How can statistical analysis prove the existence of an independent foreign language skill?

Ágnes Dévény[1]

The aim of my research was to prove that foreign language mediation is an independent language skill that can be measured by specific criteria and as part of a foreign language examination it contributes to a more complex assessment of the test-takers' language knowledge.

Besides other methods that are beyond the scope of this paper, the most important part of the research was the descriptive and mathematical statistical analysis of the results of the 18 examination periods covering more than 27000 test results from Budapest Business School Examination Centre. I wanted to present in this paper how I used statistical analyses to prove the existence of an independent foreign language skill.

My research verified my hypothesis that written mediation examination task measures a segment of language knowledge independently; skills (subskills) appearing in it are not present in other examination tasks. Foreign language mediation is an independent language skill; therefore bilingual language examinations can measure skills that are not measureable by monolingual examination systems.

Keywords: assessing foreign language competence, language testing, pedagogy

1. Introduction

Is foreign language mediation an independent language skill? Is it part of our conception of language proficiency and a segment of the language learners'/users' language competence? There have been heated debates in professional circles on the role and function of mediation and the debates have not been settled yet (Szabari 2001; Bárdos 1997, 2005; Heltai 2001). Experts interpret and explain the concept of mediation in many different ways and their opinions vary, reflecting the insecurity of specialists in the field of bilingualism/multilingualism. Debates become even more heated when experts have to decide whether a mediation task (e.g. written mediation test) can be considered as legitimate language examination subtest (task) (Klaudy 1984, 1986a, 1986b, 1990; Fekete 2001, 2002). The scientific investigation of bilingualism/multilingualism in education and testing, as well as the interpretation

[1] Ágnes Dévény, PhD, associate professor, Budapest Business School, Institute of Foreign Languages and Communication, Faculty of Commerce, Catering and Tourism, Hungary (Budapest)

of the notion of foreign language mediation, are especially delicate topics as they conceal the diversity and distinctness of opinions and theories of language skills and abilities. Opponents of bilingual examinations do not regard any type of mediation as an independent language skill and an element of language knowledge, and do not accept the concept that by testing and measuring this skill we can get a more complex idea of the candidate's foreign language proficiency. On the contrary, many of them fear of the negative washback effect of the mother tongue on foreign language acquisition (Nikolov et al. 1999). Contradictory views are reflected in the fact that whereas mediation task was completely left out from school-leaving (maturity) examination tasks, several bilingual and monolingual examination systems have been accredited in Hungary since the beginning of the accreditation procedures in 1999 (Einhorn 1998; Nikolov et al. 1999; Alderson 2001). The acceptance of bilingual examination systems was justified only by some needs analyses and Hungarian language examination traditions, but there was not any scientific research verifying that foreign language mediation is an independent language skill, and it is part of the candidate's language competence.

The principal argument of the opponents of the mediation task is its negative washback on teaching practice, and the difficulties of its evaluation arising from the complex nature of the skill, which definitely includes several subskills.

The negative impact cannot be considered as a generally acceptable counter-argument in all cases. According to Bachman & Palmer (2000) a language test can be useful and reasonable if it corresponds in demonstrable ways to language use in a specific target language use domain and is based on certain procedures of needs analysis.

A national survey (Teemant et al. 1993) and some surveys of smaller scope (Major 2000; F. Silye 2004; Dévény − Szőke 2007) were conducted mainly in the field of professional language usage during the last decade. The results of these surveys show that mediation is verified as a real-life domain of language usage, so it is justifiable as a test task, but they do not produce any evidence of acknowledging it as an independent language skill.

2. Posing (raising) the problem

The aim of my research was to prove that foreign language mediation is an independent language skill. It is part of the candidate's foreign language competence, and the written mediation task in the language examination system, examined in my survey, is suitable for evaluating this skill just like other subtests each using its measurement criteria respectively. This language skill and its measurement behaves fundamentally in the same way in the test battery as other subtests, and its application in the examination system contributes to obtaining a

more complete and more objective picture of the candidates' foreign language proficiency.

