Questionnaire Evaluation with Factor Analysis and Cronbach's Alpha

An Example

- Melanie Hof -

1. Introduction

The pleasure writers experience in writing considerably influences their motivation and consequently their writing performance (Hayes, 1996). Low-motivated writers perform worse, since they spend less time on a writing task, are less engaged in a writing task, study less thoroughly instructional material, and are less willing to attend writing training sessions.

Although someone's pleasure in writing needs to be taken into account in order to draw valid conclusions about the factors influencing someone's writing performance, it is rather difficult to measure such attitude (O'Keefe, 2002). We cannot look into writers' minds. We just can ask them to externalize the attitude we are interested in, but then we probably do not get a truthful answer (Thurstone, 1977). Assuming that a writing researcher highly values writing, writers probably present their attitude more positively than it is. Besides, if we just ask writers whether they like writing or not, we cannot get insight in the aspects that are related to that attitude (O'Keefe, 2002). For example, the amount of writing experience they have. Since writing is less laborious when you have a lot of experience, highly experienced writers generally like it more (Hayes, 1996).

To avoid socially preferred answers and be able to receive information about an attitude and aspects related to an attitude, researchers prefer the use of questionnaires asking for a person's degree of agreement with evaluative statements about the object of attitude and related aspects (O'Keefe, 2002). However, the use of such method does not necessarily mean that reliable and valid indications of someone's attitude can be obtained. In the end, some items can measure a completely different construct than the attitude of interest (Ratray & Jones, 2007).

In this paper, two statistical methods are discussed extensively with which the validity and reliability of a questionnaire measuring an attitude and attitude related aspects can be tested: exploratory factor analysis and Cronbach's alpha (Bornstedt, 1977; Ratray & Jones, 2007). To show how these tests should be conducted and the results interpreted, a questionnaire used to determine Dutch seventh graders' pleasure in writing will be evaluated.

2. Data

In the school years 2010-2011 and 2011-2012, the Centre for Language, Education and Communication of the University of Groningen has conducted an experiment to test whether writing instruction in secondary school content courses improves the writing skills, writing attitude and content knowledge of seventh graders. 114 Seventh graders received instruction in writing an expository text in the Dutch class. Afterwards, the 57 seventh graders in the experimental group wrote that genre three times in the history and three times in the science class. The 57 seventh graders in the control group followed the normal procedure in the history and science classes: making workbook exercises about the topics of interest. Before and after the intervention, all seventh graders had to write an expository text about the same subject to test their skill change, had to do a content knowledge exam to test their knowledge change, and had to fill in a questionnaire to test their attitude change. The effects of the intervention could be determined by comparing the skill, knowledge and attitude changes of both groups.

p01	I love writing.
p02	Writing is my favorite school subject.
p03	When I write, I feel well.
p04	I hate writing.
p05	I write as soon as I get the chance.
p06	I make sure that I have to write as less as possible.
p07	I write more than my class mates.
p08	When I write, I prefer to do something different.
p09	Writing gives me pleasure.
p10	I just write, when I can get a good grade for it.
p11	Writing is boring.
p12	I like different kinds of writing.
p13	When I have the opportunity to determine on my own what I do in the Dutch class, I usual do a writing task.
p14	I write even if the teacher does not assign a writing task.
p15	I would like to spend more time on writing.
p16	Writing is a waste of time.
p17	I always look forward to writing lessons.
p18	I write because I have to at school.
p19	I like it to write down my thoughts.
p20	I would like to write more at school.

Table 1. Questionnaire to seventh' graders pleasure in writing

The attitude measure focused on two attitudes: a writer's self-efficacy in writing and a writer's pleasure in writing. Participants had to indicate their level of agreement with 40 evaluative statements about writing on a 5-point Likert scale (ranging from strongly agree to strongly disagree). The level of agreement with the first 20 statements revealed participants' self-efficacy, the level of agreement with the other 20 statements participants' level of pleasure in writing. 7 of each 20 items were formulated negatively instead of positively to force students to evaluate every statement in its own right. When all items are formulated in the same direction, people seem to evaluate them equally (Ratray & Jones, 2007). In this paper, just the reliability and validity check of the second part of the questionnaire - students' pleasure in writing - is discussed (for the questionnaire see Table 1).

3. Factor analysis

With factor analysis, the construct validity of a questionnaire can be tested (Bornstedt, 1977; Ratray & Jones, 2007). If a questionnaire is construct valid, all items together represent the underlying construct

well. Hence, one's total score on the twenty items of the questionnaire of interest should represent one's pleasure in writing correctly. Exploratory factor analysis detects the constructs - i.e. factors - that underlie a dataset based on the correlations between variables (in this case, questionnaire items) (Field, 2009; Tabachnik & Fidell, 2001; Rietveld & Van Hout, 1993). The factors that explain the highest proportion of variance the variables share are expected to represent the underlying constructs. In contrast to the commonly used principal component analysis, factor analysis does not have the presumption that all variance within a dataset is shared (Costello & Osborne, 2005; Field, 2009; Tabachnik & Fidell, 2001; Rietveld & Van Hout, 1993). Since that generally is not the case either, factor analysis is assumed to be a more reliable questionnaire evaluation method than principal component analysis (Costello & Osborne, 2005).

