Technical report of EFA for SAQ dataset

1. Objective Statement

SPSS software is a very practical and easy-to-use statistical analysis software, learning and using SPSS software is an indispensable step in learning statistical analysis, so the frequency of using SPSS software is very high among scholars who study statistical disciplines or decision science. The SPSS anxiety questionnaire (SAQ) investigates the anxiety of SPSS users through 23 questions.

EFA aims to identify those highly correlated variables as factors for subsequent analysis by analyzing the structure of the interrelationships of many variables. In this paper, in order to define the underlying structure of the questionnaire in the analysis and to pave the way for further subsequent regression analysis or discriminant analysis, i.e., to perform data summarization and data reduction, conducted an Exploration Factors analysis.

2. Methodology

2.1. Data description

In this paper, we use 2571 questionnaires completed by SAQ as the dataset, excluding 9 missing data, the number of Observation is 2562. The questionnaire consists of 23 questions, i.e., 23 variables to examine the anxiety level of SPSS users. The 23 questions contain a five-point Likert scale, which are 'strongly disagree', 'disagree', ' neither agree nor disagree', 'agree or disagree', 'agree or disagree', and 'agree or disagree'. neither agree nor disagree', 'agree', and 'strongly agree' (SD, D, N, A, and SA, respectively). Thus, all 23 variables are ordinal variables using ordinal scale measurement. Increasing values of the variables mean that users are more anxious about using SPSS. In this paper, we plan to initially use the 23 variables for EFA analysis and subsequently perform some variable deletion based on the reported results to allow each Factor to better represent its dimension.

These 23 variables are named Question 1 through Question 23, and their respective descriptions are shown in Table 1

Table 1 Variable Description

Variable Name	**Variable Description**
Question 1	Statistics makes me cry
Question 2	My friends will think I'm stupid for not being able to cope with SPSS
Question 3	My friends will think I'm stupid for not being able to cope with SPSS
Question 4	I dream that Pearson is attacking me with correlation coefficients
Question 5	I don't understand statistics
Question 6	I have little experience of computers
Question 7	All computers hate me
Question 8	I have never been good at mathematics
Question 9	My friends are better at statistics than me
Question 10	Computers are useful only for playing games
Question 11	I did badly at mathematics at school
Question 12	People try to tell you that SPSS makes statistics easier to understand but it doesn't
Question 13	I worry that I will cause irreparable damage because of my incompetence with computers
Question 14	Computers have minds of their own and deliberately go wrong whenever I use them
Question 15	Computers are out to get me
Question 16	I weep openly at the mention of central tendency
Question 17	I slip into a coma whenever I see an equation
Question 18	SPSS always crashes when I try to use it
Question 19	Everybody looks at me when I use SPSS
Question 20	I can't sleep for thoughts of eigenvectors
Question 21	I wake up under my duvet thinking that I am trapped under a normal distribution
Question 22	My friends are better at SPSS than I am
Question 23	If I'm good at statistics my friends will think I'm a nerd

2.2. Analysis methods

In this paper, we mainly adopt Exploration Factors analysis to perform dimension reduction of the variables in the dataset. EFA is a good starting point in inference statistics.In EFA, the researcher does not presuppose relationships between variables and factors but rather analyzes the data to identify these relationships. Specifically, EFA attempts to find the minimum number of factors that explain the observed co-variability among variables. Each factor corresponds to one or more variables, and these factors explain the co-variation of the variables.

The steps of an EFA typically include variates and model specification, pre-test for factorability assessment, factor extraction, factor rotation, interpretation of the results, and result validation.

2.3. Sample size requirement

To perform EFA analysis, the absolute size of the dataset should be at least greater than 100 observations, preferably greater than 200, and the dataset used in this paper is very satisfied with this requirement.

In addition, the ratio of observations to variables should be at least 5 to 1. This paper has 2,562 observations and 23 variables, the ratio also meets the sample size requirements for EFA.

