**Qing Shu - BUS-Z 798 - Assignment 2**

**Part 1 - EFA Analysis**

**1.1** **Which items would you recommend dropping from your scale? Re-run the initial factor analyses with the dropped items excluded; how did the absence of those items changed the results? Provide any required evidence from the analyses to support this.**

**Firstly, Y23 (“When I get up in the morning, I feel like going to work”) and Y9 (“Most days I am enthusiastic about my work”) should be excluded *with no doubt* for three reasons.** First, after the varimax rotation, Y23 and Y9 provide little correlation with factors for which they are designed. Their high uniqueness also indicates that they share very little information with other items, thus they are hard to be assigned a ‘common factor’ (Table 1). Second, by checking their description in the questionnaire, Y23 seems not measuring the working engagement. Because it is a natural feeling for most people at the moment of getting up that they don’t want to go to work. By ‘natural feeling’, I imply that it is not that people don’t want to work but people don’t want to get up early at all. Therefore, this item cannot accurately measure the engagement. As for Y9, it is measuring the state of motivation of work rather than the personal evaluation of the work. The question designers probably think that a strong motivation maybe the result of high satisfaction, thus can reflect the level of satisfaction. But such operation can raise problems because Y9 is quite similar with Y35 (“I am enthusiastic about my job”). Therefore, I won’t retain this item even if its statistics looked normal. Meanwhile, I feel unexpected to see that Y9 is not highly corrected with Y35 in the data. This raises some of my concern about the data quality. ( I notice there are two Y29 in the questionnaire. May one of them be Y36?)

**Table 1 Rotated factor loadings (Part of the whole table)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Factor1 | Factor2 | Factor3 | Factor4 | Factor5 | Factor6 | … | Uniqueness |
| Y38 | -0.1557 | -0.1065 | 0.5022 | 0.0036 | -0.0712 | 0.6687 |  | 0.2471 |
| Y39 | -0.0202 | -0.1901 | 0.4535 | 0.0320 | -0.0715 | 0.2562 |  | 0.4004 |
| Y3 | 0.0731 | -0.0957 | 0.8276 | -0.0477 | -0.0854 | -0.0458 |  | 0.1836 |
| Y6 | 0.0456 | -0.0557 | 0.9305 | -0.0223 | 0.0386 | 0.1552 |  | 0.0650 |
| Y31 | 0.8059 | 0.0996 | 0.0173 | 0.1187 | 0.0110 | -0.0034 |  | 0.2830 |
| Y32 | 0.8853 | 0.0494 | 0.0628 | -0.0075 | 0.0070 | -0.0707 |  | 0.1628 |
| Y35 | 0.6413 | 0.1137 | -0.0881 | 0.1793 | 0.1515 | -0.0130 |  | 0.3362 |
| Y36 | 0.6961 | 0.0198 | -0.0282 | 0.1021 | -0.0048 | -0.0269 |  | 0.3504 |
| Y20 | 0.1540 | 0.0970 | -0.0853 | 0.6353 | 0.0710 | 0.0457 |  | 0.4832 |
| Y23 | 0.1550 | 0.1268 | -0.0217 | 0.3525 | 0.1300 | 0.0227 |  | **0.6944** |
| Y27 | 0.2363 | 0.1389 | -0.0353 | 0.6734 | 0.0005 | -0.0354 |  | 0.4262 |
| Y28 | 0.7144 | 0.0615 | 0.0742 | 0.1938 | 0.1536 | -0.0526 |  | 0.3391 |
| Y29 | 0.6705 | 0.0868 | 0.0499 | 0.0353 | 0.1707 | -0.0187 |  | 0.3832 |
| Y9 | 0.0553 | 0.2960 | -0.1161 | 0.0056 | 0.2684 | 0.0411 |  | **0.6746** |
| Y10 | 0.2565 | 0.2258 | -0.0588 | 0.0681 | 0.6407 | -0.0884 |  | 0.4503 |
| Y11 | 0.0799 | 0.6634 | -0.0484 | 0.0277 | 0.0313 | -0.0791 |  | 0.3636 |
| Y12 | 0.0763 | 0.7660 | -0.1144 | 0.1183 | 0.0438 | 0.0378 |  | 0.2905 |
| Y13 | 0.0910 | 0.8783 | -0.0866 | 0.0481 | 0.0876 | -0.0679 |  | 0.1723 |

**Secondly, Y20 (“My job inspires me”) and Y27 (“I feel happy when I am working intensively”) *need to* be excluded.** The items’ description designed per se imply no obvious problem. I tried to retain this two until I re-ran the EFA with dropping the Y23 and Y9 and found no significant change. At this moment, I choose to believe what the data want to show me – Y20 and Y27 contribute together to factor 4 (Table 1) which is distinct from factor 1 (Table 1). It probably because the respondents have some tendency to misunderstand such descriptions. For, example, maybe they think ‘My job inspires me’ is asking whether their job can provide them with inspirations.