3.1. Prerequisites

In order to conduct a reliable factor analysis the *sample size* needs to be big enough (Costello & Osborne, 2005; Field, 2009; Tabachnik & Fidell, 2001). The smaller the sample, the bigger the chance that the correlation coefficients between items differ from the correlation coefficients between items in other samples (Field, 2009). A common rule of thumb is that a researcher at least needs 10-15 participants per item. Since the sample size in this study is 114 instead of the required 200-300, we could conclude that a factor analysis should not be done with this data set. Yet, it largely depends on the proportion of variance in a dataset a factor explains how large a sample needs to be. If a factor explains lots of variance in a dataset, variables correlate highly with that factor, i.e. load highly on that factor. A factor with four or more loadings greater than 0.6 "is reliable regardless of sample size." (Field, 2009, p. 647). Fortunately, we do not have to do a factor analysis in order to determine whether our sample size is adequate, the Kaiser-Meyer-Okin measure of sampling adequacy (KMO) can signal in advance whether the sample size is large enough to reliably extract factors (Field, 2009). The KMO "represents the ratio of the squared correlation between variables to the squared partial correlation between variables." (Field, 2009, p. 647). When the KMO is near 0, it is difficult to extract a factor, since the amount of variance just two variables share (partial correlation) is relatively large in comparison with the amount of variance two variables share with other variables (correlation minus partial correlation). When the KMO is near 1, a factor or factors can probably be extracted, since the opposite pattern is visible. Therefore, KMO "values between 0.5 and 0.7 are mediocre, values between 0.7 and 0.8 are good, values between 0.8 and 0.9 are great and values above 0.9 are superb." (Field, 2009. p. 647). The KMO value of this dataset falls within the last category (KMO=0.922).

Another prerequisite for factor analysis is that the variables are measured at an *interval level* (Field, 2009). A Likert scale is assumed to be an interval scale (Ratray & Jones, 2007), although the item scores are discrete values. That hinders the check of the next condition: the data should be *approximately normally distributed* to be able to generalize the results beyond the sample (Field, 2009) and to conduct a maximum likelihood factor analysis to determine validly how
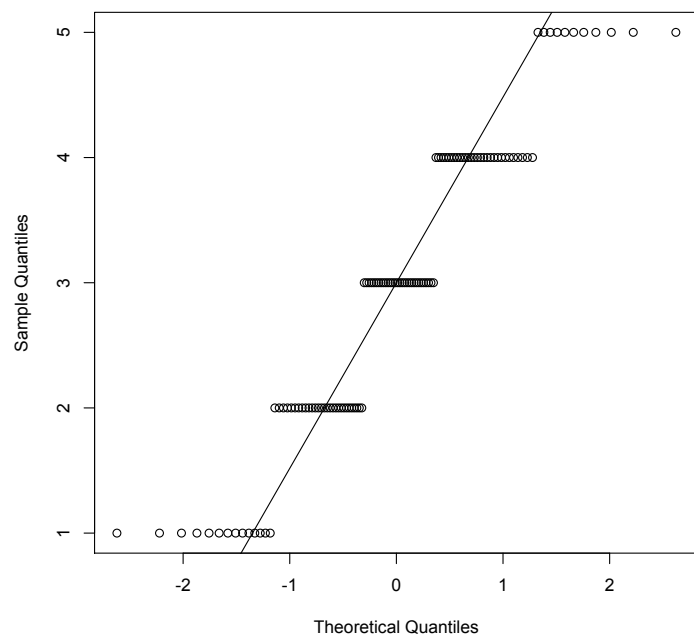

Figure 1. Q-Q plot of the variable p06: 'I make sure that I have to write as less as possible.'

3

many factors underlie the dataset (see §3.3.1) (Costello & Osborne, 2005). Normality tests seem not to be able to test normality of distribution in a set of discrete data. The normality tests signal non-normality of distribution in this dataset by rendering p-values far lower than 0.05, although we can see a pattern of normal distribution in the Q-Q plots with a bit of fantasy. For example, the sixth item in the questionnaire "I make sure that I have to write as less as possible." is far from normally distributed according to the Shapiro-Wilk test (W = 0.9136, p = 1.777e-06), but does seem normally distributed in Figure 1. The exact values are centered in five groups, of which the centers are on the Q-Q line. Most points are centered at the middle of the line, there are a bit less points with values of 2 or 4, and very few points with values of 1 or 5. Based on the Q-Q plots, I concluded that the dataset is approximately normally distributed, and therefore usable in a maximum likelihood factor analysis and generalizable to the population of Dutch seventh graders. If we want to generalize for a larger population, we need to conduct the same survey among other (sub)groups in the population as well (Field, 2009).

