

CFA

Anton Cronet, Gustaf Strid

2025-04-21

To determine what factors relate to job performance, a data set containing nine job performance-related variables regarding 1000 employees was chosen. The aim is to confirm the latent structures suggested by theory of which influence job performance. To test these latent constructs a confirmatory factor analysis was conducted. Where we found that the first measurement model that was proposed was not sufficient to explain the variation in the data.

Introduction

In the context of worklife, it is important to understand the factors that contribute to job performance. Analyzing these factors, such as with Confirmatory Factor Analysis (CFA), can help draw conclusions to a vague concept. Job performance can be measured by many metrics, such as revenue, or hours worked, but these only measure the extent of these variables. What actually is important when looking at job performance? What factors play which roles in the whole concept of job performance?

The purpose of this report is to employ Confirmatory Factor Analysis to test theoretical predictions regarding the factors that contribute to job performance. This involves assessing whether the proposed measurement model adequately captures the latent structure of job performance as hypothesized based on prior research and theories. Specifically, we aim to verify if the observed variables scores on psychological tests, client and superior satisfaction, education levels, IQ scores, work hours, hours spent on personal culture outside of work, and project completion percentages can by creating a three factor model consisting of the three latent structures motivation, intellect and social ability sufficiently explain the variation in the data.

Data description

The data set sets a score from 0-100 for most variables. The only exceptions are IQ, and time variables such as hours worked in a week, or hours spent towards personal culture. The data measures different aspects of either the employees themselves or the effect they are having in their workplace. The data consists of 1000 observations.

There is a lack of information around the data set, however it could be assumed that these observations are sampled in a similar way, seeing as these are transformed into scores relative to each other. With 1000 observations, there is plenty of data to fit a Factor Analysis model. The data was not transformed any further.

Overview of the 9 variables with some observations

```
## 'data.frame':    1000 obs. of  9 variables:
##  $ PsychTest1: int   62 46 68 55 51 51 55 55 50 55 ...
##  $ PsychTest2: int   78 27 75 56 32 66 40 64 33 36 ...
##  $ YrsEdu      : int    5 2 2 4 4 3 0 1 2 0 ...
##  $ IQ          : int   97 93 96 103 98 102 98 95 96 94 ...
##  $ HrsTrn      : int    6 7 5 7 5 10 8 7 10 6 ...
##  $ HrsWrk      : int   33 54 47 80 53 56 54 75 69 68 ...
##  $ ClientSat   : int   84 55 70 63 55 63 61 58 63 66 ...
##  $ SuperSat    : int   59 38 68 81 39 68 53 71 61 56 ...
##  $ ProjCompl   : int   34 56 38 78 56 57 53 75 62 63 ...
```

The PsychTest variables refer to 2 psychological tests, and the score the employees had on these.

YrsEdu is the amount of higher education that the employees have, with the data ranging from 0 to 5.

IQ is IQ test score.

HrsTrn is the amount of hours a week dedicated to “personal culture” outside of work hours, this could for example be reading relevant articles and self-studying, all related to the employees work.

HrsWrk refers to the amount of hours an employee works in a week.

ClientSat and SuperSat refer to the satisfaction of the employees work from the clients and superiors respectively.

Lastly, ProjCompl measures the percentage completion of the respective employees’ projects. The dataset comes from Kaggle, and is compiled to represent different aspects of job performance

Data suitability and assumptions

Next we need to investigate if the data is suitable for a factor analysis and if the assumptions are fulfilled. First inspection of the data regarding the suitability is to check the correlation in the data. If the data does not contain sufficient correlations the data is not suitable for factor analysis, since it relies on the interrelationships among variables to identify underlying latent structures. A start is to look at the correlation matrix.

Correlation matrix between all the variables

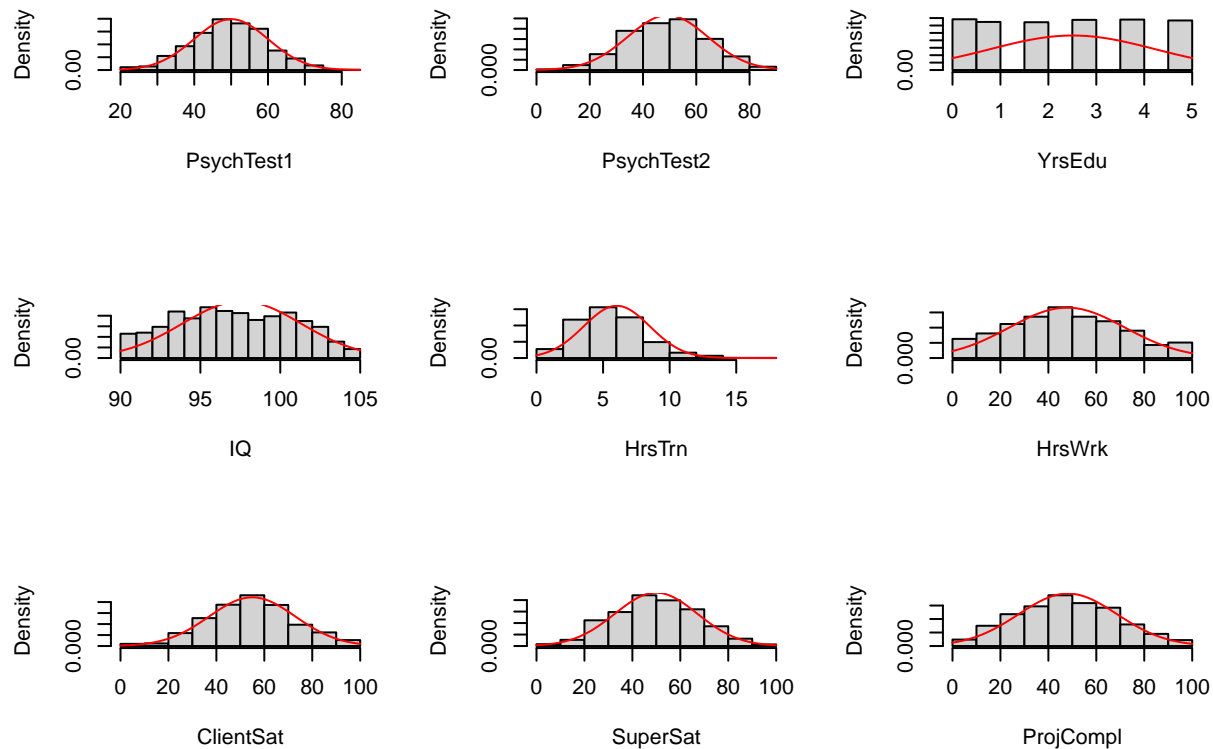
##	PsychTest1	PsychTest2	YrsEdu	IQ	HrsTrn	HrsWrk	ClientSat	SuperSat
## PsychTest1	1.000	0.640	0.049	0.023	-0.034	-0.043	0.876	0.307
## PsychTest2	0.640	1.000	0.034	-0.004	-0.046	-0.067	0.532	0.468
## YrsEdu	0.049	0.034	1.000	0.464	-0.053	-0.051	0.173	0.037
## IQ	0.023	-0.004	0.464	1.000	0.025	0.014	0.084	0.135
## HrsTrn	-0.034	-0.046	-0.053	0.025	1.000	0.816	0.155	0.683
## HrsWrk	-0.043	-0.067	-0.051	0.014	0.816	1.000	0.111	0.823
## ClientSat	0.876	0.532	0.173	0.084	0.155	0.111	1.000	0.391
## SuperSat	0.307	0.468	0.037	0.135	0.683	0.823	0.391	1.000
## ProjCompl	-0.041	-0.068	0.021	0.053	0.791	0.976	0.119	0.809
##	ProjCompl							
## PsychTest1	-0.041							
## PsychTest2	-0.068							
## YrsEdu	0.021							
## IQ	0.053							
## HrsTrn	0.791							
## HrsWrk	0.976							
## ClientSat	0.119							
## SuperSat	0.809							
## ProjCompl	1.000							