I strove to answer two sets of questions in my study: *1. from the point of view of the examination; 2. from the point of view of language teaching*. In this paper I will only deal with the first point of view and present my research questions concerning this point:

− What is the construct validity of the written mediation task in the language examination?

− Is mediation a reliable examination task in the language examination system?

− Are there any subskills in written mediation tasks that appear in other subtests as well?

− What does the inter-subtest correlation of written mediation task with other examination tasks show?

− Based on these research questions the following *hypothesis* was created:

− The special language skills that appear and can be measured in written mediation tasks are not present in other subtests, therefore bilingual language examinations can measure skills that are not measureable by monolingual examination systems.

In my research I used statistical analyses of the examination scores to examine the relationship between the results of the subtests. Although different types of statistical analyses e.g. confirmatory factor analysis (CFA) has been used in psychology and other social sciences to support measuring abilities, personality traits etc., the concept of using statistical analyses to prove the existence of an independent foreign language skill is not a usual approach in language pedagogy as until now in classical and modern test theories statistical analyses were used for solving different testing and test construction problems but did not serve as a method of proving the existence of a language skill.

3.   The background of the research

For the research I chose the examination system of Budapest Business School, a bilingual, criterion-referenced, LSP[2] examination system, where along with the traditional language skills (speaking, writing, reading comprehension, listening comprehension) mediation skill is also measured at different levels of the examination. Examination tasks of complex intermediate (B2) exam at the time of the research were the following:

---

[2] LSP = language for specific purposes

1. Written examination:
   - grammar test
   - reading comprehension test
   - writing test (generally business letter)
   - written mediation test from Hungarian to foreign language
2. Listening comprehension test
3. Oral examination:
   - introductory conversation
   - oral mediation task (mediation of a newspaper article into Hungarian)
   - speaking test (conversation on vocational topics
   - situation

The Written mediation test under research as an intermediate examination task meant the following: the examinee had to mediate in written form a Hungarian text of approximately 150 words, with an output of approximately 100 words. The requirement was not a word for word translation of the text but it involved a summarizing element. Candidates were not allowed to use a dictionary.

4. Methods of the research

4.1. *The characteristics of the sample*

For the statistical analysis of the examination results I used the *examination scores* of the candidates who took the intermediate LSP exam from 2000 to 2007 in BBS[3] Examination Centre. The members of the sample were 17-35 year-old men and women, typically college and university students (97-98%), 2-3% of them were employees from different areas of the economy. The size of the sample was slightly different in each examination period, but it ranged typically from 1000 to 2000, summing up to 27,832 candidates in 18 examination periods.

4.2. *Methods of data analysis*

As a method of data analysis I used descriptive and inferential statistics calculating by the help of SPSS (inter-subtest correlations, subtest − final test score correlations, reliability analysis (index) of the subtests, factor analysis, multiple regression analysis, cluster analysis).

---

[3] Budapest Business School

5. The results of the research

Based on the statistical analyses I wanted to support my hypothesis claiming that written mediation examination task can measure a skill or a complexity of subskills that cannot be measured with other examination tasks, thus besides with the traditional language skills an *independent foreign language mediation skill* exists. Statistical analyses in this paper will be illustrated on the result of one examination period (January 2005) and the results of all 18 exam periods will be summarized separately. Figure 1 shows the main characteristics of the chosen examination period.

Figure 1. Descriptive statistics of the examination test scores
(Sample period: January 2005) (n=1669)

	Grammar test	Reading Comp.	Writing test	Written mediation	Listening Comp.	Introct. conv.	Speaking	Situation	Oral mediation
Mean	10.78	12.12	12.31	5.81	9.52	15.53	13.21	14.72	7.26
SE of Mean	.088	.090	.093	.049	.096	.076	.101	.100	.050
Median	11	12	12	6	9	16	14	15	8
Mode	10	13	12	6	9	18	14	20	8
SD	3.610	3.680	3.814	1.994	3.917	3.093	4.116	4.073	2.038
Variance	13.030	13.544	14.549	3.975	15.341	9.564	16.943	16.588	4.153
Skewness	-.170	-.148	-.103	-.207	.385	-.643	-.455	-.757	-.585
SE of Skewness	.060	.060	.060	.060	.060	.060	.060	.060	.060
Kurtosis	-.561	-.415	-.414	-.280	-.412	,085	-.284	.268	-.110
SE of Kurtosis	.120	.120	.120	.120	.120	.120	.120	.120	.120
Range	19	19	20	10	19	16	20	20	10
Minimum	1	1	0	0	1	4	0	0	0
Maximum	20	20	20	10	20	20	20	20	10