The final step before a factor analysis can be conducted is generating the correlation matrix and checking whether *the variables do not correlate too highly or too lowly with other variables* (Field, 2009). If variables correlate too highly (r > 0.8 or r < -.8), "it becomes impossible to determine the unique contribution to a factor of the variables that are highly correlated." (Field, 2009, p. 648). If a variable correlates lowly with many other variables (-0.3 < r < 0.3), the variable probably does not measure the same underlying construct as the other variables. Both the highly and lowly correlating items should be eliminated. As can be seen in Table 2, none of the questionnaire items correlates too highly with other items, but some correlate too lowly with several other items. That does not necessarily mean that the items should be eliminated: the variables with which they do not correlate enough could constitute another factor. There is one objective test to determine whether the items do not correlate too lowly: Barlett's test. However, that test tests a very extreme case of non-correlation: all items of the questionnaire do not correlate with any other item. If the Barlett's test gives a significant result, we can assume that the items correlate anyhow, like in this data set: $\chi2$ (190) = 1263.862, p = 7.117332e-158. Since the Barlett's test gives a significant result and the items correlate at most with a third of the items too lowly, items were not excluded before the factor analysis was conducted.

	p01	p02	p03	p04	p05	p06	p07	p08	p09	p10	p11	p12	p13	p14	p15	p16	p17	p18	p19	p20
p01	1.00	0.67	0.61	-0.56	0.41	-0.39	0.37	-0.46	0.62	-0.37	-0.52	0.43	0.49	0.57	0.47	-0.41	0.60	-0.47	0.55	0.52
p02	0.66	1.00	0.62	-0.42	0.35	-0.30	_0.25_	-0.34	0.58	_-0.20_	-0.34	0.42	0.47	0.53	0.50	_-0.29_	0.61	-0.30	0.50	0.59
p03	0.61	0.62	1.00	-0.52	0.47	-0.34	0.43	-0.39	0.59	-0.33	-0.40	0.51	0.49	0.60	0.60	-0.39	0.69	-0.40	0.51	0.57
p04	-0.56	-0.42	-0.52	1.00	_-0.22_	0.51	-0.32	0.47	-0.47	0.46	0.61	-0.44	-0.31	-0.31	-0.35	0.67	-0.43	0.54	-0.36	-0.44
p05	0.41	0.35	0.47	_-0.22_	1.00	_-0.22_	0.48	-0.34	0.54	_-0.26_	_-0.27_	_0.24_	0.33	0.46	0.48	_-0.21_	0.50	_-0.23_	_0.25_	0.43
p06	-0.39	-0.30	-0.34	0.51	_-0.22_	1.00	-0.34	0.56	-0.32	0.44	0.46	-0.31	_-0.28_	-0.30	-0.38	0.54	-0.40	0.45	-0.40	-0.40
p07	0.37	_0.25_	0.43	-0.32	0.48	-0.34	1.00	-0.38	0.35	_-0.25_	_-0.24_	_0.27_	_0.22_	0.34	0.47	-0.31	0.52	_-0.20_	_0.29_	0.41
p08	-0.46	-0.34	-0.40	0.47	-0.34	0.56	-0.38	1.00	-0.42	0.46	0.43	_-0.26_	_-0.22_	-0.41	-0.38	0.46	-0.40	0.47	-0.39	-0.39
p09	0.62	0.58	0.60	-0.47	0.55	-0.32	0.35	-0.42	1.00	-0.45	-0.41	0.55	0.46	0.61	0.54	-0.34	0.64	-0.45	0.54	0.62
p10	-0.37	_-0.21_	-0.34	0.46	_-0.26_	0.44	_-0.25_	0.46	-0.45	1.00	0.40	_-0.26_	_-0.22_	-0.40	-0.33	0.42	-0.38	0.63	_-0.29_	-0.43
p11	-0.53	-0.39	-0.41	0.61	_-0.27_	0.46	_-0.24_	0.43	-0.41	0.40	1.00	-0.38	_-0.25_	-0.35	_-0.24_	0.47	-0.33	0.45	_-0.25_	-0.39
p12	0.42	0.42	0.51	-0.44	_0.24_	-0.31	_0.27_	_-0.26_	0.55	_-0.27_	-0.38	1.00	0.38	0.49	0.40	-0.38	0.44	_-0.18_	0.45	0.41
p13	0.49	0.48	0.47	-0.31	0.33	_-0.28_	_0.22_	_-0.22_	0.46	_-0.22_	_-0.25_	0.38	1.00	0.50	0.45	_-0.23_	0.54	_-0.18_	0.51	0.59
p14	0.57	0.53	0.60	-0.31	0.46	-0.30	0.34	-0.41	0.61	-0.39	-0.35	0.49	0.50	1.00	0.51	_-0.28_	0.61	-0.38	0.45	0.56
p15	0.47	0.50	0.60	-0.35	0.48	-0.37	0.47	-0.38	0.54	-0.33	-0.24	0.40	0.45	0.51	1.00	-0.44	0.70	_-0.28_	0.49	0.59
p16	-0.41	_-0.29_	-0.39	0.67	_-0.21_	0.54	-0.31	0.46	-0.34	0.42	0.47	-0.38	_-0.23_	_-0.28_	-0.44	1.00	-0.39	0.53	-0.38	-0.40
p17	0.60	0.61	0.69	-0.43	0.50	-0.40	0.52	-0.39	0.64	-0.38	-0.33	0.44	0.54	0.61	0.70	-0.39	1.00	-0.35	0.59	0.75
p18	-0.47	-0.30	-0.40	0.53	_-0.23_	0.45	_-0.20_	0.47	-0.45	0.63	0.45	-0.42	_-0.18_	-0.38	_-0.28_	0.53	-0.35	1.00	-0.35	-0.40
p19	0.55	0.51	0.51	-0.36	_0.25_	-0.40	_0.29_	-0.39	0.54	-0.30	_-0.25_	0.45	0.51	0.45	0.49	-0.38	0.59	-0.35	1.00	0.56
p20	0.52	0.59	0.57	-0.44	0.43	-0.40	0.41	-0.39	0.63	-0.44	-0.40	0.41	0.59	0.56	0.59	-0.40	0.75	-0.40	0.56	1.00