2.4. Variates and Model specification

Ade

The variates of EFA are formed to maximize their explanation of the entire variable set, they are expressed as a linear combination of underlying factors. Mathematically, the factor model of standardized variables is represented as:

$$Question_i = A_{i1}F_1 + A_{i2}F_2 + \cdots + A_{im}F_m + V_iU_i$$

Where, $Question_i = ith$ standardized variable

A_{ij} =the standardized coefficient of variable i on common factor j

F = common factor

V_i =the standardized coefficient of variable i on unique factor i

U_i =the unique factor for variable i m

= number of common factor

The common factors can be expressed as:

$$F_i = W_{i1}Question_1 + W_{i2}Question_2 + \cdots + W_{i23}Question_{23}$$

Where, $F_i = the$ estimate of the i th factor

W_i =weight or factor score coefficient

2.5. Factorability assessment

This paper will examine the overall measures of intercorrelation among the variables and the Individual variable's measure of sampling adequacy to conduct some pre-tests on the factorability assessment for EFA. These include visual inspection, Bartlett's test of sphericity, and KMO measure of sampling adequacy (MSA). Visual inspection mainly examines whether the correlation of these variables is greater than

0.03 and the partial correlation of them is larger than 0.5 to identify the appropriateness for factors analysis. For Bartlett's test of sphericity, the null hypothesis is that the correlation matrix is an identity matrix, and the alternative hypothesis is that the correlation matrix is not an identity matrix. This is to examine whether the correlation between the factors is zero or not. KMO MSA mainly indicate the appropriateness for performing factor analysis, the larger the value of KMO is, the more suitable it is for performing factor analysis, especially when the value of KMO is greater than 0.9.

2.6. Factor extraction

After the pre-test, if the dataset is suitable for EFA, this paper will perform factor extraction, which is to extract the representative factors from these variables. There are mainly two factor extraction methods, they are principal component analysis and common factor analysis. The difference between them is whether to extract all variance or only common variance. This paper will compare the results of these two methods, i.e., comparing how much the extracted variance explains the factor to determine a suitable method. This paper will also select the appropriate stopping rule, which is a criterion for determining the number of factors to retain in the results.

2.7. Factors interpretation

This paper will adopt a suitable rotation method to expand the cross-loading between factors to enhance the explanatory ability of the factors. Rotation methods are divided into orthogonal factor rotation and oblique factor rotation. The difference between them is whether the rotation is carried out at an angle of 90 degrees to maximize cross-loading.

Since the strength of factor analysis results is measured by communality. Therefore, this paper will carry out the interpretation of the communality results of each variable. Then, this paper will re-specify the factor model by the results obtained from the above methods for data reduction, i.e., using a suitable surrogate method to represent factors. Finally, this paper will explain the component matrix and rotated component matrix with some explanation and description of each factor.

2.8. Result validation

Result validation is also very important, which determines whether the EFA results in this paper are replicable and generalizable. This paper will validate the results of the EFA as well. This is mainly done by random sampling, where the sample is split into an estimation sample and a validation sample by about 50% to compare their results to whether they are robust or not.

3. Results and interpretation

3.1. Factorability assessment

For visual inspection, we mainly check the values of correlation and partial correlation between variables, as shown in Table 2 and Table 3.

Table 2 Correlation Matrix

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23
Q1	1																						
Q2	-0.099	1																					
Q3	-0.337	0.318	1																				
Q4	0.436	-0.112	-0.38	1																			
Q5	0.402	-0.119	-0.31	0.401	1																		
Q6	0.217	-0.074	-0.227	0.278	0.257	1																	
Q7	0.305	-0.159	-0.382	0.409	0.339	0.514	1																
Q8	0.331	-0.05	-0.259	0.349	0.269	0.223	0.297	1															
Q9	-0.092	0.315	0.3	-0.125	-0.096	-0.113	-0.128	0.016	1														
Q10	0.214	-0.084	-0.193	0.216	0.258	0.322	0.284	0.159	-0.134	1													
Q11	0.357	-0.144	-0.351	0.369	0.298	0.328	0.345	0.629	-0.116	0.271	1												
Q12	0.345	-0.195	-0.41	0.442	0.347	0.313	0.423	0.252	-0.167	0.246	0.335	1											
Q13	0.355	-0.143	-0.318	0.344	0.302	0.466	0.442	0.314	-0.167	0.302	0.423	0.489	1										
Q14	0.338	-0.165	-0.371	0.351	0.315	0.402	0.441	0.281	-0.122	0.255	0.325	0.433	0.45	1									
Q15	0.246	-0.165	-0.312	0.334	0.261	0.36	0.391	0.3	-0.187	0.295	0.365	0.332	0.342	0.38	1								
Q16	0.499	-0.168	-0.419	0.416	0.395	0.244	0.389	0.321	-0.189	0.291	0.369	0.408	0.358	0.418	0.454	1							
Q17	0.371	-0.087	-0.327	0.383	0.31	0.282	0.391	0.59	-0.037	0.218	0.587	0.333	0.408	0.354	0.373	0.41	1						

