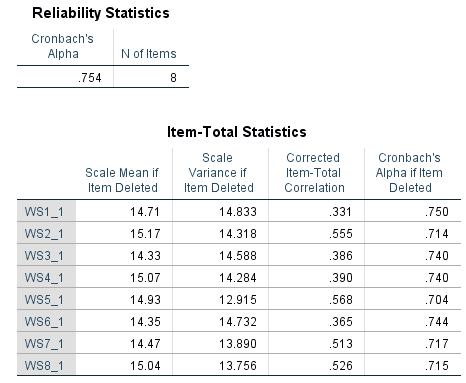
**Question 1: Internal Consistency and Test-Retest Reliability in SPSS**

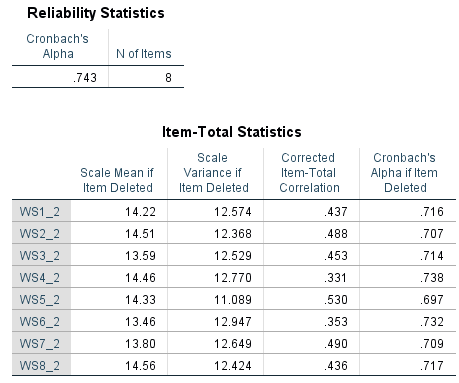
1. *Internal consistency and b. Suggested item-level improvements*

Time 1 internal consistency as expressed by Cronbach’s Alpha is adequate, measuring .754:



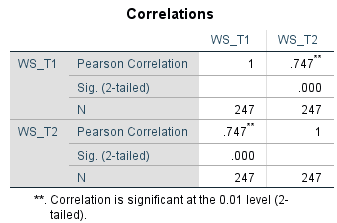
For Time 1, internal consistency as measured by Cronbach’s Alpha would not be improved by removing any of the items.

Time 2 internal consistency as expressed by Cronbach’s Alpha is adequate, measuring .743:



For Time 2, internal consistency as measured by Cronbach’s Alpha would not be improved by removing any of the items.

1. *What is the test-retest reliability of the measure?*



Test-retest reliability measured as 0 .747. This is an adequate level of reliability.

1. *95% Confidence Intervals*

What is the 95% confidence interval for any score X obtained at Time 1?

Syntax utilized: 

Confidence interval for a given score in Time 1 is +/- 0.25

What is the 5% confidence interval for any score X obtained at Time 2?

Syntax utilized: 

Confidence interval for a given score in Time 1 is +/- 0.27

**Question 2: Factor Analysis in R**

1. *MAP Test*

For the MAP test, I used the VSS() function from the psych package. Package documentation suggests extracting “more [factors] than hypothesized.” Parallel analysis suggested 14 factors are present, so in the VSS() syntax, I extracted 15 factors. MAP test resulted in a maximum suggested factor number of 13.

Syntax used:



Abridged output:

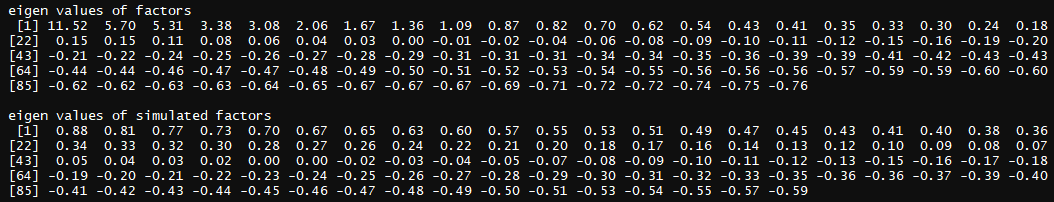


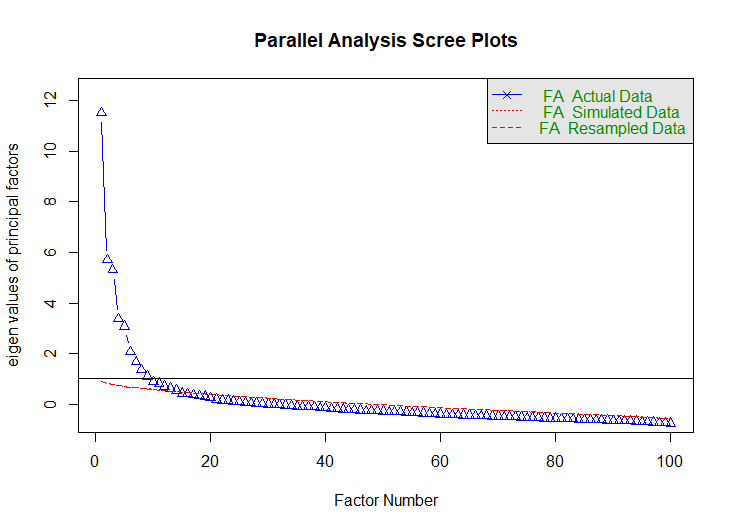
1. *Parallel Analysis*

I used the fa.parallel() function from the psych package to conduct parallel analysis. Results suggest that 14 factors are present.