Source: own calculations using SPSS

5.1. Internal correlations as a way of assessing the construct validity of tests

Construct validity is a form of test validation which essentially involves assessing to what extent the test is successfully based upon its underlying theory. (Alderson et al. 1995) In order to reveal the construct validity of the mediation task I calculated the inter-subtest and the subtest − final test scores correlation coefficients of the examination tasks.

5.1.1. Inter-subtest correlations

The idea behind inter-subtest correlation of test components is that if two test-components would correlate very highly with each other, we might assume that the two test components are *not* testing different traits or skills. It means that one of them might be superfluous.

At first I used Pearson Correlation to see the overlapping between the subtests. The optimal value in the case of inter-subtest correlation is between .3 – .5, that means a 9 to 25% overlap of variance between two subtests (Alderson et al. 1995). The results (Figure 2.) show that except for *Grammar test* all other subtests' inter-subtest indices are in the ideal interval. But even in the case of *Grammar test* the overlap is only about 30% of variance of the two subtests.

The problematic point of the measurement is that test results do not always show a normal distribution, as examinees who sit for a criterion-referenced proficiency exam are supposed to be well-prepared – so the distribution curve can be negatively skewed (see Figure 1). That is why the non-parametric correlation was counted as well. The correlation matrix showed nearly the same results as the parametric analysis.

Figure 2. Inter-subtest correlation matrix of written mediation (L1 $\rightarrow$ L2)[4] task. Comparison of the results of parametric and non-parametric analysis. (January 2005) (n=1669)

Subtests	M	SD	Parametric analysis				Non-parametric analysis			
			r Pearson Corr.	p	R^2	Overlap of variance %	r Spearman's rho	p	R^2	Overlap of variance %
Grammar test	10.78	3.610	**.561****	.000	.315	31	**.550****	.000	.302	30
Reading Comp.	12.12	3.680	**.424****	.000	.179	18	**.416****	.000	.173	17
Writing test	12.31	3.814	**.459****	.000	.210	21	**.445****	.000	.198	20
Listening Comp.	9.52	3.917	**.461****	.000	.212	21	**.450****	.000	.202	20
Introductory Conv.	15.53	3.093	**.367****	.000	.134	13	**.361****	.000	.130	13
Speaking test	13.21	4.116	**.341****	.000	.116	12	**.345****	.000	.119	12
Situation	14.72	4.073	**.386****	.000	.148	15	**.394****	.000	.155	16
Oral mediation	7.26	2.038	**.338****	.000	.114	11	**.340****	.000	.116	12

** Correlation is significant at the 0.01 level (2-tailed).
Source: own calculations using SPSS

The means of inter-subtest correlations of 18 examination periods show a desirable overlapping between the different subtests – an approximately 18 – 25% overlap of variance (Figure 3). The only exception is the *Grammar test* which shows an average of 36% overlap of variance. *Grammar test,* as it does not measure an independent language skill, behaves in the same way in all examination periods.

---

[4] L1=Language1 (generally mother tongue); L2=Language2 (foreign language)

Figure 3. Overlapping of skills measured by written mediation (L1 → L2) task with skills measured by other subtests
[Mean of inter-subtest correlations of 18 examination periods (%)]

Subtests	r	p	R^2	Overlap of variance (%)
Grammar test	.598	.000	.357	**36**
Reading Comp.	.493	.000	.243	**24**
Writing test	.507	.000	.257	**26**
Listening Comp.	.461	.000	.212	**21**
Introductory Conv.	.417	.000	.173	**17**
Speaking test	.426	.000	.181	**18**
Situation	.437	.000	.190	**19**
Oral mediation (L2-L1)	.420	.000	.176	**18**

Source: own calculations using SPSS

5.1.2. Subtest – total test scores correlation

According to classical test theory of language testing the correlations between each subtest and the whole test might be expected to be higher – possibly around +.7 or more – since the overall score is taken to be a more general measure of language ability then each individual component score (Alderson et al 1995). Subtest – total test scores correlation (Figure 4) is calculated because it shows the impact of the subtest on the total (final, overall) test scores of the examination.