Q18	0.347	-0.164	-0.375	0.382	0.322	0.513	0.501	0.28	-0.15	0.293	0.373	0.493	0.533	0.498	0.343	0.422	0.376	1					
Q19	-0.189	0.203	0.342	-0.186	-0.165	-0.167	-0.269	-0.159	0.249	-0.127	-0.2	-0.267	-0.227	-0.254	-0.21	-0.267	-0.163	-0.257	1				
Q20	0.214	-0.202	-0.325	0.243	0.2	0.101	0.221	0.175	-0.159	0.084	0.255	0.298	0.204	0.226	0.206	0.265	0.205	0.235	-0.249	1			
Q21	0.329	-0.205	-0.417	0.41	0.335	0.272	0.483	0.296	-0.136	0.193	0.346	0.441	0.374	0.399	0.3	0.421	0.363	0.43	-0.275	0.468	1		
Q22	-0.104	0.231	0.204	-0.098	-0.133	-0.165	-0.168	-0.079	0.257	-0.131	-0.162	-0.167	-0.195	-0.17	-0.168	-0.156	-0.126	-0.16	0.234	-0.1	-0.129	1	
Q23	-0.004	0.1	0.15	-0.034	-0.042	-0.069	-0.07	-0.05	0.171	-0.062	-0.086	-0.046	-0.053	-0.048	-0.062	-0.082	-0.092	-0.08	0.122	-0.035	-0.068	0.23	1

Table 3 Anti-image Matrices

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23
Q1	.930a																						
Q2	-0.02	.875a																					
Q3	0.053	-0.157	.951a																				
Q4	-0.167	-0.041	0.084	.955a																			
Q5	-0.156	0.01	0.037	-0.134	.960a																		
Q6	0.02	-0.053	-0.042	-0.007	-0.035	.891a																	
Q7	0.023	0.016	0.072	-0.087	-0.044	-0.275	.942a																
Q8	-0.049	-0.033	-0.007	-0.075	-0.027	0.024	-0.015	.871a															
Q9	-0.016	-0.193	-0.142	0.03	-0.02	0.011	-0.03	-0.099	.834a														
Q10	-0.012	-0.012	-0.016	0.006	-0.093	-0.116	-0.033	0.051	0.043	.949a													
Q11	-0.041	0.038	0.064	-0.022	-3.27E-05	-0.084	0.045	-0.412	0.037	-0.092	.906a												
Q12	-0.007	0.031	0.087	-0.154	-0.058	0.045	-0.043	0.033	-0.003	-0.019	-0.005	.955a											
Q13	-0.085	-0.008	-0.032	0.023	0.004	-0.164	-0.039	0.002	0.061	-0.06	-0.099	-0.198	.948a										
Q14	-0.04	0.023	0.069	-0.004	-0.026	-0.099	-0.054	-0.023	-0.043	-0.012	0.035	-0.082	-0.099	.967a									
Q15	0.089	0.037	0.008	-0.062	0.014	-0.128	-0.077	-0.033	0.068	-0.093	-0.052	-0.026	-0.008	-0.093	.940a								
Q16	-0.264	-0.011	0.081	-0.036	-0.096	0.102	-0.02	-0.006	0.052	-0.082	0.005	-0.04	0.026	-0.081	-0.232	.934a							
Q17	-0.047	-0.029	0.035	-0.035	-0.018	0.041	-0.08	-0.296	-0.068	0.012	-0.23	0.006	-0.09	-0.028	-0.085	-0.076	.931a						
Q18	-0.023	0.018	0.039	-0.025	0.002	-0.244	-0.087	0.024	-0.006	-0.026	-0.022	-0.146	-0.17	-0.145	0.038	-0.09	-0.034	.948a					