When examining the correlation matrix, it is clear that the data have sufficient levels of correlation, with many coefficients exceeding 0.3. Additionally, there seems to be no issue regarding perfect multicollinearity. Furthermore, Bartlett's Test of Sphericity is another way of testing the correlation in the data. Bartlett's Test of Sphericity tests the null that all variables are uncorrelated. Rejecting the null does not necessarily mean that the correlations are good, but it is an indication that the data might be suitable. The result from Bartlett's Test (see appendix) is yet another good indication.

Regarding the assumptions for both exploratory and confirmatory factor analysis it is challenging to test all assumptions in advance, like the assumption of no correlated error terms. The assumption of no correlated error terms for the exploratory factor analysis(EFA) will however to a large degree be indirectly evaluated through the model fit evaluation of the implied model in the confirmatory analysis. Other assumptions are possible to check beforehand like the assumption that the variables must be on an interval or ratio scale. Although it is not technically met due to the inclusion of the variables ClientSat and SuperSat since these are on likert scale. However, likert scale variables are often used in factor analysis

since they tend to approximate interval measurement well enough. The inclusion of years of education and IQ are also a bit questionable regarding directionality. It is unlikely that these variables are outcomes of job performance. Despite that, it is reasonable to assume that these variables contribute to some latent structure like intellect which could be a factor of job performance.

Distribution of the data for each variable with a fitted normal distribution



Based on looking at the histograms, with the exception of YrsEdu and maybe IQ, the variables show distributions that are fairly symmetrical and bell-shaped, which is a good indication of univariate normal distribution for the variables. While a few of the variables are not perfectly symmetrical they do not seem to exhibit too large skewness or kurtosis to the point that it substantially violates the assumption of normality. With the exception of the variables IQ and YearEdu, it would not be unreasonable to assume they are sampled from a normal distribution. The distribution of IQ in our sample is not clearly normal distributed, but it should be sampled from a normal distribution since IQ is normal.

Univariate normal distributions do not guarantee that joint distribution of the variables is multivariate normal.

Method

The main statistical method that will be used in this assignment is a confirmatory factor analysis. CFA is a statistical method in the more general framework structural equation

modeling. CFA is a confirmatory method used to test hypotheses regarding the underlying factors that manifest in the observed indicator variables. That is CFA is used to test the fit of a pre proposed model, often specified by theory. The model that is being tested aims to express the relationship between the observed variables and their underlying latent constructs. In factor analysis the latent structures are measured indirectly through the covariance or correlation among the observed variables to estimate the strength of the relationship between the indicator variables and their corresponding factors.

The choice of using CFA as the method for this assignment is based on the fact that job performance is oftentimes multidimensional, it is for the most part near impossible to measure the performance directly (Some really specific factory jobs might have output as the only metric). Job performance is a latent construct, making it highly suitable for a factor analysis. Furthermore the topic of job performance is a well researched subject where there already are some established theories on what factors impact job performance.

However, the data at hand does not exactly mimic previous theories. Therefore we will also conducting an EFA to aid the formulation of the measurement model. EFA is a method used for identifying underlying structures. Unlike CFA that tests an already existing model, EFA aims to discover patterns within the data.

Once a model is found, modification indices can be used to determine what modification of the model would have the greatest impact. These will be computed using the lavaan package in R. The model refinement process involved evaluating these modifications and only incorporating those that align with the theoretical expectations and improved model fit.

Choosing the measurement model

The measurement model for a confirmatory factor analysis can be created based on either an EFA, theory, empirical evidence, subject-matter knowledge, experience or a combination of these. As stated earlier, there is previous research regarding the topic of job performance. Despite that, there is not a pre-proposed measurement model that is directly applicable to the data at our disposal. However previous theory and subject matter knowledge can still be of aid to help formulate our measurement model. Caillier (2010) researches factors affecting job performance, where attitude is one category of variables. NG and Feldman (2009) use education level as a prediction variable in their research. Neither of the mentioned research performs a CFA but still gives a good indication of what dimension they believe constitutes job performance.

