PL KNOWLEDGE scaling

Szymon Talaga

28.04.2015

Introduction

The scaling procedure that is presented here uses the Mokken Scale technique, which is an non-parametric IRT technique that allows prodcution of high-quality ordinal scales. Mokken scales are useful because they posses two very important properties: monotonous homogeneity and double monotonicity. Monotonous homogeneity means that in a proper Mokken scale it is assured that an intenisty of a respondent's trait (in this case it would be economic/financial knowledge) is positively correlated with a probability of he or she giving the correct anwer to a question. It ensures correct ordinal measurement of a trait in which the sum of the correct answers is a sufficient estimator of the trait.

Double monotonicity means that for every person the item ordering by difficulty is invariant. In other words it means that the ordering of the items from the easiest to the most difficult stays the same regardless of a respondent in question. As a consequence this property means that the fraction of incorrect answers to an item is a sufficient estimator of its difficulty level.

The main aim of this script is to test whether this two assumption hold in the case of the PL sample and what subset(s) yield the best scale(s). All the tests conducted here will be based on the inspection of four statistics:

- #ac: theoretical maximal number of the possible violations of the model
- #vi: number of the actual violations in the data
- #zsig : number of violations that are considered statistically significant
- #crit: special summary statistic that combines the former three; values above 40 implies moderate violations of the model assumptions; values above 80 implies strong violation

The aim of the first part of the analysis is to detect and remove poorly scaling items. Then it will be possible to use inductive techniques to find items that form the best scale or scales.

```
## Loading required package: scatterplot3d
## Loading required package: MASS
```

Scaling using the Mokken Scale approach

First we use inductive technique based on a genetic algorithm to find subset of items that form the best scales.

```
##     Scale
## k1      3
## k2      0
## k3      0
## k4      0
## k5      0
## k6      0
## k7      0
```

```
## k8      1
## k10     3
## k11     0
## k12     0
## k13     0
## k14     2
## k15     0
## k16     0
## k17     0
## k18     0
## k19     0
## k20     0
## k21     1
## k22     1
## k23     0
## k24     0
## k25     2
## k26     0
## k27     0
## k28     0
## k29     0
## k30     1
```

It is clear that the data does not allow creation of long and good Mokken Scales. The only possibility is to make a 4-items Mokken Scale out of items: k8, k21, k22 and k30. Therefore now we will asses the quality of this scale.

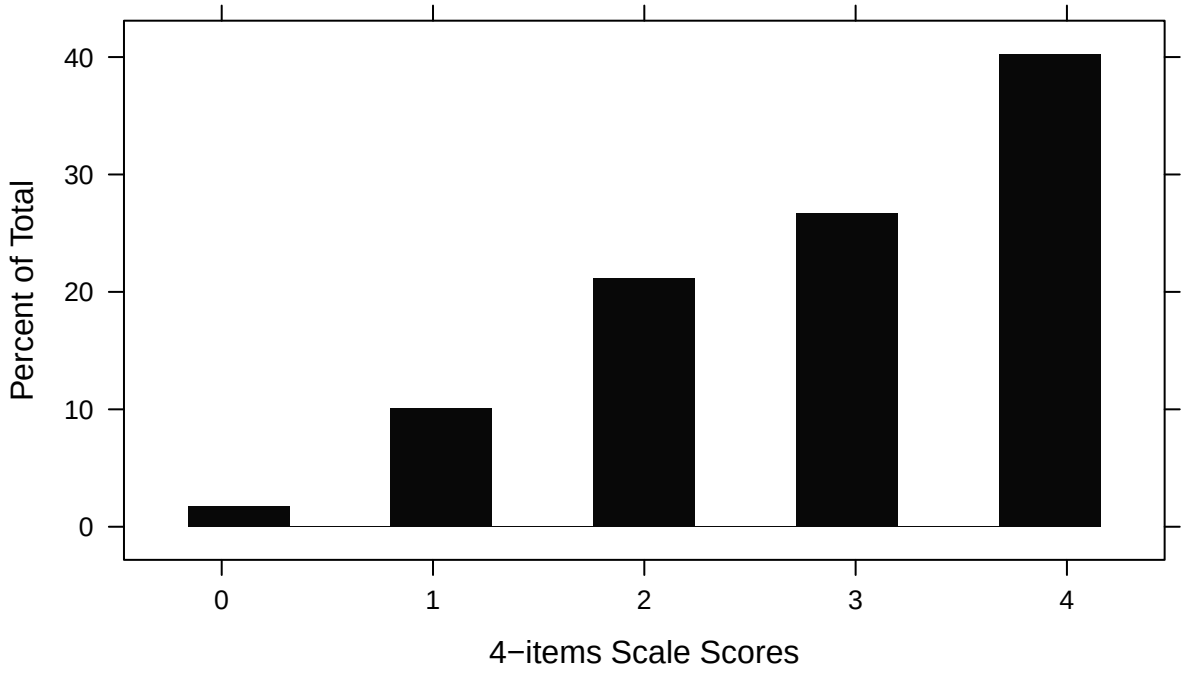
```
##     ItemH #ac #vi #vi/#ac maxvi sum sum/#ac zmax #zsig crit
## k8   0.47   1   0       0     0   0       0    0     0    0
## k21  0.34   3   0       0     0   0       0    0     0    0
## k22  0.32   3   0       0     0   0       0    0     0    0
## k30  0.31   3   0       0     0   0       0    0     0    0
```