Table 2. Correlation matrix of the dataset. The underlined correlations are too low (-0.3 < r < 0.3).

3.2. The factor analysis

3.2.1. Factor extraction

At heart of factor extraction lies complex algebra with the correlation matrix, it reaches beyond the scope of this paper to explain that comprehensively and in full detail. I refer the interested reader to chapter 6 and 7 of *Statistical Techniques for the Study of Language and Language Behaviour* (Rietveld & Van Hout, 1993).

The algebraic matrix calculations finally end up with eigenvectors (Field, 2009; Tabachnik & Fidell, 2001; Rietveld & Van Hout, 1993). As can be seen in Figure 2, these eigenvectors are linear representations of the variance variables share. The longer an eigenvector is, the more variance it explains, the more important it is (Field, 2009). We can calculate an eigenvector's value by counting up the loadings of each variable on the eigenvector. As demonstrated in Figure 3, just a small proportion of the 20 eigenvectors of the correlation matrix in this study has a considerable eigenvalue: many reach 0 or have even lower values. We just want to retain the eigenvectors - or factors - that explain a considerable amount of the variance in the dataset, by which value do we draw the line?

Statistical packages generally retain factors with eigenvalues greater than 1.0 (Costello & Osborne, 2005). Yet, then there is a considerable change that too many factors are retained: in 36% of the samples Costello and Osborne studied (2005), too many factors were retained.

A more reliable and rather easy method is to look at the scree plot, as the graph in Figure 3 is called (Costello & Osborne, 2005). The factors with values above the point at which the curve flattens out should be retained. The factors with values at the break point or below should be eliminated. The statistical package R helps to determine where the break point is by drawing a straight line at that point. Thus, looking at Figure 3, two factors should be retained.

However, the best method to determine how many factors to retain is a maximum likelihood factor analysis, since that measure tests how well a model of a particular amount of factors accounts for the variance within a dataset (Costello & Osborne, 2005). A high eigenvalue does not necessarily mean that the factor explains a hugh amount of the variance in a dataset. It could explain the variance in one cluster of variables, but not in another one. That cluster probably measures another underlying factor
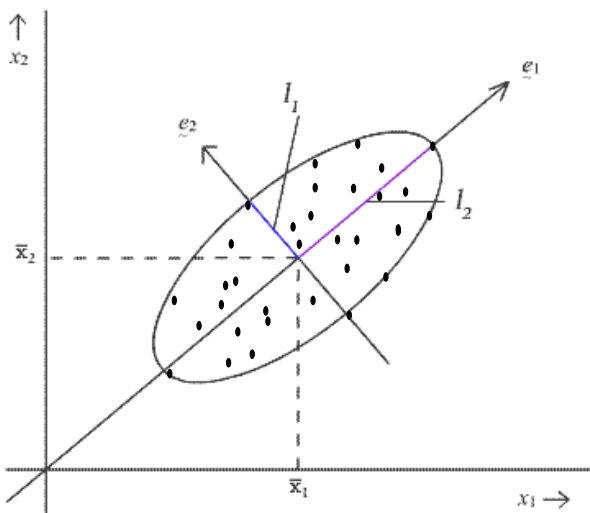

