,et al.Characterizing Facial Skin Ageing in Humans: Disentangling Extrinsic from Intrinsic Biological Phenomena[J].Biomed Res Int, 2015, 2015:318586.DOI:10.1155/2015/318586.

[10]. Tsukahara K , Sugata K , Osanai O ,et al.Comparison of age-related changes in facial wrinkles and sagging in the skin of Japanese, Chinese and Thai women[J].Journal of Dermatological Science, 2007, 47(1):19-28.DOI:10.1016/j.jdermsci.2007.03.007.

[11]. Ezure T , Yagi E , Kunizawa N ,et al.Comparison of sagging at the cheek and lower eyelid between male and female faces[J].Skin Research & Technology, 2011, 17(4):510-515.DOI:10.1111/j.1600-0846.2011.00526.x.

[12]. Thu Q Nguyen et al., 'A Randomized, Double-Blind, Placebo- Controlled Clinical Study Investigating the Efficacy and Tolerability of a Peptide Serum Targeting Expression Lines' 14, no. 5 (2021).

[13]. Fong S , Pabis K , Latumalea D ,et al.Principal component-based clinical aging clocks identify signatures of healthy aging and targets for clinical intervention[J].Nature Aging, 2024, 4(8):1137-1152.DOI:10.1038/s43587-024-00646-8.

[14]. Rossi A B , Nkengne A , Stamatas G ,et al.Development and validation of a photonumeric grading scale for assessing lip volume and thickness[J].Journal of the European Academy of Dermatology and Venereology, 2011, 25(5):523-531.DOI:10.1111/j.1468-3083.2010.03816.x.

[15]. Jason W Osborne, 'Best Practices in Exploratory Factor Analysis', n.d.

[16]. Koo T K , Li M Y .A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research[J].Journal of Chiropractic Medicine, 2016, 15(2):155-163.DOI:10.1016/j.jcm.2016.02.012.

[17] Henry F. Kaiser, 'An Index of Factorial Simplicity', Psychometrika 39, no. 1 (March 1974): 31–36, https://doi.org/10.1007/BF02291575.