We see that the scale conforms well to the monotonous homogeneity assumption. Now we will check the double monotonicity assumption.

```
##     ItemH #ac #vi #vi/#ac maxvi sum sum/#ac zmax #zsig crit
## k8   0.47   3   0       0     0   0       0    0     0    0
## k21  0.34   3   0       0     0   0       0    0     0    0
## k22  0.32   3   0       0     0   0       0    0     0    0
## k30  0.31   3   0       0     0   0       0    0     0    0
```

It appears it also conforms well to the double monotonicity assumption. Therefore it means that the scale provides ordinal measurement that allows treating raw sums of correct answers as sufficient estimators of the true level of the trait (economic/financial knowledge). Moreover items ordering by difficulty levels is ivariant. Summing up, it implies that although the scale is of rather low quality it still is a formal Mokken Scale and may be used to asses the level of economic/financial knowledge.

However one thing is still to be done. The discrimiatory power of the scale has to be assesed. This can be done (at basic level) by inspecting the distribution of the scale scores (see the plot below).



The distribution is strongly left-skewed what implies that the scale does not discriminate between respondents with high economic/financial knowledge, but do discriminate between those of the poor knowledge.

The analysis proves that the data does not conform to the Mokken model in general. Only the 4-items scale has barely satisfactory global scaling properties and allows some basic discrimination between respondents in regard to the economic/financial knowledge. Therefore we will also try to scale the KNOWLEDGE items set using the Classical Test Theory approach.

Scaling using the Classical Test Theory Approach

We check the realibility of the entire set of items and we determine which items should be deleted to obtain a better scale.

```
## Warning in alpha(data): Some items were negatively correlated with total
## scale and were automatically reversed.

##  raw_alpha std.alpha
##       0.66      0.66

##      raw_alpha std.alpha
## k1        0.66      0.67
## k2        0.65      0.65
## k3        0.65      0.65
## k4        0.64      0.65
## k5-       0.67      0.67
## k6        0.65      0.66
## k7        0.65      0.65
## k8        0.63      0.64
## k10       0.66      0.66
## k11-      0.67      0.67
## k12       0.65      0.66
```

```
## k13       0.65      0.66
## k14       0.67      0.67
## k15       0.64      0.65
## k16       0.66      0.66
## k17       0.66      0.66
## k18       0.65      0.65
## k19       0.66      0.67
## k20-      0.65      0.66
## k21       0.65      0.65
## k22       0.64      0.64
## k23       0.65      0.65
## k24-      0.66      0.67
## k25       0.65      0.65
## k26       0.64      0.65
## k27       0.66      0.66
## k28       0.64      0.64
## k29       0.64      0.65
## k30       0.64      0.64
```