The purpose of this report is to conduct a CFA, not an EFA. However, doing an EFA can still help aid the formulation of the measurement model. As such an EFA was conducted (see appendix for results). Doing EFA on the same data as the CFA is sub-optimal since it will lead to bias. Hence, the data was split at random into two parts one data set for the EFA and one for the CFA. As the EFA is primarily a supplement to the theoretical in constructing the measurement model and the main focus is on the CFA, a 70/30 split was made. Noting that the results from the EFA were not without concerns, a quite clear structures emerge with three factors. The indicator variables HrsTrn, HrsWrk, SuperStat and ProjCompl load very highly on the first factor. The variables PsychTest1, PsychTest2

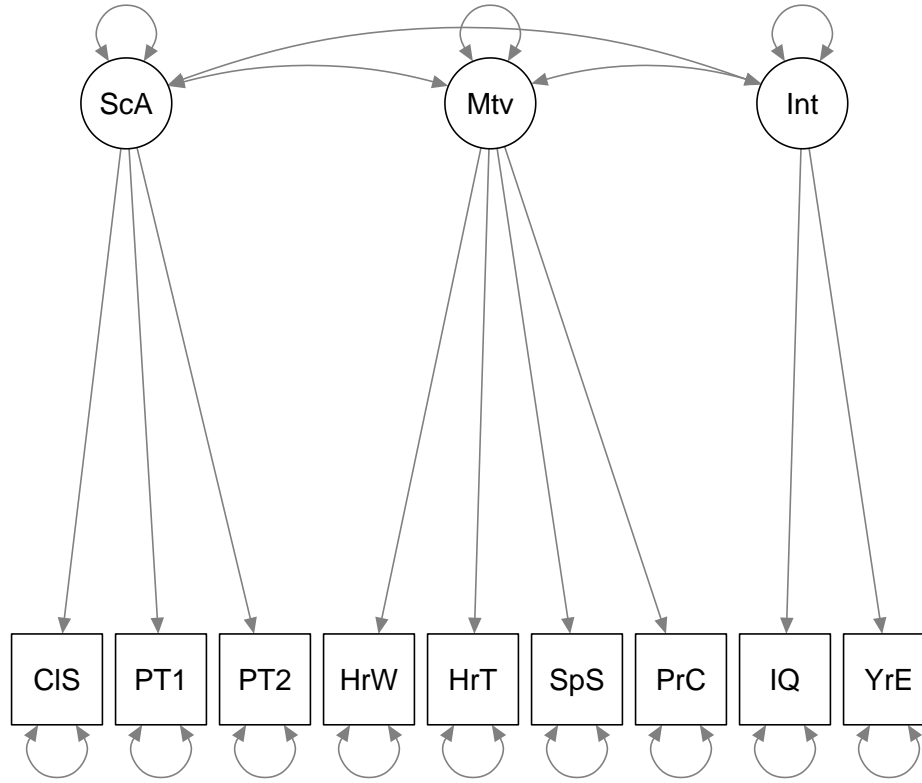
together with ClientSat load the highest on the second factor. And the variables IQ and YsrEdu make up the third factor.

The Suggested model fits quite well with the dimensions suggested from previous research above. The variables IQ and YsrEdu could quite intuitively be indicators of the latent construct of intellect. The same goes for the variables HrsTrn, HrsWrk, SuperStat and ProjCompl, they could likely be indicators of motivation akin to attitude. Assuming that the variables PsychTest1 and PsychTest2 have some association to how social a person is, they together with ClientSat are possible indicators of social ability.

The hypothesis behind our model, and which will be tested is if these grouping of variables belong together in modeling different aspects of job performance. All variables linked to the employee's involvement in the work were grouped into the category "motivation", these variables being the hours put into work, into informings one's self to relevant events, project completion and the satisfaction from the superiors. These all show the time and work an employee dedicates, as well as superior satisfaction being an external metric related to all of these, supposedly. Social ability should be a conceptualized as latent construct in the context of several reasons. Social ability encompasses a range of work place related skills, such as communication, empathy, teamwork, and conflict resolution. lastly intellect as a factor for job performance seems reasonable, as it facilitates for important things like problem solving, learning and application of knowledge.

The goal with testing the measurement model is to test a model that makes sense. Social ability, motivation and intellect are three dimensions of job performance that intuitively seem reasonable. Combined with that fact, the reasoning above, the EFA and previous research our proposed measurement model will be:

Measurement model



Our measurement model consists of three latent variables, social ability (ScA), Motivation (Mtv) and Intellect (Int) all allowed to be correlated. ScA has three observed variables ClientSat (CIS) PsychTest1 (PT1) PsychTest2 (PT2) connected to it, Mtv has the observed variables HrsWrk (HrW), HrsTrn (HrT) SuperSat (SpS) and ProjCompl (PrC). lastly the third latent variable Int has the observed variables IQ and YrsEdu (YrE). No cross loadings and no correlated errors.

Regarding identification. A necessary but unfortunately not sufficient condition is the t-rule, that states that the number unique variance and covariance need to be greater than the number of free parameters in the model. Number of unique variance and covariance:

$$9(9+1)/2 = 45$$

Number of free parameters in the model are 24. Since the $45 > 24$ the t-rule is fulfilled.