Figure 4. Subtest – total test scores (minus itself) correlation matrix
(January 2005) (n=1669)

	Max. test score	r Total test score minus itself	p	M of subtests	SD of subtests	M Total test score minus itself	SD Total test score minus itself
Grammar test	20	**.671****	.000	11.81	3.427	88.75	22.276
Reading Comp.	20	**.691****	.000	12.34	4.150	88.21	21.654
Writing test	20	**.619****	.000	12.33	3.469	88.23	22.406
Written mediation (L1-L2)	10	**.726****	.000	5.55	1.996	95.00	23.218
Listening Comp.	20	**.637****	.000	10.24	3.656	90.31	22.214
Introductory Conv.	20	**.778****	.000	14.90	3.617	85.66	21.788
Speaking test	20	**.780****	.000	12.38	4.484	88.17	21.047
Situation	20	**.773****	.000	14.24	4.364	86.31	21.177
Oral mediation (L2-L1)	10	**.763****	.000	6.76	2.349	93.80	22.866

** Correlation is significant at the 0.01 level (2-tailed).
Source: own calculations using SPSS

The correlation coefficients in case of Grammar test (r = .671), Reading comprehension test (r = .691), Writing test (r = .616) and Listening comprehension test (r = .637) are on the low side. In the case of all other tests including Written mediation test (r = .726) are above the optimal .7 value. These results show a rather strong correlation between the subtests and the total scores and show how important these subtests are in the test battery. Non-parametric analysis in each case shows a

lower correlation coefficient of Written mediation test but it is near to the desired value.

Figure 5. shows the Subtest – total score (minus self) correlation of Written mediation task (L1 →L2) in 18 examination periods. The correlation coefficients are generally between .6 and .7, a bit lower than the optimal, but convincingly higher than inter-subtest correlation coefficients.

Figure 5. Subtest – total score (minus self) correlation of Written mediation task (L1 →L2) in 18 examination periods

		r	p	M	SD	M Total test score minus Mediation score	SD Total test score minus Mediation score	N
1	Sept. 2000	**.676**	.000	4.88	2.237	89.95	26.069	229
2	Jan. 2001	**.668**	.000	5.88	2.241	91.92	27.253	360
3	May 2001	**.642**	.000	6.27	2.134	98.46	21.735	2072
4	May 2002	**.646**	.000	5.75	2.140	97.63	21.599	2754
5	Jan 2003	**.605**	.000	5.86	2.182	96.21	20.480	1630
6	May 2003	**.595**	.000	5.70	2.034	97.86	20.306	2577
7	Sept. 2003	**.608**	.000	5.56	2,065	94.02	19.519	1180
8	Jan. 2004	**.601**	.000	5.20	2.156	96.91	20.378	1532
9	May 2004	**.584**	.000	5.39	1.928	94.90	20.367	2540
10	Sept. 2004	**.606**	.000	6.07	1.927	93.23	19.681	1280
11	Jan. 2005	**.726**	.000	5.55	1.996	95.00	23.218	1669
12	March 2005	**.581**	.000	5.81	1.994	95.45	20.551	319
13	May 2005	**.596**	.000	5.75	2.047	97.95	21.670	2286
14	Sept. 2005	**.656**	.000	5.91	1.9.50	89.99	19.449	1399
15	Jan. 2006	**.632**	.000	5.92	2.193	97.81	22.802	1596
16	May 2006	**.706**	.000	6.16	2.124	98.41	22.931	1922
17	Sept. 2006	**.608**	.000	5.24	2.191	90.28	24.404	1103
18	Jan 2007	**.672**	.000	5.74	2.226	98.62	23.655	1384

Source: own calculations using SPSS

As a result of the correlation analyses it is observable that the Written mediation test does not behave in a different way when compared with the other examination tasks, and does not show such a high common proportion of variance with them that would indicate that the mediation task measures the same skills as the other examination tasks (subtests). It was obvious that neither the written mediation task nor the other examination tasks can be omitted from the test battery.