Figure 2. Scatterplot of two variables (x_1 and x_2). The lines e_1 and e_2 represent the eigenvectors of the correlation matrix of variables x_1 and x_2. The eigenvalue of an eigenvector is the length of an eigenvector measured from one end of the oval to the other end.
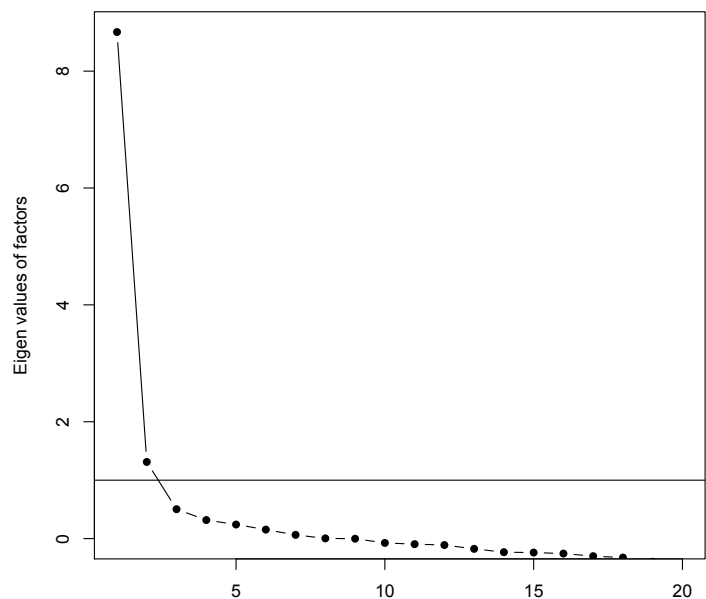

Figure 3. Screeplot of factors underlying the dataset. Every point represents one factor.

```
Call:
factanal(x = na.omit(passion), factors = 1)

Loadings:
     Factor1
p01  0.767                          Factor1
p02  0.710       SS loadings        8.635
p03  0.786       Proportion Var     0.432
p04 -0.637
p05  0.557
p06 -0.542       Test of the hypothesis that 1 factor is sufficient.
p07  0.520       The chi square statistic is 336.46 on 170 degrees of freedom.
p08 -0.572       The p-value is 4.89e-13
p09  0.782
p10 -0.529
p11 -0.545
p12  0.599
p13  0.605
p14  0.719
p15  0.714
p16 -0.558
p17  0.830
p18 -0.559
p19  0.674
p20  0.783

Call:
factanal(x = na.omit(passion), factors = 2, rotation = "oblimin")

Loadings:
     Factor1 Factor2
p01  0.547  -0.289                        Factor1 Factor2
p02  0.747            SS loadings         6.141   3.534
p03  0.727            Proportion Var      0.307   0.177
p04          0.802    Cumulative Var      0.307   0.484
p05  0.625
p06          0.641
p07  0.463            Factor Correlations:
p08 -0.133   0.558            Factor1 Factor2
p09  0.702  -0.115    Factor1   1.000  -0.642
p10          0.597    Factor2  -0.642   1.000
p11          0.680
p12  0.412  -0.243
p13  0.719   0.122    Test of the hypothesis that 2 factors are sufficient.
p14  0.739            The chi square statistic is 197.76 on 151 degrees of freedom.
p15  0.758            The p-value is 0.00636
p16          0.771
p17  0.917
p18          0.739
p19  0.623
p20  0.771

Call:
factanal(x = na.omit(passion), factors = 3, rotation = "oblimin")

Loadings:
     Factor1 Factor2 Factor3
p01  0.565  -0.263   0.293                   Factor1 Factor2 Factor3
p02  0.767           0.279    SS loadings      5.963   3.539   0.585
p03  0.720  -0.106           Proportion Var    0.298   0.177   0.029
p04          0.781  -0.121    Cumulative Var    0.298   0.475   0.504
p05  0.600          -0.134
p06          0.672   0.166
p07  0.438  -0.153  -0.336
p08 -0.130   0.575
```

Table 3. Output of a factor analysis in R with 1, 2 or 3 extracted factors

```
p09  0.695  -0.118   0.125
p10          0.605              Factor Correlations:
p11          0.656  -0.195           Factor1 Factor2 Factor3
p12  0.411  -0.235   0.146     Factor1  1.0000 -0.6148 -0.0385
p13  0.700   0.104             Factor2 -0.6148  1.0000  0.0596
p14  0.725                     Factor3 -0.0385  0.0596  1.0000
p15  0.738          -0.275
p16          0.794   0.136
p17  0.902          -0.169     Test of the hypothesis that 3 factors are sufficient.
p18          0.718             The chi square statistic is 153.68 on 133 degrees of freedom.
p19  0.611                     The p-value is 0.106
p20  0.750
```