Q19	0.012	-0.029	-0.121	-0.034	-0.018	-0.015	0.068	0.047	-0.111	-0.009	-0.006	0.04	0.009	0.044	0.009	0.047	-0.047	0.03	.941a				
Q20	-0.016	0.059	0.078	-0.004	-0.011	0.051	0.048	0.021	0.038	0.043	-0.082	-0.065	0.018	0.001	-0.037	-0.005	0.015	-0.003	0.091	.889a			
Q21	0.006	0.041	0.07	-0.086	-0.046	0.039	-0.208	-0.02	-0.031	0.017	-0.005	-0.079	-0.033	-0.063	0.035	-0.085	-0.041	-0.072	0.031	-0.323	.929a		
Q22	0.001	-0.121	-0.007	-0.033	0.035	0.04	0.013	-0.023	-0.126	0.019	0.034	0.018	0.052	0.029	0.025	-0.003	0.01	-0.024	-0.115	-0.011	-0.024	.878a	
Q23	-0.059	-0.002	-0.076	-0.017	-0.005	0.018	-0.008	0.002	-0.092	0.015	0.01	-0.028	-0.03	-0.026	-0.024	0.023	0.055	0.023	-0.038	-0.028	0.013	-0.176	.766a

a Measures of Sampling Adequacy(MSA)

As can be seen from Table 2, the absolute value of the correlation between some variables and most other variables is less than 0.3, that is, there is not enough relationship to form an interpretable factor, such as Q5 and Q10. They can be used as candidates for deletion. For Table 3, The partial correlation of most variables is greater than 0.5, and even basically hits 0.8. This means that the dataset is suitable for EFA.

Similarly, this article conducts KMO and Bartlett's Test to examine overall measures of intercorrelation. The results are reported in Table 4

Table 4 KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.93
Bartlett's Test of Sphericity	Approx. Chi-Square	19334.492
	df	253
	Sig.	0

From Table 4 we can see that the chi-square statistic value of Bartlett's Test of Sphericity is 19334.492 and the p-value is 0.000. Therefore, the null hypothesis, which is his correlation matrix is an identity matrix, is significantly rejected, and it can be concluded that the relationship between factors is significantly different from 0. The value of KMO is 0.93 which is greater than 0.9, so we can demonstrate that performing factor analysis is marvelous.

3.2. Factor extraction

By comparing the two extraction methods, principal component analysis and common factor analysis, this article finds that the total explained variance using the Principal Component analysis method is more, so this article chooses the Principal

Component analysis method. The total explained variance table is reported in the table below.

Table 5 Total explained variance

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Varianc	Cumulative %	Total	% of Varianc	Cumulative %	Total	% of Varianc	Cumulative %
1	5.819	30.624	30.624	5.819	30.624	30.624	3.159	16.628	16.628
2	1.708	8.988	39.613	1.708	8.988	39.613	2.63	13.843	30.471
3	1.281	6.74	46.352	1.281	6.74	46.352	2.379	12.519	42.99
4	1.168	6.146	52.499	1.168	6.146	52.499	1.807	9.509	52.499
5	0.898	4.729	57.228						
6	0.827	4.353	61.58						
7	0.771	4.058	65.638						
8	0.758	3.99	69.628						
9	0.72	3.792	73.419						
10	0.684	3.599	77.018						
11	0.656	3.451	80.469						
12	0.603	3.175	83.644						
13	0.561	2.955	86.599						
14	0.509	2.676	89.276						
15	0.491	2.586	91.862						
16	0.433	2.281	94.143						
17	0.409	2.152	96.295						
18	0.371	1.95	98.245						
19	0.334	1.755	100						
Extraction Method: Principal Component Analysis.									

As can be seen from Table 5, through the principal component analysis method, 4 factors were extracted, and the cumulative square loadings were 52.499. This is not a very high percentage, which means that the extracted factors explained approximately 52.5% of the 23 variables. information.

This article uses the scree test criterion as a stopping rule, which is based on the pattern of eigenvalues of the extracted factor. Figure 1 shows that eigenvalues are examined to find and “elbow” in the pattern denoting subsequent factors that are not distinctive. From the scree plot It can be seen that when four factors are extracted, the eigenvalues corresponding to the factors are very small, and their contribution to the total variance is relatively small. Therefore only 4 factors are extracted.