5.2. *Reliability analysis of the subtests*

Reliability of tests is the extent to which test scores are consistent. According to classical item analysis in test construction it is useful to calculate the reliability index (coefficient Alpha) of test items in order to know which test item increases or decreases the reliability of the whole test. Those items that decrease the reliability of the test should be omitted.

In this case I considered each subtest as an item of the whole test (examination) battery and calculated the reliability indices of subtests (Figure 6).

Figure 6. Reliability analysis of subtests
(January 2005) (n=1669)

| Item-Total Statistics	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Grammar test	90.47	369.586	.656	**.850**
Reading Comp.	89.13	377.324	.579	**.858**
Writing test	88.94	381.048	.524	**.864**
Written mediation (L1-L2)	95.44	422.064	.581	**.862**
Listening Comp.	91.73	376.122	.541	**.862**
Introductory Conv.	85.72	382.069	.678	**.849**
Speaking test	88.04	351.327	.683	**.848**
Situation	86.53	350.619	.698	**.846**
Oral mediation (L2-L1)	93.99	413.115	.681	**.856**

Full Test Reliability Statistics			
	Cronbach's Alpha	N of Cases	N of Items
	.869	1669	9

Source: own calculations using SPSS

As the results show (Figure 7) if *Written mediation test* were deleted from the test battery the reliability index (Cronbach alpha) of the test batteries would decrease in all examination periods. It means that the total scores of the whole examination and the examination itself would be less reliable.

Figure 7. Omitting written mediation (L1 → L2) task from the test battery reduces reliability of the language examinations (18 examination periods)
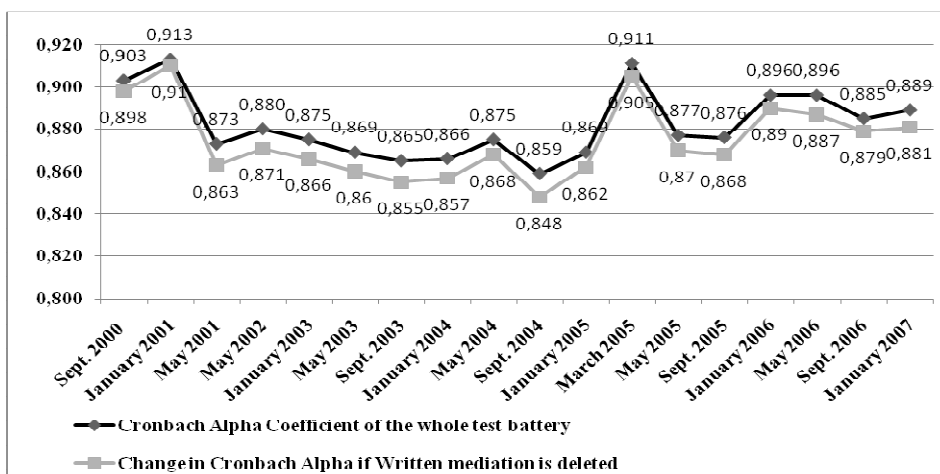
Source: own calculations

5.3. *Factor analysis*

The procedure of factor analysis is used to reduce the number of variables accounting for test performance by identifying the common underlying factor (or factors) shared by a series of tests in the test battery.

The results of the KMO (KMO≥0.8) and Bartlett's Test showed that the data were suitable for factor analysis. I used PCA method as it „gives us a way of discovering factors that underlie language performance and ways of testing the relationship among them." (Hatch − Lazaroton 1991) PCA initially attempts to fit as much of the data from the correlation matrix of all variables entered, into a single principle component, in other words, it attempts to explain through the first factor as much of the variability in the data as possible. Once it has done this it trawls through data again, looking for the second component which will explain as much of the remaining variance as possible (Green – Weir 2001).