**At last, I have difficulty in deciding whether Y10 (“I feel fairly satisfied with my present job”) should be excluded or not (Even though I have to drop it).** First of all, I see no any problem of this item in measuring the job satisfaction. I also tend to exclude the possibility that respondent would misunderstand the meaning of this item, because its meaning is very similar with Y12 (“I find real enjoyment in my work”). The EFA results show Y10 contributes alone to factor 5 (Table 1). After I dropped Y27 and Y20, Y10 shifts to factor 1 (Table 2), which leads to a ‘ideal’ three-factor model but is contradictory with the design of the questionnaire. **I think there must be some problem with the data. If the data provided to me is from a survey. I will go to check the survey process.**

**Table 2 Rotated factor loadings after dropping Y27 and Y20 (Part of the whole table)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Factor1 | Factor2 | Factor3 | … | … | Uniqueness |
| Y38 | -0.3547 | 0.5513 | 0.3075 |  |  | 0.2841 |
| Y39 | -0.2327 | 0.5808 | 0.1386 |  |  | 0.4323 |
| Y3 | -0.1139 | 0.7530 | 0.3037 |  |  | 0.1785 |
| Y6 | -0.1281 | 0.7688 | 0.4346 |  |  | 0.0855 |
| Y31 | 0.7509 | 0.2598 | -0.1028 |  |  | 0.2686 |
| Y32 | 0.7564 | 0.3259 | -0.1401 |  |  | 0.2056 |
| Y35 | 0.7134 | 0.1155 | -0.1078 |  |  | 0.3251 |
| Y36 | 0.6665 | 0.2295 | -0.1957 |  |  | 0.3398 |
| Y28 | 0.7155 | 0.2569 | -0.0814 |  |  | 0.3588 |
| Y29 | 0.6716 | 0.2269 | -0.0524 |  |  | 0.3939 |
| Y10 | **0.4541** | -0.1084 | 0.1070 |  |  | 0.5577 |
| Y11 | 0.3659 | -0.3299 | 0.5137 |  |  | 0.3893 |
| Y12 | 0.3678 | -0.3769 | 0.5325 |  |  | 0.3588 |
| Y13 | 0.4469 | -0.4240 | 0.6333 |  |  | 0.1742 |

**When all above problematic items is excluded, changes include:** The turn pointing of the scree plot shift from the number 5 (Figure 1) to the number 4 (Figure 2). Besides, the remaining items can fit well with the three factors mirroring the three constructs (Table 3).

**Figure 1 Scree plot before items are dropped**



**Figure 2 Scree plot with all problematic items dropped**



**Table 3 Factor loadings with all problematic items dropped (Part of the whole table)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Factor1 | Factor2 | Factor3 | … | … | Uniqueness |
| Y38 | -0.3340 | 0.5622 | 0.2996 |  |  | 0.3121 |
| Y39 | -0.2105 | 0.5941 | 0.1383 |  |  | 0.4256 |
| Y3 | -0.0881 | 0.7510 | 0.2948 |  |  | 0.2147 |
| Y6 | -0.1103 | 0.7703 | 0.4176 |  |  | 0.1221 |
| Y31 | 0.7655 | 0.2256 | -0.0842 |  |  | 0.2936 |
| Y32 | 0.7839 | 0.2961 | -0.1254 |  |  | 0.1853 |
| Y35 | 0.7108 | 0.0872 | -0.0956 |  |  | 0.3592 |
| Y36 | 0.6823 | 0.1981 | -0.1775 |  |  | 0.3640 |
| Y28 | 0.7209 | 0.2295 | -0.0727 |  |  | 0.3722 |
| Y29 | 0.6664 | 0.1985 | -0.0441 |  |  | 0.4514 |
| Y11 | 0.3447 | -0.3396 | 0.5137 |  |  | 0.4227 |
| Y12 | 0.3471 | -0.3923 | 0.5502 |  |  | 0.3483 |
| Y13 | 0.4184 | -0.4364 | 0.6378 |  |  | 0.1959 |

**1.2 How many factors would you recommend retaining? Provide any required evidence from the analyses to support this.**

First, the first scree plot (Figure 1) shows that no more than 4 factors should be retained. However, it doesn’t mean I have to retain exact the first 4 factors, because the theoretical background implies 3 factors to be the ideal situation.