Table 3. Output of a factor analysis in R with 1, 2 or 3 extracted factors

which should not be ignored. The null hypothesis in a maximum likelihood factor analysis is that the number of factors fits well the dataset, when the null hypothesis is rejected a model with a larger amount of factors should be considered. As can be seen in Table 3, a model with one factor is rejected at an α-level of 0.01, a model with two factors is rejected at an α-level of 0.05 and a model with three factors at none of the usual α-levels (p=.106). If we would set our α-level at 0.05 as common in social-scientific research (Field, 2009), a model with three factors would be the best choice. However, the third factor seems rather unimportant, it just explains 2.9% of the variance in the dataset and is based on just one variable p07. Variables with loadings lower than 0.3 are considered to have a non-significant impact on a factor, and need therefore to be ignored (Field, 2009). It seems more appropriate to set our α-level at 0.01 and assume that two factors should be retained. The second factor accounts for a considerable amount of the variance in the dataset: 17,7%. All variables load highly on that factor, except for the ones that load higher on the other factor and therefore seem to make up that one. Finally, the scree test rendered the same result.

3.2.2. Factor rotation

When several factors are extracted, the interpretation of what they represent should be based on the items that load on them (Field, 2009). If several variables load on several factors, it becomes rather difficult to determine the construct they represent. Therefore, in factor analysis, the factors are rotated towards some variables and away from some other. That process is illustrated in Figure 4. In that
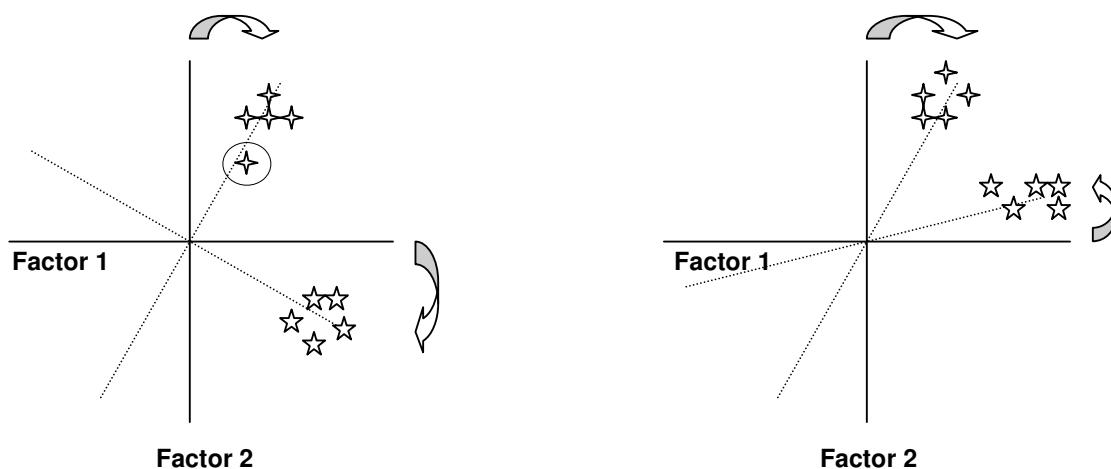

Figure 4. Graphical presentation of factor rotation: orthogonal rotation (left) and oblique rotation (right). The axes represent the extracted factors, the stars the original variables (graph source: Field, 2000, p. 439)

analysis, two factors were retained. One is represented as the x-axis, the other one as the y-axis. The variables (the stars) get their positions in the graph based on their correlation coefficients with both factors. It is rather ambiguous to which the circled variable belongs (left graph). It loads just a bit more on factor 1. However, by rotating both factors, the ambiguity gets solved: the variable loads highly on factor 1 and lowly on factor 2.

As can be seen in Figure 4, there are two kinds of rotation. The first kind of rotation 'orthogonal rotation' is used, when the factors are assumed to be independent (Field, 2009; Tabachnik & Fidell, 2001; Rietveld & Van Hout, 1993). The second kind of rotation 'oblique rotation' is used, when the factors are assumed to correlate. Since it was assumed that all 20 items in this questionnaire measured the same construct, we may expect that an oblique rotation is appropriate. This can be checked after having conducted the factor analysis, since statistical packages always give a correlation matrix of the factors when you opt an oblique rotation method (oblimin or promax). Therefore, it is highly recommended to always do a factor analysis with oblique rotation first, even if you are quite sure that the factors are independent (Costello & Osborne, 2005). The factors in this study certainly correlate with each other, although negatively: r=-0.64.

3.3.3. Factor interpretation

Ignoring the variables that load lower than 0.3 on a factor (see §3.3.1), we can conclude based on the output of the factor analysis with two extracted factors (see Table 3) that the positively formulated items in this questionnaire make up the first factor and the negatively formulated items the second factor. It is a rather common pattern that reverse-phrased items load on a different factor (Schmitt & Stults, 1985), since people do not express the same opinion when they have to evaluate a negatively phrased item instead of a positively phrased one (Kamoen, Holleman & Van den Bergh, 2007). People tend to express their opinions more positively when a questionnaire item is phrased negatively (Kamoen, Holleman & Van den Bergh, 2007). However, it can be expected that the two factors measure the same underlying construct, since they correlate considerably in a negative direction. It is after all expected that seventh graders who score highly on the negatively phrased items (hence, dislike writing), do not score highly on the positively phrased ones.