In the case of the presented examination period all subtests loaded positively on the first factor with .642 or above (Figure 9), which can be considered to indicate a substantial link between them. They all load on the same factor as the first factor represents *general linguistic ability* (Green – Weir 2001).

Figure 8. Factor analysis (January 2005) (n=1669)

Component Matrix(a)

	Component				
	1	2	3	4	5
Grammar test	**.763**	.274	-.152	-.359	-.040
Reading Comp.	**.687**	.267	.441	-.379	.198
Writing test	**.642**	.326	-.523	.166	.398
Written mediation (L1-L2)	**.690**	.371	-.151	.067	-.563
Listening Comp.	**.656**	.319	.422	.488	.076
Introductory Conv.	**.751**	-.453	-.012	.158	-.009
Speaking test	**.759**	-.481	-.049	-.088	.030
Situation	**.776**	-.466	.028	.012	-.050
% of Variance	51.435	14.320	8.683	7.187	6.564
% Total Variance Explained by 5 factors					
88.188					

Extraction Method: Principal Component Analysis. KMO = .863
a. 5 components extracted.
Source: own calculations using SPSS

More complex procedures can be followed such as rotation of the factors to see if any clearer solutions present themselves in order to reveal the underlying factors or components. In analysing test results Varimax rotation is the most commonly used procedure (Green – Weir 2001). The result of the Varimax rotation of test scores (Figure 9) showed that the different subtests representing different language skills fell on different factors with rather high factor loadings, except for *Grammar test* that does not represent a separate language skill.

Figure 9. Factor analysis – Varimax rotation (January 2005) (n=1669)

Rotated Component Matrix(a)

	Component				
	1	2	3	4	5
Grammar test	.287	**.586**	**.485**	**.388**	-.019
Reading Comp.	.218	**.875**	.116	.084	.294
Writing test	.191	.140	.187	**.925**	.169
Written mediation (L1-L2)	.199	.159	**.900**	.175	.230
Listening Comp.	.200	.239	.206	.167	**.887**
Introductory Conv.	**.842**	.064	.131	.144	.208
Speaking test	**.858**	.223	.106	.147	.013
Situation	**.862**	.179	.161	.078	.127
% of Variance	30.423	16.206	14.875	14.012	12.673

% Total Variance Explained by 5 factors
88.188

Extraction Method: Principal Component Analysis.   KMO = .863
Rotation Method: Varimax with Kaiser Normalization
  a. Rotation converged in 6 iterations
Source: own calculations using SPSS

Written mediation appears in the 3rd factor in the observed examination period (January 2005) and accounts for 15% of variance of the test battery.

In the 18 examination periods foreign language mediation tests appear in the 2nd to 5th factors. Most of the times it falls on the 2nd and 3rd factor. Factor loadings are between .730 and .911. On average, in the 18 examination periods, foreign language mediation tests account for 17% of variance. In all cases the factors correlate well with total test scores.

Summarising the results of the factor analysis show that foreign language mediation skill represented by Written mediation tasks is one of the components of general language ability. It can be clearly separated from other language skills with its high loading of variance and on average it explains 17% of the variance of the language skills in the test battery so it represents an important part of language knowledge.

5.4. *Multiple linear regression analysis (Method FORWARD)*

Multiple regression analysis reveals the common subskills between mediation tests and other subtests.

The problematic point was to keep reliability of variables in the test battery on nearly the same level. (The test battery contains subtests that are subjectively scored, so it was important to elaborate and use a system of evaluation criteria that increases the objectivity of evaluation of these tests. Surveying the elaboration of this system is beyond the scope of this paper.)

In this paper I cannot present the whole process and all the steps of multiple linear regression analysis of the test scores of the examination period under research,

so I will show only the Model Summary (Figure 10) representing the most important result of the analysis.