Second, if applying the eigenvalues-gearter-than-one rule, we should retain 3 factors (Table 4). Conway and Huffcutt (2003) don’t recommend this rule in their paper, but (Hayton et al., 2004) also claims that this rule can be used to decide the upper bound for the number of factors to be retained.

**Table 4 Eigenvalue of factors before items are dropped (Part of the whole table)**

Factor analysis/correlation Number of obs = 578

Method: iterated principal factors Retained factors = 17

Rotation: (unrotated) Number of params = 153

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Factor | Eigenvalue | Difference | Proportion | Cumulative |
| Factor1 | 4.35981 | 1.69253 | 0.3728 | 0.3728 |
| Factor2 | 2.66728 | 1.14347 | 0.2281 | 0.6009 |
| Factor3 | 1.52381 | 0.63343 | 0.1303 | 0.7312 |
| Factor4 | 0.89038 | 0.42831 | 0.0761 | 0.8074 |
| Factor5 | 0.46207 | 0.08434 | 0.0395 | 0.8469 |
| Factor6 | 0.37772 | 0.10353 | 0.0323 | 0.8792 |
| Factor7 | 0.27419 | 0.01628 | 0.0234 | 0.9026 |
| Factor8 | 0.25791 | 0.06213 | 0.0221 | 0.9247 |
| Factor9 | 0.19578 | 0.03716 | 0.0167 | 0.9414 |
| Factor10 | 0.15863 | 0.02812 | 0.0136 | 0.9550 |
| Factor11 | 0.13051 | 0.01301 | 0.0112 | 0.9662 |
| Factor12 | 0.11750 | 0.01821 | 0.0100 | 0.9762 |
| Factor13 | 0.09929 | 0.00750 | 0.0085 | 0.9847 |
| Factor14 | 0.09179 | 0.04370 | 0.0078 | 0.9926 |
| Factor15 | 0.04810 | 0.01155 | 0.0041 | 0.9967 |
| Factor16 | 0.03655 | 0.03377 | 0.0031 | 0.9998 |
| Factor17 | 0.00278 | 0.00312 | 0.0002 | 1.0000 |
| Factor18 | -0.00034 | - | -0.0000 | 1.0000 |

LR test: independent vs. saturated: chi2(153) = 3857.06 Prob>chi2 = 0.0000

**1.3 How were the results different (or similar) using orthogonal and non-orthogonal rotation? What does this tell you about your data?**

After using non-orthogonal rotation (promax, Table 5), comparing with orthogonal rotation (varimax, Table 1), all poorly performing items remain their poor performance, such as Y9, Y10, Y20, Y23, and Y27, but as for previously weakly performing items (Y38, Y39, Y35), their low loadings are worsened. This may again reveal problems related to quality of this data set.