4. Cronbach's alpha

When the questionnaire at issue is reliable, people completely identical - at least with regard to their pleasure in writing - should get the same score, and people completely different a completely different score (Field, 2009). Yet, it is rather hard and time-consuming to find two people who are fully equal or unequal. In statistics, therefore, it is assumed that a questionnaire is reliable when an individual item or a set of some items renders the same result as the entire questionnaire.

The simplest method to test the internal consistency of a questionnaire is dividing the scores a participant received on a questionnaire in two sets with an equal amount of scores and calculating the correlation between these two sets (Field, 2009). A high correlation signals a high internal consistency. Unfortunately, since the correlation coefficient can differ depending on the place at which you split the dataset, you need to split the dataset as often as the number of variables in your dataset, calculate a correlation coefficient for all the different combinations of sets and determine the questionnaire's reliability based on the average of all these coefficients. Cronbach came up with a faster and comparable method to calculate a questionnaire's reliability:

$\alpha = (N^2 M(Cov))/(\sum s^2 + \sum Cov)$

Assumption behind this equation is that the unique variance within variables (s^2) should be rather small in comparison with the covariance between scale items (Cov) in order to have an internal consistent measure (Cortina, 1993).

```
Reliability analysis
Call: alpha(x = passion)

    alpha average_r mean   sd
    0.93    0.42    3.3   0.36

Reliability if an item is dropped:              Item statistics
        alpha      average_r                        n    r    r.cor mean sd
p01     0.93       0.41                    p01     114 0.77  0.76 3.5 1.19
p02     0.93       0.42                    p02     114 0.68  0.67 3.9 1.07
p03     0.93       0.41                    p03     114 0.77  0.76 3.6 1.06
p04-    0.93       0.42                    p04-    114 0.70  0.69 3.3 1.29
p05     0.93       0.43                    p05     113 0.57  0.55 3.8 0.90
p06-    0.93       0.42                    p06-    114 0.61  0.59 3.0 1.19
p07     0.93       0.43                    p07     114 0.55  0.52 3.4 0.82
p08-    0.93       0.42                    p08-    114 0.63  0.60 2.6 1.18
p09     0.93       0.41                    p09     114 0.78  0.77 3.6 0.98
p10-    0.93       0.43                    p10-    114 0.59  0.57 2.5 1.23
p11-    0.93       0.42                    p11-    114 0.60  0.58 3.1 1.26
p12     0.93       0.42                    p12     114 0.62  0.59 3.0 1.00
p13     0.93       0.42                    p13     114 0.60  0.57 3.4 1.08
p14     0.93       0.42                    p14     114 0.71  0.70 3.7 1.01
p15     0.93       0.42                    p15     114 0.71  0.70 3.8 0.95
p16-    0.93       0.42                    p16-    113 0.63  0.62 3.4 1.12
p17     0.93       0.41                    p17     113 0.80  0.80 3.8 0.85
p18-    0.93       0.42                    p18-    114 0.62  0.61 2.6 1.17
p19     0.93       0.42                    p19     114 0.67  0.66 3.2 1.06
p20     0.93       0.41                    p20     114 0.77  0.76 3.8 0.94
```

Table 4. Output of a Cronbach's alpha analysis in R

4.1. Prerequisite

Before the Cronbach's alpha of a questionnaire can be determined, the scoring of reverse-phrased items of a questionnaire needs to be reversed (Field, 2009). Hence, a score of 5 on a negatively formulated item in this questionnaire should be rescored in 1, a score of 4 in 2, et cetera. Assuming that a seventh grader who really likes writing strongly agrees with the statement 'I like writing' and strongly disagrees with the statement 'I hate writing', item scores can differ substantially between students as long as the scores of the reverse-phrased items are not reversed. Given that covariances between such scores are negative, the use of reverse-phrased items will finally lead to a lower and consequently incorrect Cronbach's alpha, since the top half of the Cronbach's alpha equation incorporates the average of all covariances between items. Fortunately, R reverses scores automatically. Since reverse-phrased items load negatively on an underlying factor, R detects them easily by determining that underlying factor.