Figure 10. Model summary of regression analysis (January 2005) (n=1669)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.561(a)	.315	.315	1.650
2	.613(b)	.376	.375	1.576
3	.630(c)	.397	.396	1.549
4	.637(d)	.405	.404	1.539
5	.638(e)	.407	**.405**	1.537

a Predictors: (Constant), Grammar test
b Predictors: (Constant), Grammar test, Listening Comp.
c Predictors: (Constant), Grammar test, Listening Comp., Writing test
d Predictors: (Constant), Grammar test, Listening Comp., Writing test, Situation
e Predictors: (Constant), Grammar test, Listening Comp., Writing test, Situation, Reading Comp.
f **Dependent Variable: Written mediation (L1-L2)**
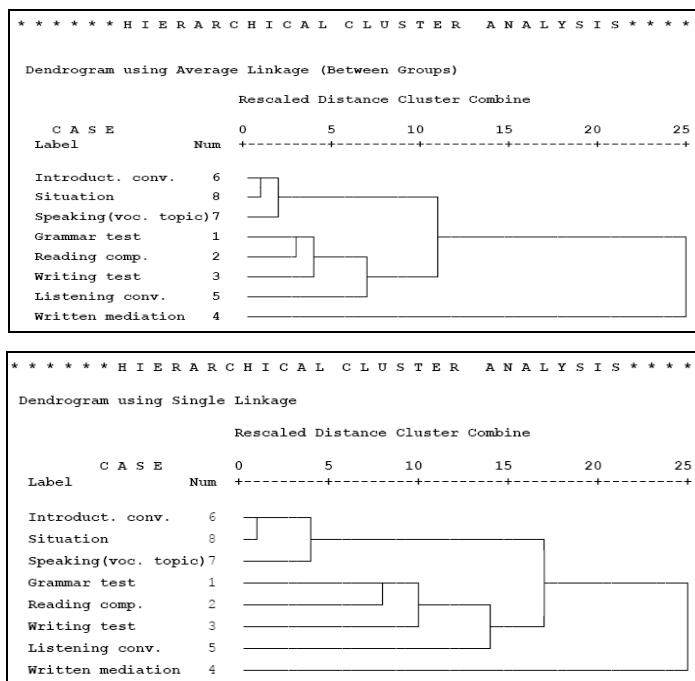*Source:* own calculations using SPSS

In this case (January 2005) Introductory conversation, and Speaking test on vocational topic were dropped out from the model. The Model Summary reveals that approximately 41% of the variance of Written mediation test can be explained by the help of other variables. It indicates the presence of common subskills. The remaining 59% indicates the existence of subskills that are present only in mediation skill.

Multiple regression analyses revealed the common subskills (the proportion of variation in the dependent variable explained by the regression model) between mediation tasks and other tasks. On average 54% of the variance of the mediation task is not explained by other examination tasks so it represents subskills that can be attributed only to foreign language mediation skills.

5.5. Cluster analysis

Clustering is the assignment of a set of observations into subsets (called *clusters*) so that observations in the same cluster are similar in some sense. Cluster analysis sorts through the raw data and groups them into clusters. A cluster is a group of relatively homogeneous cases or observations. Objects in a cluster are similar to each other. They are also dissimilar to objects outside the cluster, particularly objects in other clusters. So I decided to use this method to see how the different subtests relate to each other (Figure 11).

Figure 11. Cluster analysis – Dendrograms using Average and Single Linkage
(January 2005) (n=1662)

Source: own calculations using SPSS

When observing the 18 examination periods in each case *Written mediation test* formed an independent cluster.

Summarising the results of cluster analyses, they also confirmed that foreign language mediation is an independent language skill as mediation tasks are sharply separated from other examination tasks forming an independent cluster.

6. Conclusions

My research verified my hypothesis that written mediation examination task measures a segment of language knowledge independently; skills (subskills) appearing in it are not present in other examination tasks. Foreign language mediation is an independent language skill; it is as legitimate as traditional basic language skills and its usage as an examination task contributes to a more complex notion of the test-takers' language knowledge therefore bilingual language examinations can measure skills that are not measureable by monolingual examination systems.

References

Alderson, Ch. J. 2001: The lift is being fixed. You will be unbearable today. Script of Plenary Session. 10th Hungarian Macmillan Conference, Budapest.

Alderson, Ch. J., Clapham, C. és Wall, D. 1995: *Language Test Construction and Evaluation*. Cambridge University Press, Cambridge.