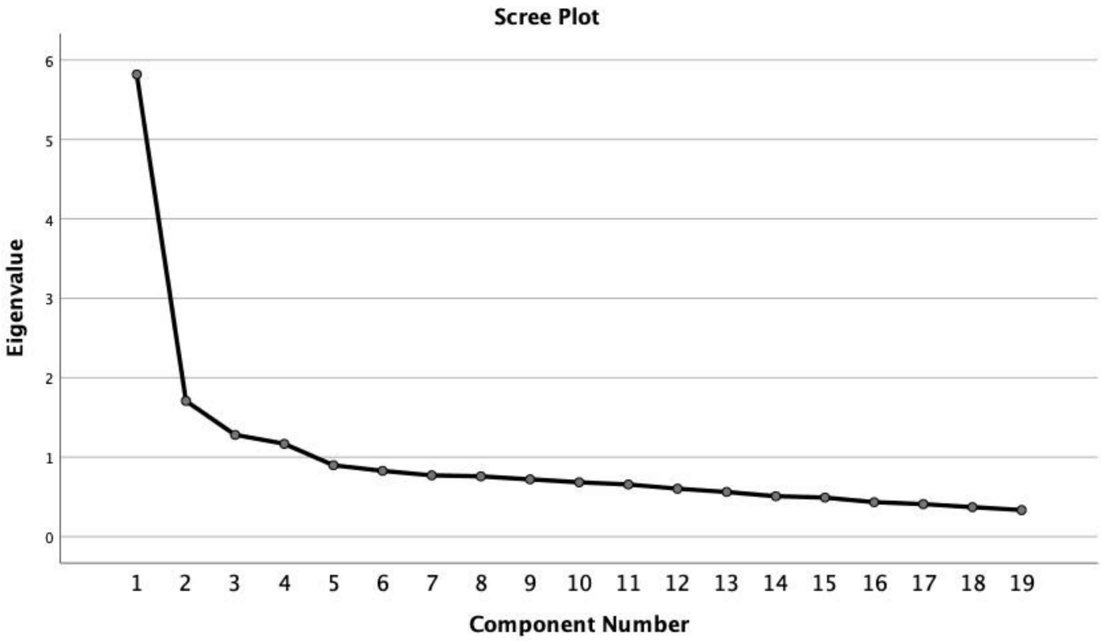

Figure 1 Scree plot

The common factor variance of each variable is reported in Table 6. We can demonstrate that the common factor variance after extraction is mostly around 0.5, which is not large. This should be the reason why the total variance explained of the factors is smaller.

Table 6 Communality

	Initial	Extraction
Q1	1	0.435
Q2	1	0.414
Q3	1	0.53
Q4	1	0.469
Q5	1	0.343
Q6	1	0.654
Q7	1	0.545
Q8	1	0.739
Q9	1	0.484
Q10	1	0.335

Q11	1	0.69
Q12	1	0.513
Q13	1	0.536
Q14	1	0.488
Q15	1	0.378
Q16	1	0.487
Q17	1	0.683
Q18	1	0.597
Q19	1	0.343
Q20	1	0.484
Q21	1	0.55
Q22	1	0.464
Q23	1	0.412
Extraction Method: Principal Component Analysis.		

3.3. Factors interpretation

The component matrix and rotated component matrix that extracted four factors are shown in Table 7 and Table 8 respectively. In the component matrix, we can see that some variables have cross-loading, and the cross-loading is small, which may cause the factors to not explain the variables well. After rotation by the varimax method, cross-loading still exists in some variables. Therefore, consider deleting these variables to expand cross-loading and improve the explanatory power of the factors.

Therefore, this article deletes the four variables Q12, Q4, Q16, and Q1 from the rotated component matrix, and then re-specifies factor analysis.