**Table 5 Factor loadings with promax rotation (Part of the whole table)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Factor1 | Factor2 | Factor3 | Factor4 | Factor5 | Factor6 | … | Uniqueness |
| Y38 | -0.0249 | 0.0208 | 0.1927 | -0.0344 | 0.0198 | -0.0906 |  | 0.2471 |
| Y39 | 0.0174 | -0.0292 | 0.0904 | 0.0304 | -0.0174 | 0.0777 |  | 0.4004 |
| Y3 | -0.0177 | 0.0195 | 0.7491 | -0.0320 | 0.0190 | -0.0844 |  | 0.1836 |
| Y6 | 0.0195 | -0.0223 | 0.9916 | 0.0337 | -0.0203 | 0.0879 |  | 0.0650 |
| Y31 | 0.7139 | 0.0069 | 0.0079 | -0.0043 | 0.0050 | -0.0222 |  | 0.2830 |
| Y32 | 0.9420 | 0.0126 | 0.0179 | -0.0190 | -0.0046 | -0.0495 |  | 0.1628 |
| Y35 | 0.3682 | -0.0049 | -0.0107 | 0.0250 | 0.5784 | 0.0384 |  | 0.3362 |
| Y36 | 0.4569 | 0.0072 | 0.0101 | -0.0083 | 0.0443 | -0.0286 |  | 0.3504 |
| Y20 | 0.0150 | -0.0101 | -0.0169 | 0.6607 | -0.0096 | 0.0465 |  | 0.4832 |
| Y23 | 0.0119 | -0.0050 | -0.0110 | 0.1218 | -0.0068 | 0.0378 |  | 0.6944 |
| Y27 | 0.0006 | 0.0236 | 0.0283 | 0.7355 | 0.0177 | -0.0795 |  | 0.4262 |
| Y28 | 0.5117 | -0.0123 | -0.0174 | 0.0375 | -0.0096 | 0.0616 |  | 0.3391 |
| Y29 | 0.4291 | -0.0164 | -0.0243 | 0.0265 | -0.0152 | 0.0844 |  | 0.3832 |
| Y9 | 0.0135 | 0.0204 | -0.0234 | 0.0231 | -0.0133 | 0.1211 |  | 0.6746 |
| Y10 | 0.0028 | 0.0390 | 0.0327 | -0.0347 | 0.0210 | 0.7361 |  | 0.4503 |
| Y11 | -0.0168 | 0.4199 | 0.0280 | -0.0303 | 0.0181 | -0.0790 |  | 0.3636 |
| Y12 | 0.0093 | 0.6062 | -0.0159 | 0.0193 | -0.0085 | 0.0396 |  | 0.2905 |
| Y13 | 0.0070 | 0.9219 | -0.0104 | 0.0106 | -0.0060 | 0.0290 |  | 0.1723 |

**1.4 How did the results differ when using factor analysis vs. PCA?**

Results from the two methods doesn’t differ essentially.

**Table 6 Factor loadings with PCA (Part of the whole table)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Comp1 | Comp2 | Comp3 | Comp4 | Comp5 | Comp6 | **……** |
| Y38 | -0.1634 | 0.3126 | 0.3367 | 0.0963 | 0.0791 | 0.1266 |  |
| Y39 | -0.1156 | 0.3671 | 0.2132 | 0.1106 | 0.0301 | 0.1818 |  |
| Y3 | -0.0662 | 0.4284 | 0.2822 | -0.1047 | -0.0088 | -0.0430 |  |
| Y6 | -0.0671 | 0.4169 | 0.3616 | -0.0878 | 0.0093 | -0.1570 |  |
| Y31 | 0.3417 | 0.1881 | -0.1007 | -0.0808 | -0.1042 | 0.1065 |  |
| Y32 | 0.3291 | 0.2252 | -0.1308 | -0.1711 | -0.1039 | 0.0130 |  |
| Y35 | 0.3384 | 0.0998 | -0.1048 | 0.0156 | 0.0163 | 0.1048 |  |
| Y36 | 0.3066 | 0.1777 | -0.1832 | -0.0396 | -0.1141 | 0.0988 |  |
| Y20 | 0.1826 | -0.0229 | 0.0646 | 0.5950 | 0.0718 | 0.1732 |  |
| Y23 | 0.1677 | -0.0123 | 0.1271 | 0.4353 | 0.1086 | -0.6754 |  |
| Y27 | 0.2195 | 0.0128 | 0.0780 | 0.5102 | -0.0602 | 0.2936 |  |
| Y28 | 0.3387 | 0.1881 | -0.0785 | -0.0388 | 0.0353 | -0.0131 |  |
| Y29 | 0.3115 | 0.1697 | -0.0681 | -0.1799 | 0.0300 | -0.1351 |  |
| Y9 | 0.1283 | -0.1711 | 0.2149 | -0.1695 | 0.6580 | 0.3891 |  |
| Y10 | 0.2371 | -0.0763 | 0.0925 | -0.1163 | 0.5727 | -0.3039 |  |
| Y11 | 0.1895 | -0.2225 | 0.3808 | -0.1387 | -0.2871 | -0.1681 |  |
| Y12 | 0.1978 | -0.2532 | 0.4014 | -0.0588 | -0.1959 | 0.1691 |  |
| Y13 | 0.2198 | -0.2623 | 0.4011 | -0.1399 | -0.2308 | 0.0280 |  |

**Part 2 - Parallel Analysis**

**2.1 What do the results of the parallel analysis tell you?**

It tells that all factors’ eigenvalues are higher than their random counterparts (Figure 3), thus the parallel analysis cannot decide that they are trivial.

**Figure 3**



**Reference**

Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, *6*(2), 147-168. <https://doi.org/10.1177/1094428103251541>

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191-205. <https://doi.org/10.1177/1094428104263675>