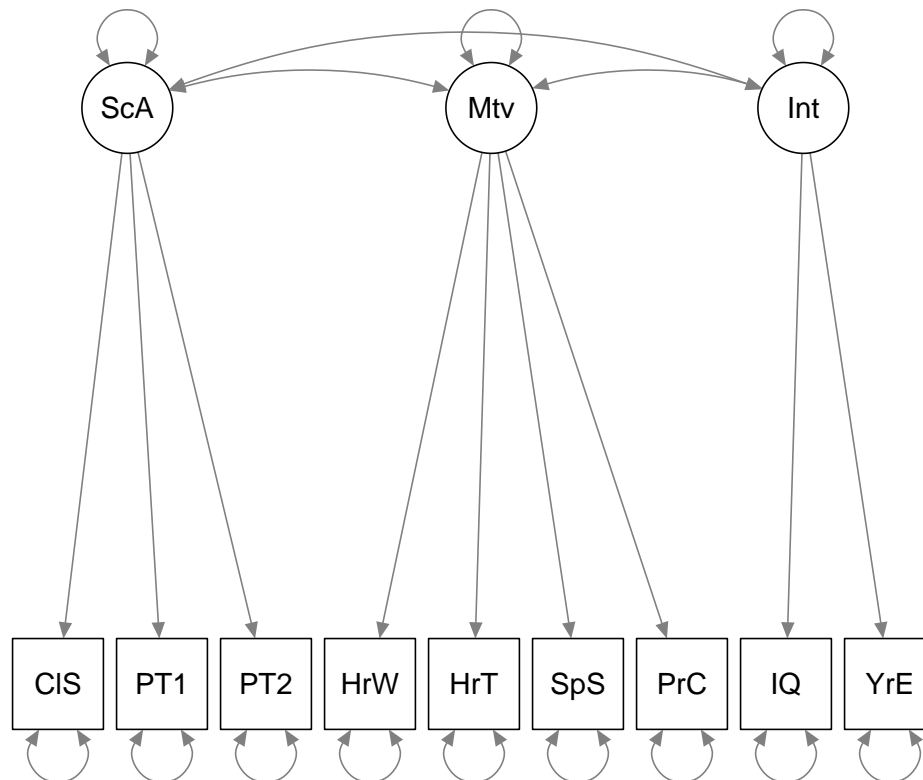
A condition that is sufficient but not necessary is the three-indicator rule. That is the model does not need to fulfill this condition to be identified, but if it does it will be. The three-indicator rule states that all factors has at least three indicator variables, their should be no correlated errors and all indicator variables loads to only one factor. In the measurement model each indicator variable only loads to one factor each, their is no cross loading and assumed their is no correlated errors terms of each indicator variable (these represent the unique variance in the indicator variables that are not explained by the latent factors). However, the intellect factor only have two indicators so the three-indicator rule is not fulfilled. Despite intellect only having two indicator variables, the model seem identified

fulfilling every other condition. Only having two indicator is not enough to conclude that the model is not identifiable, but it could potentially be problematic for assessing model fit.

Results

To start off we had our theoretical model. This model uses all the available variables as described in the previous section:

Figure showing the links between the factors and their respective latent variables:



Fit measures of the model:

##		Metric	Value	Threshold	Comparison
##	chisq	Chi-square/df ratio	82.9323757	<3	Fail
##	cfi	CFI	0.7248258	>0.95	Fail
##	tli	TLI	0.5872388	>0.95	Fail
##	rmsea	RMSEA	0.3421202	<0.06	Fail
##	srmr	SRMR	0.1386905	<0.08	Fail
##	pvalue	Chi-square p-value	0.0000000	>0.05	Fail

To evaluate the fit of our measurement model we need to look at the fit indices, i.e. evaluate if the data supports the proposed model. The chi-square test is a actual statistical test that test the model implied covariance matrix against the sample covariance matrix, neither is actually known but can be estimated from the data. The null hypothesis is the chi squared test is that the model implied covariance matrix equals the sample covariance matrix (typical we want P-value above 0.05). however, the chi-squared test is sensitive to sample size, for small samples, rejecting the null can be difficult even if the model is bad and for large samples

even small departures from the null can be statistically significant. Since the chi-squared test has its problems, other methods for evaluating the model fit are needed, and many different fit indices have been put forward. Unfortunately, many of these tests are not formal statistical tests, and therefore act more of a rule of thumb or suggested cut-off points. Because of the limitations of the chi-squared test and the rule-of-thumb nature of other fit indices, it is important to look at many different fit indices to be able to make a nuanced and holistic evaluation of the model fit. One commonly used fit index that tries to mitigate some of the problems with the chi-squared test is the chi-squared divided by its degrees of freedom, where a cut-off point of less than or equal to 3 has been suggested. Other fit indices that are commonly used are the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI), which are comparative fit measures. They both compare the fit of the measurement model to a baseline model, which assumes no relationships among the variables. Their suggested cut-off point is 0.95. The RMSEA and the SRMR, along with the TLI and CFI, are not associated with any statistical tests, and the suggested cut-off point for RMSEA and SRMR are 0.06 and 0.08 respectively.