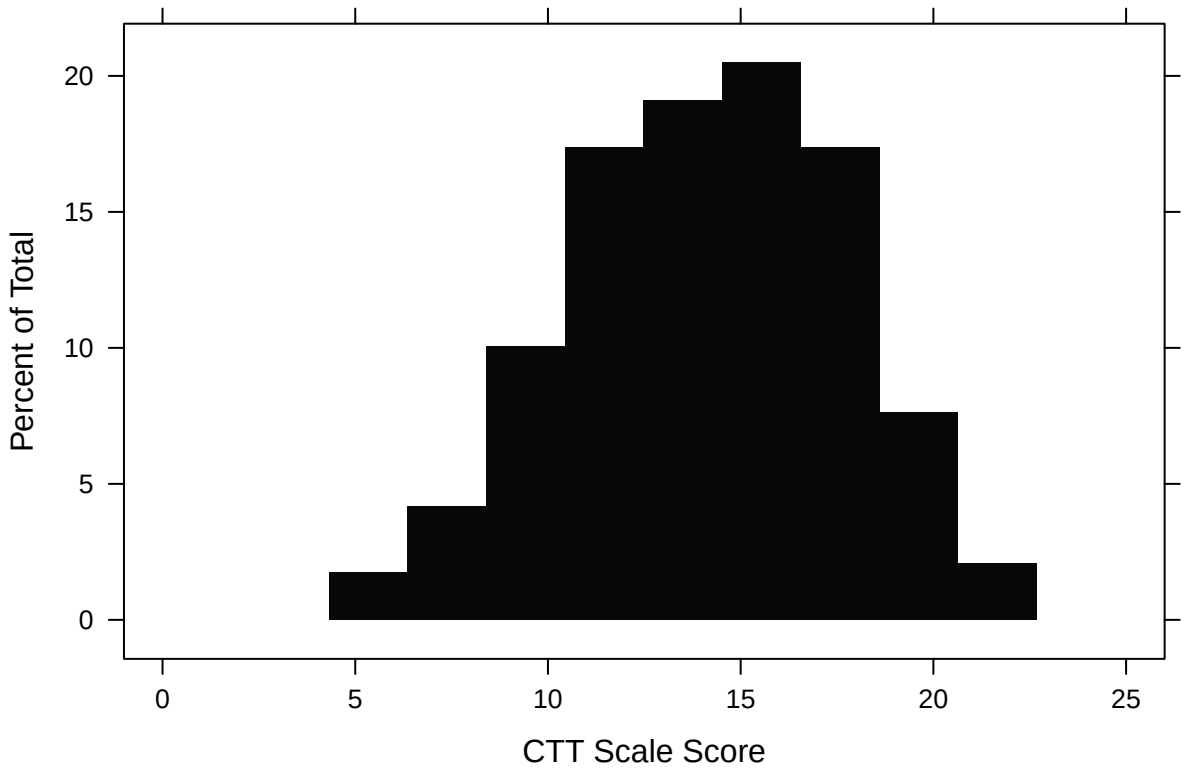
The scale has at most satisfactory reliability. It is also clear that several items need to be removed. These are items k5, k11, k20 and k24 due to the negative correlation with the scale and item k14 becuase it lowers the reliability of the scale.

```
##  raw_alpha std.alpha
##       0.67      0.66
```

```
##     raw_alpha std.alpha
## k1       0.67      0.67
## k2       0.65      0.65
## k3       0.65      0.65
## k4       0.65      0.64
## k6       0.65      0.65
## k7       0.66      0.65
## k8       0.63      0.63
## k10      0.67      0.67
## k12      0.66      0.66
## k13      0.66      0.66
## k16      0.66      0.66
## k17      0.66      0.66
## k18      0.65      0.65
## k19      0.67      0.67
## k21      0.65      0.65
## k22      0.64      0.64
## k23      0.65      0.65
## k26      0.65      0.65
## k27      0.66      0.66
## k28      0.65      0.64
## k29      0.65      0.65
## k30      0.64      0.63
```

Now no further items can be removed to better the scale, thus we arrived at the final version of it.

The distribution of the scale scores is quite symmetric what implies it may discriminate better than the 4-items Mokken Scale (see the plot below). On the other hand it does not posses the specific properties (monotonous homogeneity and double monotonicity) of a Mokken Scale.




```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    5.00   12.00   14.00   14.07   17.00   22.00
```

```
##
##  Shapiro-Wilk normality test
##
## data:  CTTscale
## W = 0.98399, p-value = 0.002641
```

Therefore this scale may be considered satisfactory, but again it is rather of only low quality. It allows better discrimination amongst respondents with higher economic/financial knowledge, but it does not formally ensure that the scale scores fully and properly reflect economic/financial knowledge of respondents.