Table 7 Component Matrix

	Component			
	1	2	3	4
Q18	0.701			
Q7	0.685			
Q16	0.679			
Q13	0.673			
Q12	0.669			

Q21	0.658			
Q14	0.656			
Q11	0.652			-0.4
Q17	0.643	0.33		-0.342
Q4	0.634			
Q3	-0.629			
Q15	0.593			
Q1	0.586			
Q5	0.556			
Q8	0.549	0.401	-0.323	-0.417
Q10	0.437		0.363	
Q20	0.436		-0.404	
Q19	-0.427	0.39		
Q9		0.627		
Q2	-0.303	0.548		
Q22	-0.302	0.465		0.378
Q6	0.562		0.571	
Q23		0.366		0.507
Extraction Method: Principal Component Analysis.				
a 4 components extracted.				

Table 8 Rotated Component Matrix

	Component			
	1	2	3	4
Q6	0.8			
Q18	0.684	0.327		
Q13	0.647			
Q7	0.638	0.327		
Q14	0.579	0.36		

Q10	0.55			
Q15	0.459			
Q20		0.677		
Q21		0.661		
Q3		-0.567		0.368
Q12	0.473	0.523		
Q4	0.32	0.516	0.314	
Q16	0.334	0.514	0.313	
Q1		0.496	0.356	
Q5	0.319	0.429		
Q8			0.833	
Q17			0.747	
Q11			0.747	
Q9				0.648
Q22				0.645
Q23				0.586
Q2		-0.338		0.543
Q19		-0.372		0.428
Extraction Method: Principal Component Analysis.				
Rotation Method: Varimax with Kaiser Normalization.				
a Rotation converged in 8 iterations.				

After deletion, Table 9 shows the component matrix. There is still a lot of crossloading and exists between many variables. For example, the most obvious one, Q17, has factor loading in factors 1, 2, and 3. Therefore, it is still necessary to expand crossing-loading through rotation.

Table 9 Component Matrix after deletion

	Component			
	1	2	3	4
Q18	0.707			

Q7	0.701			
Q13	0.683			
Q11	0.671		-0.34	
Q14	0.66			
Q21	0.654			0.389
Q17	0.65	0.345	-0.323	
Q3	-0.625			
Q6	0.601		0.501	
Q15	0.6			
Q8	0.558	0.407	-0.457	
Q5	0.529			
Q10	0.449		0.359	
Q19	-0.443	0.379		
Q9		0.624		
Q2	-0.323	0.544		
Q22	-0.33	0.444		0.413
Q23		0.351		0.541
Q20	0.438		-0.389	0.459
Extraction Method: Principal Component Analysis.				
a 4 components extracted.				

After rotation, we can see from Table 10that cross-loading has been reduced a lot. Even though it still appears in variables such as Q17 and Q21, the cross-loading has increased a lot, both are greater than 2, which means that we can ignore these cross-loading. Based on the results, we can divide these 19 variables into 4 factors.
They are F1, including Q6, Q18, Q7, Q13, Q14, Q10, Q15 and Q5; F2, including Q8, Q17 and Q11; F3, including Q20, Q21 and Q3; F4, including Q2, Q19, Q22, Q23 and Q9.

Table 10 Rotated Component Matrix After deletion

	Component			
	1	2	3	4
Q6	0.793			
Q18	0.712			
Q7	0.676			
Q13	0.66			
Q14	0.622			
Q10	0.563			
Q15	0.49			
Q5	0.403			
Q8		0.851		
Q17	0.307	0.763		
Q11		0.762		
Q20			0.757	
Q21	0.367		0.65	
Q3			-0.57	
Q2			-0.47	0.442
Q19			-0.458	0.329
Q22				0.669
Q23				0.64
Q9			-0.304	0.613

According to their variable description, we can summarize the 4 factors into 4 different dimensions. The variables of F1 are basically related to computers, which we can define as anxiety about computers; the variables of F2 are all related to mathematics and equations, which can be summarized as anxiety about mathematics; the variables of F3 all mention some statistical professional names, such as Normal distribution, standard deviation, and eigenvectors, so we can understand it as anxiety about statistics; the variables in F4 are all related to friends' evaluation of themselves.

For example, using SPSS will make friends think they are stupid, so it can be summarized as peer anxiety. These four aspects of anxiety all affect different aspects of SPSS, including the hardware level, knowledge level, and external reasons. Therefore, EFA revealed through data analysis that these variables with the same characteristics can form Factors for data analysis, which is in line with our expectations and research purposes. This article adopts the most direct surrogate method, using the variable with the largest factor loading in each factor, that is, the variable that has the most influence on the factor, as the surrogate variable. In general, the summarization is shown in the Table 11.