Abridged output:

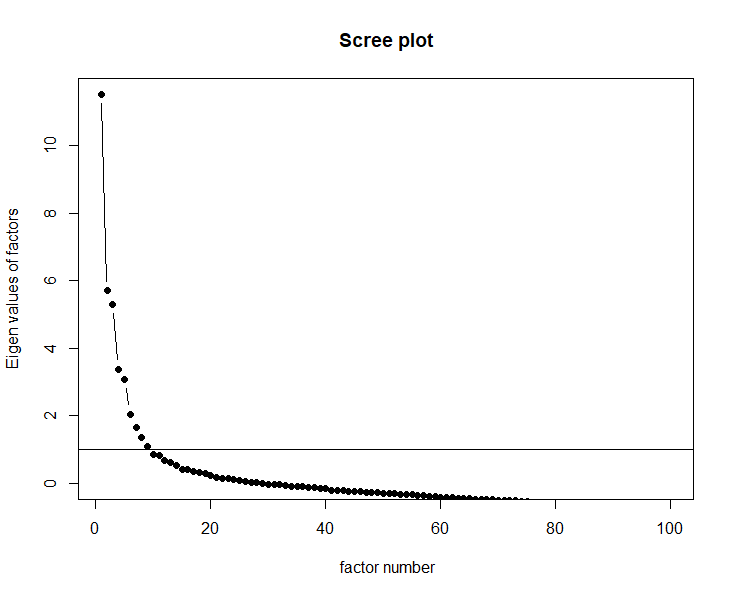






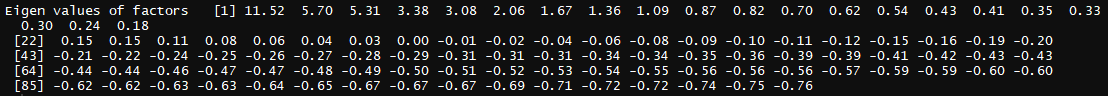
1. *Scree Plot*

Using the scree() function from the psych package, I outputted a scree plot and Eigenvalues for determining the number of factors in the dataset based on the Kaiser criterion. A visual evaluation of the Scree plot suggests presence of 9 or 10 factors.



1. *Kaiser Criterion*

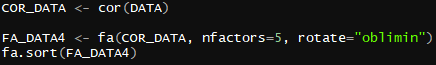
9 Eigenvalues are present over 1.0.



1. *Exploratory Factor Analysis*

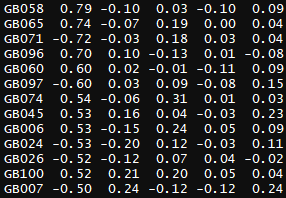
I used the fa() function from the psych package to output factor loadings. **Bowler’s Least Accurate Mid-term Exam Factor Determination Method** suggested extracting 5 factors.

Syntax used:

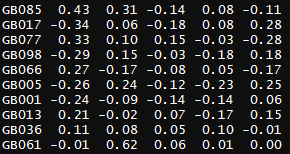


Abridged output is interspersed with conclusions:

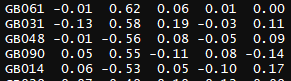




All items above retained for Factor 1, except eliminate GB074 due to double loading.



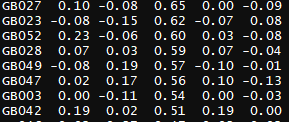
All items above eliminated due to insufficient loadings on all factors.



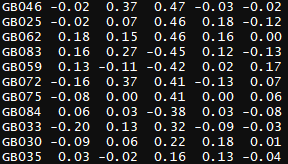
All items above retained for Factor 2.



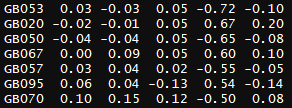
All items above eliminated due to insufficient loadings on all factors.



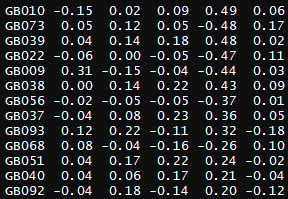
All items above retained for Factor 3.