4.2. Cronbach's alpha analysis

Generally, a questionnaire with an α of 0.8 is considered reliable (Field, 2009). Hence, this questionnaire certainly is reliable, since the α is 0.93 (see Table 4). The resulted α should yet be interpreted with caution. Since the amount of items in a questionnaire is taken into account in the equation, a hugh amount of variables can upgrade the α (Cortina, 1993; Field, 2009). For example, if we do a reliability analysis for just the items making up the first factor in our research, we get the same α, but the average correlation is 0.49 instead of 0.43. How hugh the alpha should be for a dataset with a particular amount of items is still a point of discussion (Cortina, 1993). Cortina (1993) recommends to determine the adequacy of a measure on the level of precision needed. If you want to make a fine distinction in the level of pleasure in writing someone has, a more reliable measure is needed than if you want to make a rough distinction. However, since the α of this questionnaire is far higher than 0.8, we can assume that it is reliable (Field, 2009).

Besides, a hugh Cronbach's alpha should not be interpreted as a signal of unidimensionality (Cortina, 1993; Field, 2009). Since α is a measure of the strength of a factor when there is just one factor underlying the dataset, many researchers assume that a dataset is unidimensional when the α is rather high. Yet, the α of this dataset is rather high as well, although the factor analysis revealed that the dataset is not unidimensional. Thus, if you want to measure with Cronbach's alpha the strength of a factor or factors underlying a dataset, Cronbach's alpha should be applied to all the factors extracted during a previous factor analysis (Field, 2009). Our factors turn out to be quite strong: the α of the factor made up of the positively phrased items is 0.93, the α of the factor made up of the negatively phrased items is 0.86.

The final step in the interpretation of the output of a Cronbach's alpha analysis is determining how each item individually contributes to the reliability of the questionnaire (Field, 2009). As can be seen in Table 4, R also renders the values of the α, when one of the items is deleted. If the α increases a lot when a particular item is deleted, one should consider deletion. The same counts for items which decrease the average correlation coefficient a lot, or correlate lower than 0.3 with the total score of the questionnaire (see the values under r and r.cor in Table 4; r.cor is the item total correlation corrected for item overlap and scale correlation). In this questionnaire, all items positively contribute to the reliability of the questionnaire. The α remains the same when an item is deleted, the average r almost the same, and the correlations between the total score and the item score are moderate to high.

5. Conclusion

The evaluated questionnaire seems reliable and construct valid. The items measure the same underlying construct. The extraction of two factors in the factor analysis just seems to be a consequence of the wording of the questionnaire items. After all, the two factors correlate highly with each other. The result of the reliability measure was high: α=0.93. All items contribute to the reliability and construct validity of the questionnaire: the items correlate more than 0.4 with the factors that underlie them, the Cronbach's alpha does not increase when one of the questionnaire items is deleted, and the average correlation coefficient sometimes just a bit.

Although a questionnaire is generally accepted as reliable when the Cronbach's alpha is higher than 0.8, we cannot claim that the questionnaire is valid based on the factor analysis alone (Bornstedt, 1977; Ratray & Jones, 2007). We just know that the items measure the same underlying construct. It can be expected based on the questionnaire statements that that is one's pleasure in writing. However, to prove that it measures one's pleasure in writing, the results of other measures of one's pleasure in writing should be compared (Bornstedt, 1977; Ratray & Jones, 2007). Unfortunately, it is not easy to invent such measures (O'Keefe, 2002).

Bibliography

Bornstedt, G.W. (1977). Reliability and Validity in Attitude Measurement. In: G.F. Summers (Ed.), *Attitude Measurement* (pp. 80-99). Kershaw Publishing Company: London.

Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98-104.

Costello, A.B., & Osborne, J.W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research and Evaluation, 10*, 1-9.

Field, A. (2000). *Discovering Statistics using SPSS for Windows*. Sage: London.

Field, A. (2009). *Discovering Statistics using SPSS*. Sage: London.

Kamoen, N., Holleman, B.C., & Van den Bergh, H.H. (2007). Hoe makkelijk is een niet moeilijke tekst? Een meta-analyse naar het effect van vraagformulering in tekstevaluatieonderzoek [How

easy is a text that is not difficult? A meta-analysis about the effect of question wording in text evaluation research]. *Tijdschrift voor Taalbeheersing, 29*, 314–332.

O'Keefe, D.J. (2002). *Persuasion. Theory & Research*. Sage: Thousand Oaks.

Rattray, J.C., & Jones, M.C. (2007), Essential elements of questionnaire design and development. *Journal of Clinical Nursing, 16*, 234-243.

Rietveld, T. & Van Hout, R. (1993). *Statistical Techniques for the Study of Language and Language Behaviour*. Berlin - New York: Mouton de Gruyter.

Schmitt, N. & Stults, D.M. (1985). Factors Defined by Negatively Keyed Items: The Result of Careless Respondents? *Applied Psychological Measurement, 9*, 367-373.

Tabachnik, B.G., & Fidell, L.S. (2001) *Using Multivariate Statistics* (4th ed.). Pearson: Needham Heights, MA.

Thurstone, L.T. (1977). Attitudes can be measured. In: G.F. Summers (Ed.), *Attitude Measurement* (pp. 127-141). Kershaw Publishing Company: London.