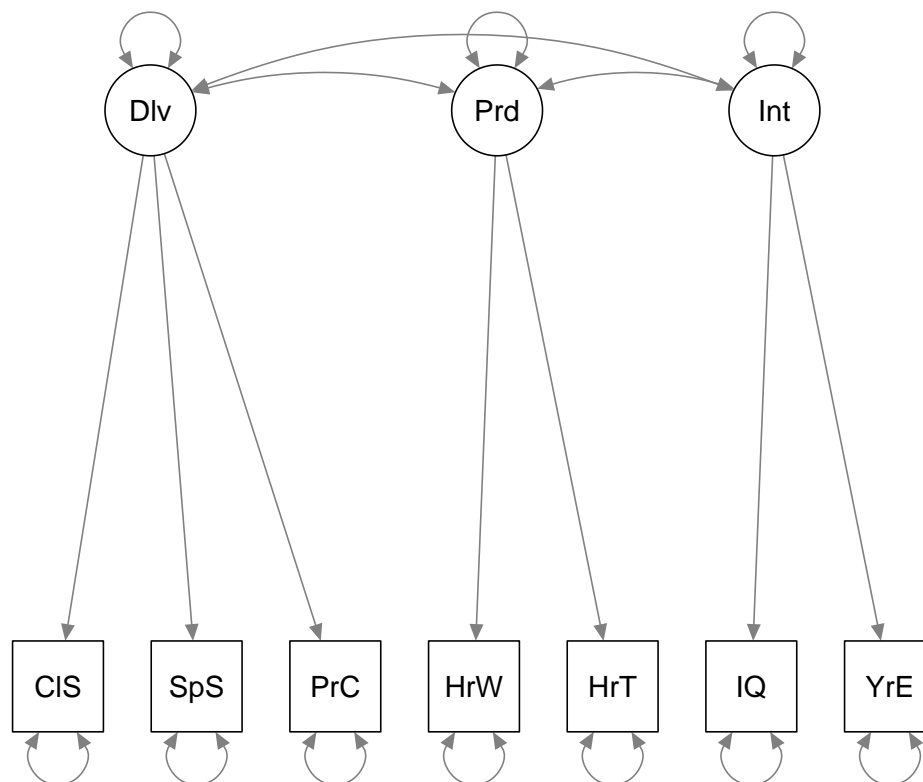
From looking at the fit indices table, it is easy to conclude that the model fit of our measurement model is not a good fit. The p-value for the chi-squared test is close to zero, suggesting a bad fit. From the other non-statistical testing-based fit indices, it is clear that the model does not imply a good fit. Chi-squared/df has a value of 82.93, which is far above the suggested cut-off point of 3 or less. Ideally, we want both the CFI and TLI over 0.95, but unfortunately, they are below that at 0.72 and 0.59 respectively. The same is true for RMSEA and SRMR, both of which indicate a poor fit. These results lead us to conclude that the specified model does not provide a sufficiently good fit to the observed data. This could mean that our theorized measurement model with the three latent variables—Social Ability, Motivation, and Intellect—does not adequately capture the actual structure of job performance.

For the Data, unfortunately, it doesn't pass the test of multivariate normal distribution (see appendix). This violates the assumption of multivariate normal distribution, which mainly affects the reliability of the maximum likelihood estimations of the factor loadings. However, it could still be argued that the data could still be used for factor analysis. While certainty of multivariate normal distribution provides more reliability of the results, not adhering perfectly to the assumptions does not mean that the results are useless; it can still provide valuable insight, just with bigger uncertainty. While not a guarantee, the central limit theorem may also give some leeway regarding the assumption. However, the violation of the normality assumption could still be a factor in the bad fit of the model.

Reworked and adjusted models

Given our initial model showed a quite bad fit, it was of interest to test if there was a model that could explain the latent construct job performance. While we recognize that doing this is not strictly a part of this assignment or suitable for a Confirmatory studies, this subsection does not aim to confirm rather be a supplement to get a better understanding of the data. Testing multiple models is not appropriate for a CFA, but given theoretical justification to ensure the validity of the subsequent models it can still be of use, while not being strictly part of the CFA.

Trying other models that make theoretical sense, we realized that the PsychTest variables were a major source of problems, causing the model to estimate negative variances, again indicating the lack of correct model identification.



Fit measures of the model:

##		Metric	Value	Threshold	Comparison
##	chisq	Chi-square/df ratio	23.40064387	3.00	Fail
##	cfi	CFI	0.94055388	0.95	Fail
##	tli	TLI	0.88651195	0.95	Fail
##	rmsea	RMSEA	0.17888801	0.06	Fail
##	srmr	SRMR	0.06966017	0.08	Pass

Number of unique variance and covariance:

$$7(7+1)/2 = 28$$

Number of free parameters in the model are 20. Since the $28 > 20$ the t-rule is fulfilled. The only part of the 3 indicator rule not fulfilled is that each latent variable has 3 factors, this is not necessary however to make the model identifiable.

The latent factor “Delivery” consists of client and superior satisfaction, in addition to percentage project completion. This variable encompasses how well the work the employee does is delivered. It is expected that these 2 variables are correlated, with more projects being completed leads to more satisfied customers and superiors.

The Productivity latent factor relates to the amount of time an employee puts into work, in addition to the time spent educating oneself outside of work, hence also providing valuable time towards relevant work projects.

The Intellect latent factor is still kept unchanged since the initial model. This one is composed of the variables that have to do with intelligence and education, with the IQ and Years of Education variables respectively. Years of education does not have to do with intelligence directly, they are however related with higher intellect perhaps meaning more education due to having an easier time.

The CFI of this model to the data is 0.940, which is notably better and by some might be deemed as a good fit, but there are ways of improving the model.

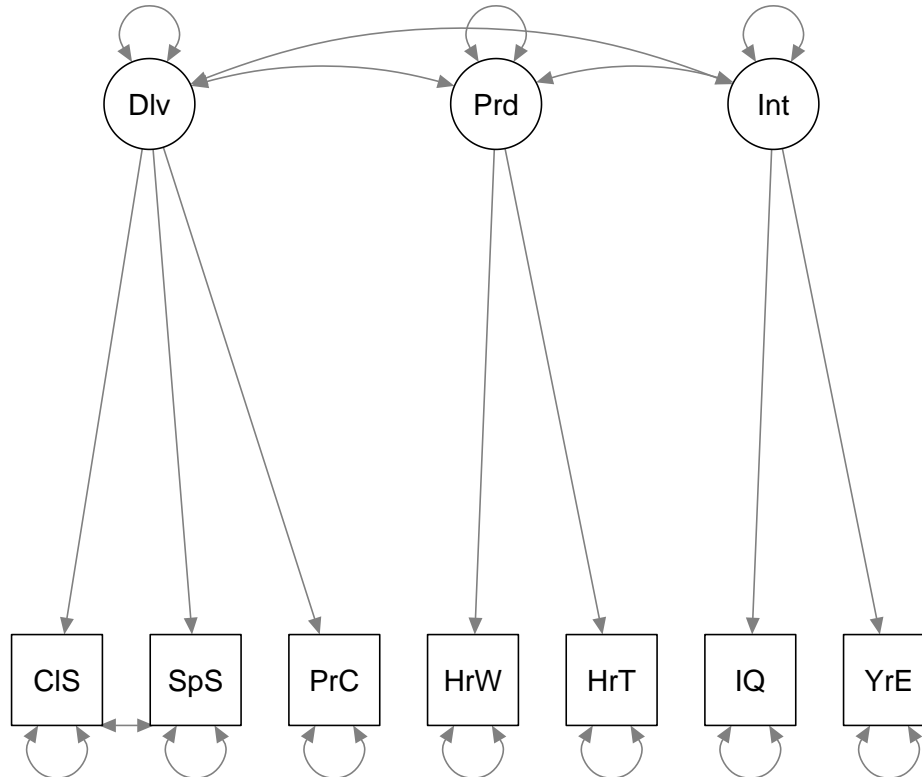
One such way is modindices, these show the changes that would most improve the model. To include such a change there needs to be a logical / theoretical reason as to why, as described in the method section. Fixing the correlation of 2 variables will always reduce the degrees of freedom when estimating the model, however there is a need for justification for these restrictions on the model.