Assesment of theoretical validity of the scales

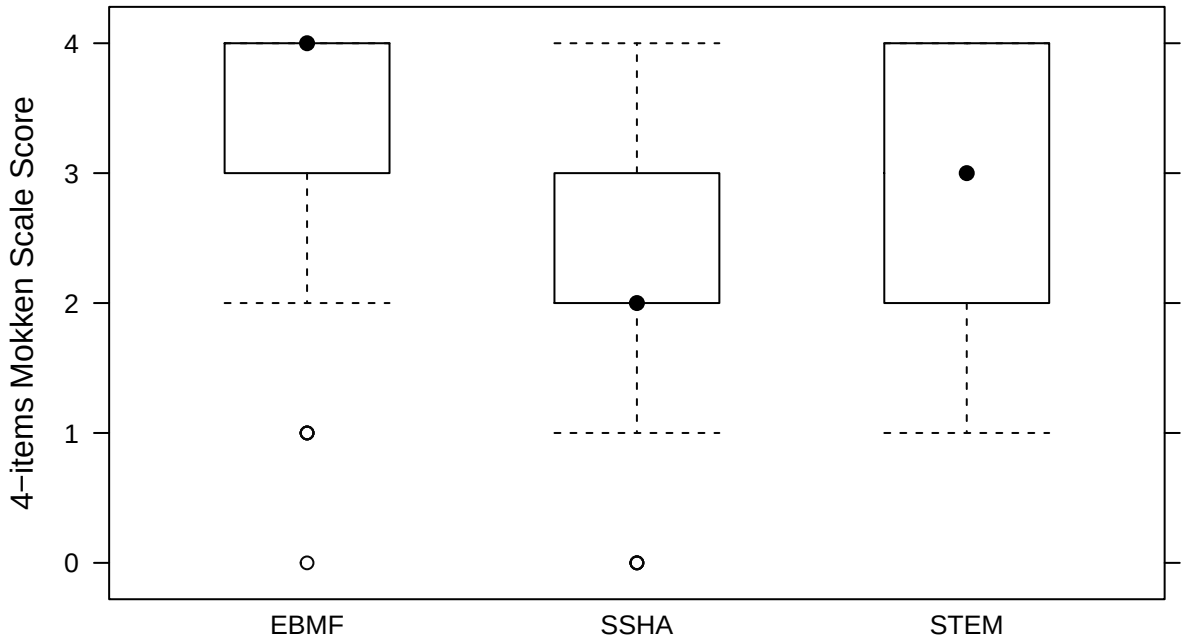
Now we move to the last stage of this analysis where theoretical validity of the scales will be assesed. The main part of this analysis is aimed at examining mean differences respondents enrolled in different types of university programmes. We compare those studying social sciences, humanities and art (SSHA), economic, bussiness, management and financial studies (EBMF) and those enrolled in STEM programmes (STEM. It is expected that EBMF will have higher mean score that the other groups.

Additionaly we will check wheter the economic/financial knowledge grow with the another years of education. We hypothesize that there should be positive (linear) trend in the EBMF group and no such trend in the others, since this kind of knowledge is taught mainly in the EBMF types of programmes.

Theoretical validity of the 4-items Mokken Scale

Numerical values of the group means suggest that indeed the EMBF group scores higher than the rest (see the plot below as well).

```
## $EBMF
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   3.000   4.000   3.434   4.000   4.000
##
## $SSHA
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   2.000   2.000   2.407   3.000   4.000
##
## $STEM
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   3.000   2.902   4.000   4.000
```