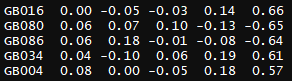
All items above eliminated due to insufficient loadings on all factors.



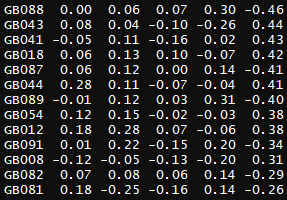
All items above retained for Factor 4.



All items above eliminated due to insufficient loadings on all factors.



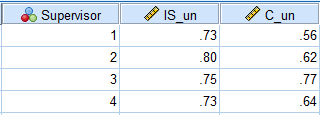
All items above retained for Factor 5.



All items above eliminated due to insufficient loadings on all factors.

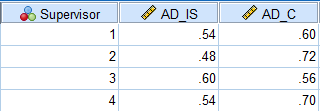
**Question 3: Inter-rater reliability and Inter-rater agreement in SPSS**

1. *Calculate and interpret rwg with a uniform distribution:*



These parameters show the level of agreement among employee ratings of supervisor “Initiating Structure” and “Consideration” performance. The critical value benchmark for the rwg measure of inter-rater agreement is, ratings that have agreement are > 0.7. According to the rwg metric, all raters agree for each rating of each supervisor regarding the rating of “Initiating Structure,” and for this metric it would be appropriate to aggregate rater scores. However, according to the rwg metric, We do not have inter-rater agreement for rater scores of “Consideration” when rating Supervisors 1, 2, or 4. It would not be appropriate to aggregate these scores. Ratings for Supervisor 3 could be aggregated.

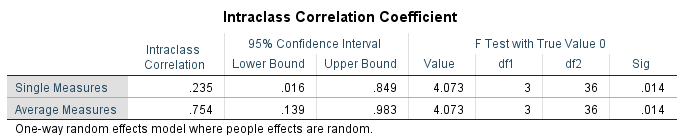
1. *Calculate and interpret ADM:*



These parameters show the level of agreement among employee ratings of supervisor “Initiating Structure” and “Consideration” performance. The critical value benchmark for the ADM measure of inter-rater agreement is, ratings that have agreement are < 0.8. According to the ADM metric, all raters agree for each rating of each supervisor, on both ratings. This means that according to this metric, it is appropriate to aggregate rater scores.

1. *ICC(1) and ICC(k)*

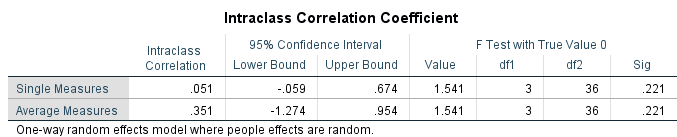
*ICC(1) and ICC(k) for Initiating Structure:*



ICC(1) is labeled above as “Single Measures – Intraclass Correlation. This parameter measures combined inter-rater reliability and inter-rater agreement for the “Initiating Structure” rating. 0.235 indicates a low level of combined IRR and IRA.

ICC(k) is labeled above as “Average Measures – Intraclass Correlation. This measure is regarded as one of the best measures of the extent to which the mean rating assigned by a group of judges is reliable. 0.754 is above our cutoff of .5, indicating a moderate to high level of reliability.

*ICC(1) and ICC(k) for Consideration:*



ICC(1) is labeled above as “Single Measures – Intraclass Correlation.” This parameter measures combined inter-rater reliability and inter-rater agreement for the “Consideration” rating. 0.05 indicates an extremely low level of combined IRR and IRA.

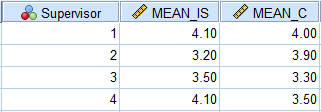
ICC(k) is labeled above as “Average Measures – Intraclass Correlation.” This measure is regarded as one of the best measures of the extent to which the mean rating assigned by a group of judges is reliable. 0.351 indicates a low to medium-low level of reliability of these raters for this rating.

*Question 3 Conclusions*

1. *Which supervisor’s scores should be aggregated?*

According to rwg and ADm, scores from the “Initiating Structure” measure can be aggregated. As well, the ICC(1) and ICC(k) procedures indicate that “Initiating Structure” measures are appropriate to aggregate, but “Consideration” measures are not.

1. *Which supervisor has the highest aggregated level of each measure?*

**

*Initiating Structure:* Supervisors 1 and 4 are tied for the highest aggregated level of this measure.

*Consideration:* While the test indicated that it is inappropriate to aggregate “Consideration” scores, the Midterm directions request we report the supervisor with the highest aggregated score. Supervisor 1 has the highest aggregated level of this measure.