Printed below are the 3 largest of these modindices. The first, and most impactful one suggests restricting the model by setting the satisfactions of the clients and superiors to covary. In the real world this would make sense, a completed and delivered project that also satisfies the client has reason to satisfy the superiors in the company as well.

##		lhs	op	rhs	mi	epc	sepc.lv	sepc.all	sepc.nox
## 35		ClientSat	~~	SuperSat	188.743	0.288	0.288	0.519	0.519
## 25		Productivity	=~	ClientSat	27.917	-2.495	-2.495	-2.497	-2.497
## 30		Intellect	=~	ClientSat	24.168	0.213	0.213	0.213	0.213

Adding this restriction to our model leads us to the final model:

Figure showing the links between the factors and their respective latent variables:



Fit measures of the new model:

##	Metric	Value	Threshold	Comparison
## chisq	Chi-square/df ratio	6.04673528	3.00	Fail
## cfi	CFI	0.98782467	0.95	Pass
## tli	TLI	0.97443180	0.95	Pass
## rmsea	RMSEA	0.08490949	0.06	Fail
## srmr	SRMR	0.04568958	0.08	Pass

Number of unique variance and covariance:

$$7(7+1)/2 = 28$$

Number of free parameters in the model are 21. Since the $28 > 21$ the t-rule is fulfilled. The only part of the 3 indicator rule not fulfilled is that each latent variable has 3 factors, this is not necessary however to make the model identifiable.

This model has a CFI of 0.988, which is considered a good fit to the data. Leaving out the PsychTest variables made the model identifiable, and with slight modification, it has good fit too.

Table 1: Factor Loadings

Factor	Delivery	Productivity	Intellect
Superior Satisfaction	0.83	0.00	0.00
Client Satisfaction	0.12	0.00	0.00
Project Completion	0.97	0.00	0.00
Hours of Work	0.00	1.01	0.00
Hours of Personal Culture	0.00	0.83	0.00
IQ	0.00	0.00	0.58
Years of Education	0.00	0.00	0.80

Regarding the factor loadings, Delivery loads project completion and Superior Satisfaction at 0.99 and 0.83 respectively, while only 0.12 for client satisfaction. This likely indicates that this factor measures the delivery of work and its effect within the company, through the satisfaction it creates for the superiors. The loading for the client's satisfaction is less important in this case.

Productivity loads hours of work and hours of personal culture at 1.01 and 0.83 respectively, it is odd to have hours of work load at 1.01, but could be due to an error in how the data was measured. Many models were tried and hours of work had some odd behavior, in this model it acted the most normal.

Intellect loads IQ at 0.58 and years of education at 0.80.

More interestingly, the covariances. Between the latent variables we have Delivery with Productivity covarying 0.99, while the other 2 covariances between Delivery-Intellect, and Productivity-Intellect both being around, with high p-values.

Method criticism

Throughout this report some concerns have been raised. The assumption of multivariate normal distributed data can not be said to be fulfilled, which effects the reliability of the maximum likelihood estimations. Transformation of the data could have been an option, but no transformations were made due to the fact that there were no transformations that we could find that significantly improved the distribution to the point that the assumption was fulfilled. Regarding the EFA, the output of the EFA made up some of the basis for the first measurement model. But as also mentioned earlier, there were some concerns regarding the data and some variables suitability. A approach that could have been appropriate wher to redo the EFA with changes accordingly, since EFA can be an iterative processes. This was not done on the basis that the focus was on the CFA and the EFA was only an aid. Some concerns can be raised about the data, Since the data is most likely created to be used in such exercises there may be no actual underlying link like we are looking for, especially if the data was created in a specific way as to uncover a different type of analysis.

we recognize that testing more than our original proposed model is not suitable for an CFA, since it leans in on a exploratory approach and can to be a cause of bias and overfitting. But since there were still strong theoretical underpinnings in the models tested, we hope it can be a way to understand the data.

Despite some concerns regarding the data and the methodology we do not believe that it renders our analysis useless. But it is important to keep in mind when interpreting the results. The analysis still gives insight to the latent structures of job performance but with higher uncertainty, which means some caution should be had when interpreting the results.

Conclusions:

Seeing as the covariances are very extreme, both high and low, the only clear cut conclusions that can be drawn regarding those is that Delivery of work has a strong relationship with Productivity, meaning how much time one puts into work and spends time related to work. Delivery and Intellect may simply not have any strong relationship which might be the case according to these results.

Productivity and Intellect do have a negative correlation which perhaps could be seen as people with more intellect need less time for their work due to deficiency, however due to the high p-value, this conclusion is not statistically significant at a 5% alpha level.

It does make sense that delivery and Productivity covary, someone who is productive likely uses their time to deliver on project. Intellect as a factor in this analysis has very little correlation to Productivity and Delivery, both of which are different metrics for job performance. The conclusion that can be drawn is that it doesn't matter how smart someone is, it is the time and work they bring that determines results and overall job performance.