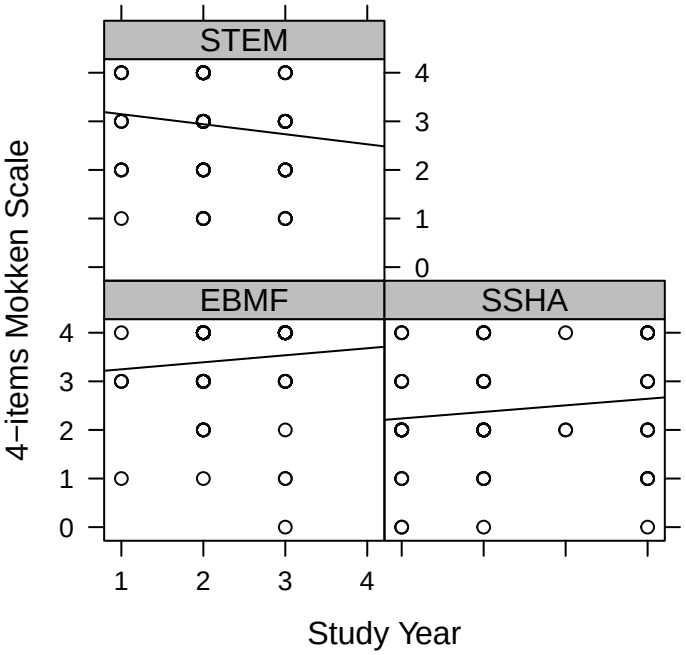


The formal assesment (using a linear model with dummy coding that compares SSHA and STEM groups against the EMBF group) proved that observed differences are statisticakky significant and 'explain' about 14.5% of the variance of the 4-items Mokken Scale.

```
##
## Call:
## lm(formula = Mokken4 ~ eduprog3, data = data4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4343 -0.4343  0.0980  0.5657  1.5926
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.4343     0.1005  34.178  < 2e-16 ***
## eduprog3SSHA  -1.0269     0.1498  -6.856 4.56e-11 ***
```

```
## eduprog3STEM  -0.5324     0.1411  -3.774 0.000196 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9998 on 279 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.145,  Adjusted R-squared:  0.1389
## F-statistic: 23.66 on 2 and 279 DF,  p-value: 3.223e-10
```

Now we turn to the change of the level of the knowledge over the years of education. Visual inspection shows that there are some trends in the three groups (in the STEM group the trend even appears to be negative).



However this observation must be confronted with formal tests. Again we use linear models in the groups in which the Mokken Scale score is the dependent variable and the year at the university the independent one (we collapsed those at their second year of MA with those at the first due to the very few respondents in the former).

```
##
## Call:
## lm(formula = Mokken4 ~ uniyear, data = data4, subset = data4$eduprog3 ==
##     "EBMF")
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5364 -0.3921  0.4636  0.6079  0.7523
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.1033     0.3723   8.336 5.17e-13 ***
## uniyear       0.1444     0.1578   0.915    0.363
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.8716 on 97 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.008555,   Adjusted R-squared:  -0.001666
## F-statistic: 0.837 on 1 and 97 DF,  p-value: 0.3625
```

```
##
## Call:
## lm(formula = Mokken4 ~ uniyear, data = data4, subset = data4$eduprog3 ==
##     "SSHA")
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6406 -0.5057 -0.2358  0.7642  1.7642
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.1009     0.2851   7.369 1.46e-10 ***
## uniyear       0.1349     0.1123   1.201    0.233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.145 on 79 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.01794,    Adjusted R-squared:  0.005514
## F-statistic: 1.444 on 1 and 79 DF,  p-value: 0.2332
```

```
##
## Call:
## lm(formula = Mokken4 ~ uniyear, data = data4, subset = data4$eduprog3 ==
##     "STEM")
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.14658 -0.73416  0.05963  1.05963  1.26584
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.3528     0.3570   9.392 2.14e-15 ***
## uniyear      -0.2062     0.1571  -1.313    0.192
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9866 on 100 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.01695,    Adjusted R-squared:  0.007116
## F-statistic: 1.724 on 1 and 100 DF,  p-value: 0.1922
```

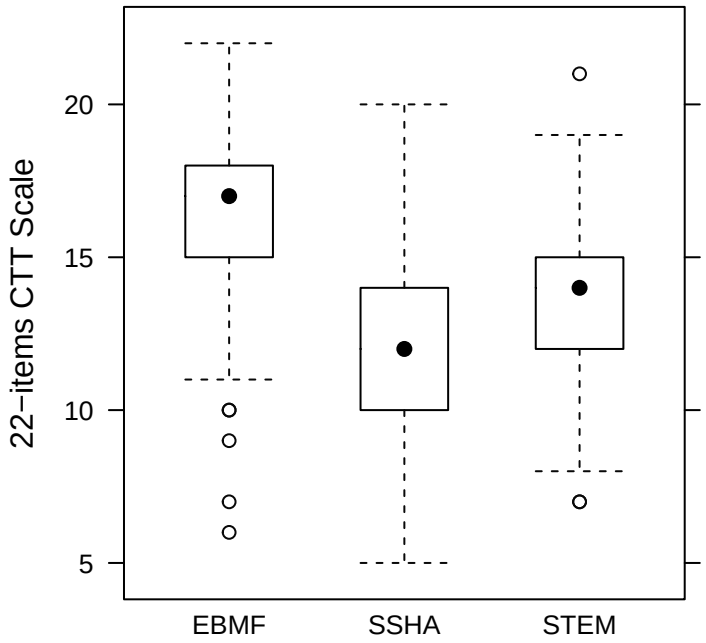
Clearly none of the trends is significant.

Summing up the theoretical validity of the scale seems to be proved to some extent, althought lack of the trend in the EBMF is somehow surprising.