Table 11 Factor summarization

Factor	Including variables	Surrogate variables	Meaning
F1	Q6, Q18, Q7, Q13, Q14, Q10, Q15, Q5	Q6	Computer anxiety
F2	Q8, Q17, Q11	Q8	Mathematics anxiety
F3	Q20, Q21, Q3	Q20	Statistics anxiety
F4	Q2, Q19, Q22, Q23, Q9	Q2	Peer anxiety

3.4. Result validation

This article randomly divides the observation into two parts of about 50%, one is the estimate sample and the other is the validation sample. By performing the abovementioned EFA analysis on the validation sample again, it is possible to determine whether the results have changed significantly to determine the validity of the EFA results.

The results of KMO and Bartlett's Test are shown in Table 12, where the Bartlett's Test result is still significant, and the KMO of 0.899 is also very close to 9. Overall, the validation sample also works for EFA.

Table 12 KMO and Bartlett's Test (Validation sample)

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.899

Bartlett's Test of Sphericity	Approx. Chi-Square	7024.907
	df	171
	Sig.	0

Table 13 and Table 14 show the component matrix and rotated component matrix of the validation Sample. It can be seen from the results of the component matrix that the variables are also extracted by 4 factors, but there is still a lot of cross-loading, and rotation is needed to optimize the results.

Table 13 Component Matrix (Validation sample)

	Component			
	1	2	3	4
Q18	0.7			
Q7	0.686			
Q13	0.677			
Q11	0.666		0.321	-0.352
Q21	0.653			0.366
Q14	0.647			
Q17	0.644	0.374		
Q3	-0.623	0.329		
Q6	0.596		-0.473	
Q15	0.581			
Q5	0.543			
Q8	0.54	0.396	0.415	-0.375
Q10	0.452		-0.38	
Q19	-0.43	0.365		
Q9		0.626		
Q2	-0.309	0.535		
Q22	-0.325	0.417		0.414
Q20	0.395	-0.321	0.477	0.324

Q23		0.306		0.509
Extraction Method: Principal Component Analysis.				
a 4 components extracted.				

After rotating using the Varimax method, the cross-loading of variables is significantly reduced, and the remaining cross-loading is greater than 2 and can be ignored. It can be seen from the rotated component matrix that it is still divided into 4 factors after extraction, and the variables contained in the 4 factors are completely consistent with the results of the above estimate sample. Therefore, we can conclude that the EFA method used in this article is valid and has replicability and generalizability.

Table 14 Rotated Component Matrix (Validation Sample)

	Component			
	1	2	3	4
Q6	0.78			
Q18	0.745			
Q7	0.684			
Q13	0.655			
Q14	0.634			
Q10	0.517			
Q15	0.476			
Q05	0.435			
Q8		0.856		
Q11		0.772		
Q17	0.345	0.748		
Q20			0.755	
Q21	0.419		0.64	
Q3	-0.305		-0.575	

Q2			-0.517	0.364
Q19			-0.409	0.329
Q22				0.671
Q23				0.635
Q9			-0.366	0.593
Extraction Method: Principal Component Analysis.				
Rotation Method: Varimax with Kaiser Normalization.				
a Rotation converged in 7 iterations.				

4. Conclusion

Through EFA, this article extracts the 23 questions (variables) of the SAQ dataset into 4 factors, representing computer anxiety, mathematics anxiety, statistics anxiety, and peer anxiety respectively. This means that the underlying structure of the SAQ dataset and the intrinsic relationship between variables are revealed through data analysis without presetting classification by question content.

This means that SPSS users' anxiety about SPSS can be representatively divided into these four types of anxiety. If you want to make some useful suggestions for alleviating anxiety for SPSS users or make some suggestions for SPSS developers' decision-making, The results of this EFA can be used as an effective reference.

After EFA, if you want to use SPSS anxiety as an independent variable or dependent variable for regression analysis or discriminant analysis, the data summarization and data reduction caused by EFA will be very helpful.