Sources

Dinno, A. (2009). Exploring the sensitivity of Horn's parallel analysis to the distributional form of simulated data. *Multivariate Behavioral Research*, 44, 362-388

James Gerard Caillier (2010) Factors Affecting Job Performance in Public Agencies, *Public Performance & Management Review*, 34:2, 139-165, DOI: 10.2753/ PMR1530-9576340201

NG, Thomas W., H. Feldman, Daniel L. C. (2009) HOW BROADLY DOES EDUCATION CONTRIBUTE TO JOB PERFORMANCE?, *Personnel psychology*, 2009-03, Vol.62 (1), p.89-134.

Data: <https://www.kaggle.com/datasets/michealronn/job-performance?fbclid=IwAR117k90QAHG176gsnIRr8VQgbaGFASLLiAlA4mZcF94wYXRrZBTolveLYo>

Appendix

Results Bartlett's Test of Sphericity

```
cortest.bartlett(data)
```

```
## R was not square, finding R from data
```

```
## $chisq
## [1] 10221.09
##
## $p.value
## [1] 0
##
## $df
## [1] 36
```

MvN test of the data:

```
## $multivariateNormality
##      Test      H      p value MVN
## 1 Royston 200.5998 5.316452e-40 NO
##
## $univariateNormality
##      Test      Variable Statistic      p value Normality
## 1 Anderson-Darling PsychTest1      0.6456 0.0919      YES
## 2 Anderson-Darling PsychTest2      0.8851 0.0236      NO
## 3 Anderson-Darling YrsEdu      28.6505 <0.001      NO
## 4 Anderson-Darling IQ      7.4032 <0.001      NO
## 5 Anderson-Darling HrsTrn      10.1394 <0.001      NO
## 6 Anderson-Darling HrsWrk      1.6349 3e-04      NO
## 7 Anderson-Darling ClientSat      0.4246 0.3167      YES
## 8 Anderson-Darling SuperSat      0.5391 0.1666      YES
## 9 Anderson-Darling ProjCompl      1.0041 0.012      NO
##
## $Descriptives
##      n      Mean      Std.Dev Median Min Max 25th 75th      Skew
## PsychTest1 1000 49.990 10.049920      50 20 81 43.00 57 -0.096469045
## PsychTest2 1000 49.286 15.019380      50 5 90 38.75 60 -0.038935412
## YrsEdu      1000 2.506 1.716074      3 0 5 1.00 4 -0.017305321
## IQ          1000 97.587 3.719175      97 90 105 95.00 101 -0.009282197
## HrsTrn      1000 6.028 2.484228      6 0 17 4.00 8 0.597744518
## HrsWrk      1000 47.810 24.080568      47 0 100 30.00 64 0.145971328
## ClientSat   1000 54.970 18.002922      55 0 100 43.00 67 -0.052232046
## SuperSat    1000 49.912 17.190747      50 0 100 38.00 62 0.027048803
## ProjCompl   1000 48.127 20.541052      48 0 100 33.00 63 0.079392150
```

```

##          Kurtosis
## PsychTest1 -0.08159950
## PsychTest2 -0.32272972
## YrsEdu      -1.28123571
## IQ          -0.91253895
## HrsTrn       0.71628778
## HrsWrk      -0.58942696
## ClientSat   -0.09460311
## SuperSat    -0.33567096
## ProjCompl   -0.53664978

## $multivariateNormality
##          Test          HZ p value MVN
## 1 Henze-Zirkler 1.124734          0 NO
##
## $univariateNormality
##          Test Variable Statistic  p value Normality
## 1 Anderson-Darling PsychTest1    0.6456 0.0919      YES
## 2 Anderson-Darling PsychTest2    0.8851 0.0236      NO
## 3 Anderson-Darling YrsEdu       28.6505 <0.001      NO
## 4 Anderson-Darling IQ           7.4032 <0.001      NO
## 5 Anderson-Darling HrsTrn       10.1394 <0.001      NO
## 6 Anderson-Darling HrsWrk        1.6349 3e-04      NO
## 7 Anderson-Darling ClientSat     0.4246 0.3167      YES
## 8 Anderson-Darling SuperSat      0.5391 0.1666      YES
## 9 Anderson-Darling ProjCompl     1.0041 0.012       NO
##
## $Descriptives
##          n   Mean   Std.Dev Median Min Max 25th 75th          Skew
## PsychTest1 1000 49.990 10.049920    50 20 81 43.00 57 -0.096469045
## PsychTest2 1000 49.286 15.019380    50 5 90 38.75 60 -0.038935412
## YrsEdu      1000 2.506 1.716074     3 0 5 1.00 4 -0.017305321
## IQ          1000 97.587 3.719175    97 90 105 95.00 101 -0.009282197
## HrsTrn      1000 6.028 2.484228     6 0 17 4.00 8 0.597744518
## HrsWrk      1000 47.810 24.080568    47 0 100 30.00 64 0.145971328
## ClientSat   1000 54.970 18.002922    55 0 100 43.00 67 -0.052232046
## SuperSat    1000 49.912 17.190747    50 0 100 38.00 62 0.027048803
## ProjCompl   1000 48.127 20.541052    48 0 100 33.00 63 0.079392150
##
##          Kurtosis
## PsychTest1 -0.08159950
## PsychTest2 -0.32272972
## YrsEdu      -1.28123571
## IQ          -0.91253895
## HrsTrn       0.71628778
## HrsWrk      -0.58942696
## ClientSat   -0.09460311
## SuperSat    -0.33567096

```

ProjCompl -0.53664978

\$multivariateNormality

##	Test	Statistic	p value	Result
## 1	Mardia Skewness	221.737326004292	0.00212346585273433	NO
## 2	Mardia Kurtosis	-4.46689474861159	7.93631505069392e-06	NO
## 3	MVN	<NA>	<NA>	NO