Theoretical validity of the 22-items Classical Test Theory Scale

Numerical values of the group means suggest again that indeed the EMBF group scores higher than the rest (see the plot below as well).

```
## $EBMF
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6.00   15.00   17.00   16.33   18.00   22.00
##
## $SSHA
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    5.00   10.00   12.00   12.23   14.00   20.00
##
## $STEM
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    7.00   12.00   14.00   13.44   15.00   21.00
```

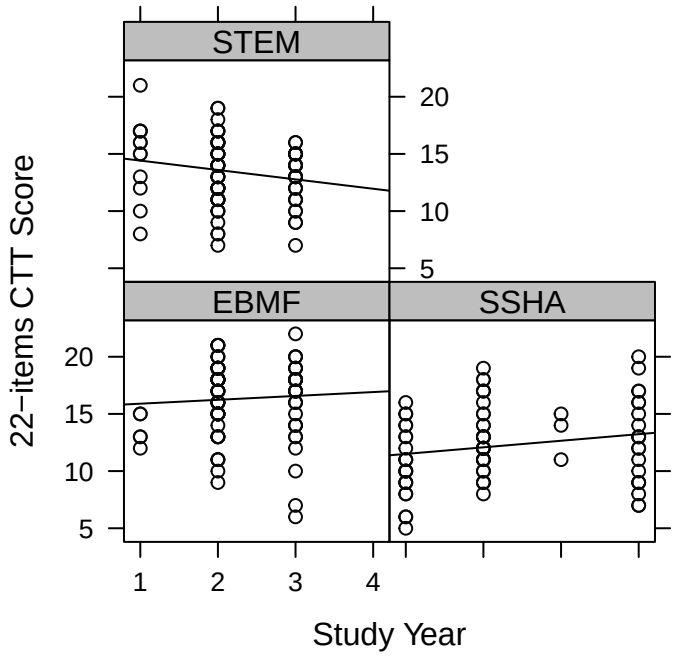


The formal test (again a linear model with dummy coding of SSHA and STEM against EBMF) proved that the observed difference is statistically significant.

```
##
## Call:
## lm(formula = CTT ~ eduprog3, data = dataCTT)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.3333  -2.2346   0.5588   1.7654   7.7654
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   16.3333     0.3041  53.717  < 2e-16 ***
## eduprog3SSHA  -4.0988     0.4533  -9.043  < 2e-16 ***
## eduprog3STEM  -2.8922     0.4268  -6.776 7.33e-11 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.025 on 279 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.243,  Adjusted R-squared:  0.2376
## F-statistic: 44.79 on 2 and 279 DF,  p-value: < 2.2e-16
```

Now we turn to the change of the level of the knowledge over the years of education. Visual inspection shows that there are some trends in the three groups (in the STEM group the trend even appears to be negative).



However this observation must be confronted with formal tests. Again we use linear models in the groups in which the Mokken Scale score is the dependent variable and the year at the university the independent one (we collapsed those at their second year of MA with those at the first due to the very few respondents in the former).

```
##
## Call:
## lm(formula = CTT ~ uniyear, data = dataCTT, subset = data4$eduprog3 ==
##     "EBMF")
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5728  -1.2341   0.7659   1.7659   5.4272
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.5566     1.3272  11.721   <2e-16 ***
## uniyear       0.3387     0.5626   0.602    0.548
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.107 on 97 degrees of freedom
```

```
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.003724,   Adjusted R-squared:  -0.006547
## F-statistic: 0.3626 on 1 and 97 DF,  p-value: 0.5485
```

```
##
## Call:
## lm(formula = CTT ~ uniyear, data = dataCTT, subset = data4$eduprog3 ==
##     "SSHA")
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.5031 -2.0783 -0.0783  2.3464  6.9217
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.9278     0.8049  13.577   <2e-16 ***
## uniyear       0.5752     0.3171   1.814   0.0734 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.234 on 79 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:   0.04,  Adjusted R-squared:  0.02785
## F-statistic: 3.292 on 1 and 79 DF,  p-value: 0.07343
```

```
##
## Call:
## lm(formula = CTT ~ uniyear, data = dataCTT, subset = data4$eduprog3 ==
##     "STEM")
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.594 -1.729  0.406  2.227  6.585
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.2353     0.9754  15.620   <2e-16 ***
## uniyear      -0.8206     0.4291  -1.912   0.0587 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.696 on 100 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.03528,    Adjusted R-squared:  0.02564
## F-statistic: 3.657 on 1 and 100 DF,  p-value: 0.05868
```

Clearly none of the trends is significant.

Summing up the theoretical validity of the scale seems to be proved to some extent, althought lack of the trend in the EBMF is (again) somehow surprising.