##

\$univariateNormality

##	Test	Variable	Statistic	p value	Normality
## 1	Anderson-Darling	PsychTest1	0.6456	0.0919	YES
## 2	Anderson-Darling	PsychTest2	0.8851	0.0236	NO
## 3	Anderson-Darling	YrsEdu	28.6505	<0.001	NO
## 4	Anderson-Darling	IQ	7.4032	<0.001	NO
## 5	Anderson-Darling	HrsTrn	10.1394	<0.001	NO
## 6	Anderson-Darling	HrsWrk	1.6349	3e-04	NO
## 7	Anderson-Darling	ClientSat	0.4246	0.3167	YES
## 8	Anderson-Darling	SuperSat	0.5391	0.1666	YES
## 9	Anderson-Darling	ProjCompl	1.0041	0.012	NO

##

\$Descriptives

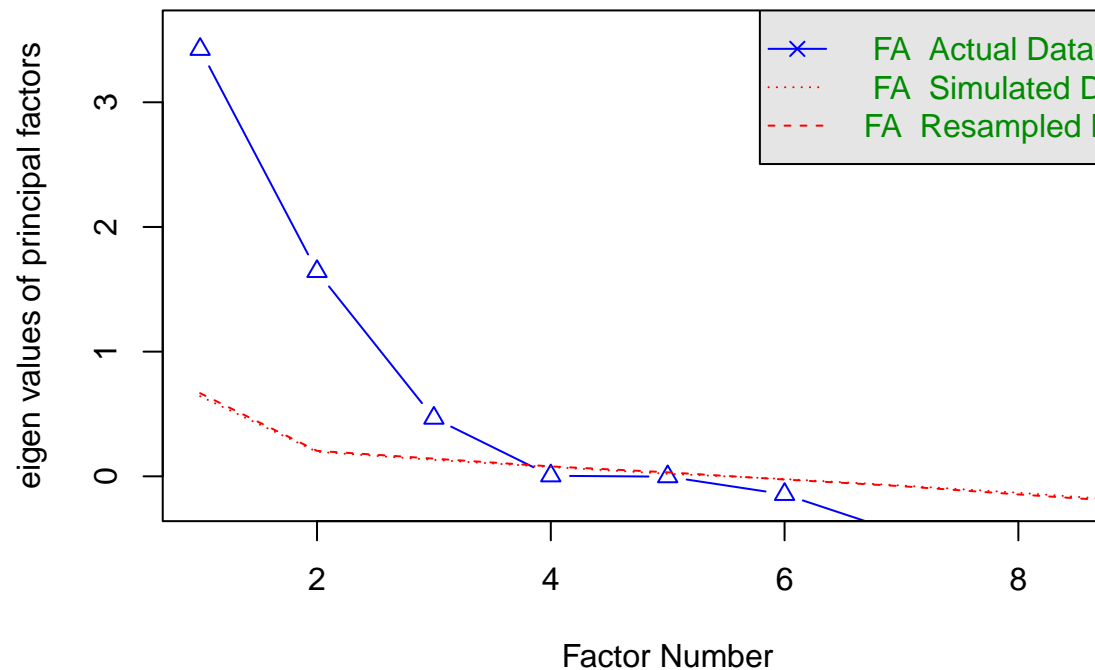
##		n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew
##	PsychTest1	1000	49.990	10.049920	50	20	81	43.00	57	-0.096469045
##	PsychTest2	1000	49.286	15.019380	50	5	90	38.75	60	-0.038935412
##	YrsEdu	1000	2.506	1.716074	3	0	5	1.00	4	-0.017305321
##	IQ	1000	97.587	3.719175	97	90	105	95.00	101	-0.009282197
##	HrsTrn	1000	6.028	2.484228	6	0	17	4.00	8	0.597744518
##	HrsWrk	1000	47.810	24.080568	47	0	100	30.00	64	0.145971328
##	ClientSat	1000	54.970	18.002922	55	0	100	43.00	67	-0.052232046
##	SuperSat	1000	49.912	17.190747	50	0	100	38.00	62	0.027048803
##	ProjCompl	1000	48.127	20.541052	48	0	100	33.00	63	0.079392150

##

Kurtosis

##	PsychTest1	-0.08159950
##	PsychTest2	-0.32272972
##	YrsEdu	-1.28123571
##	IQ	-0.91253895
##	HrsTrn	0.71628778
##	HrsWrk	-0.58942696
##	ClientSat	-0.09460311
##	SuperSat	-0.33567096
##	ProjCompl	-0.53664978

Parallel Analysis Scree Plots



Screeplot of the EFA data

Parallel analysis suggests that the number of factors = 3 and the number of components = 3

EFA result, with the oblique factor rotation promax and standardization of the variance of the factors.

```
##
## Loadings:
##           ML2    ML1    ML3
## HrsTrn      0.855  0.003 -0.161
## HrsWrk      0.996 -0.067 -0.113
## SuperSat    0.891  0.322  0.216
## ProjCompl   0.980 -0.048 -0.140
## PsychTest1 -0.001  0.894  0.033
## PsychTest2 -0.003  0.733  0.666
## ClientSat   0.153  0.979 -0.189
## YrsEdu     -0.085  0.144 -0.143
## IQ          0.028  0.008 -0.106
##
##           ML2    ML1    ML3
## SS loadings  3.51  2.426  0.617
## Proportion Var 0.39  0.270  0.069
## Cumulative Var 0.39  0.660  0.728
```

There is some concern regarding the factor loadings. Both the variables IQ and YrsEdu have quite low loading on all three factors, suggesting that they might be candidates for removal. Removing IQ and YrsEdu might also alleviate some problems with normality. Secondly, the PsychTest2 variable loads highly on both factors two and three making it in contention for removal. Adding cross factor loading could be fine if strongly motivated by theory, however it could cause problems with identification. Ideally each indicator variable loads highly on a singular factor. Lastly regarding the loadings, all associated variables of the first factor loads quite highly. The factor loading of HrsWrk is close to one, which would suggest a near perfect linear relationship with the factor.