Bachman, L. F. – Palmer, A. S. 2000: *Language Testing in Practice*. Oxford University Press, Oxford.

Bárdos, J. 1997: *A nyelvtanítás története és a módszerfogalom tartalma*. Veszprémi Egyetemi Kiadó, Veszprém.

Bárdos, J. 2005: *Élő nyelvtanítás-történet.* Nemzeti Tankönyvkiadó, Budapest.

Dévény, Á. – Szőke, A. 2007: Milyen nyelvtudást vár el a munkaadó a munkavállalótól. *Stratégiák 2007 és 2013 között. Tudományos Évkönyv 2006.* BGF, Budapest, 319-330. p.

Einhorn, Á. (szerk.). 1998: *Vizsgatárgyak, vizsgamodellek I*. Országos Közoktatási Intézet, Budapest.

F. Silye, M. 2004: A szaknyelvoktatás és a szaknyelvi tudást felhasználók igényeinek elemzése és megjelenítése egy angol szaknyelvoktatási program-modellben. PhD. Doktori Értekezés. ELTE, PPK. Neveléstudományi Doktori Iskola, Budapest.

Fekete, H. 2001: Az írásbeli nyelvvizsga a számok tükrében. Az ITK ORIGÓ angol középfok. In Fekete H. (szerk.):*Nyelvi Mérce,* Idegennyelvi Továbbképző Központ, Budapest, 2001/1-2, 54-58. p.

Fekete, H. 2002: Az írásbeli nyelvvizsga a számok tükrében. Az ITK ORIGÓ német középfok. In Fekete H. (szerk.):*Nyelvi Mérce,* Idegennyelvi Továbbképző Központ, Budapest, 2002/1-2, 45-56. p.

Green, R. – Weir, C. 2001: Using statistics for the analysis of test an evaluation data. (Manuscript).

Hatch, E.M. - Lazaraton, A. 1991: *The research manual: Design and statistics for applied linguistics.* Newbury House, New York.

Heltai, P. 2001: Iskolai tantárgy-e az idegen nyelv? *Nyelvi Mérce, ITK, Budapest* Vol. 1 Issue 1-2. 13-19. p.

Klaudy, K. 1984: Hogyan alkalmazható az aktuális tagolás elmélete a fordítás oktatásában? *Magyar Nyelvőr*, Vol. 108 Issue 3, 325-333. p.

Klaudy, K. 1986a: A fordítás helye és szerepe a nyelvoktatásban: a „valódi" fordítás a középiskolai nyelvoktatásban. In Lengyel, Zs. (szerk.): *Az idegen nyelvi nevelés-oktatás néhány iránya és lehetősége*. Oktatáskutató Intézet, Budapest, 353-371. p.

Klaudy, K. 1986b: A fordítás oktatásának elméleti alapjai. *Felsőoktatási Szemle* XXXV. 5, 277-283. p.

Klaudy, K. 1990: Átváltási műveletek a fordításban. *Idegen Nyelvek tanítása* 2., 33-45 p.

Major, É. 2000: Milyenfajta angol nyelvtudásra van szükség a nyelvigényes munkakörökben? Vizsgálat a munkaadók elvárásainak felmérésére. *Modern Nyelvoktatás*. 2000, Vol. 6 Issue 1., 33-49 p.

Nikolov, M. - Fekete, H. - Major, É. (eds.). 1999: *English language education in Hungary. A Baseline Study.* The British Council Hungary, Budapest.

Szabari, K. 2001: Anyanyelvtudás, idegennyelvtudás, nyelvi közvetítés. *Nyelvi Mérce.* Nyelvvizsgáztatók és Nyelvtanárok Lapja. ITK, Budapest, Vol. 1 Issue 1-2, 7-13. p.

Teemant, A. – Varga, Zs. – Heltai, P. 1993: *Hungary's Nationwide Needs Analysis of Vocationally-Oriented Foreign Language Learning: Student, Teacher, and Business Community Perspectives.* Művelődésügyi és Közoktatási Minisztérium, a Munkaügyi Minisztérium, az Egyesült Államok Tájékoztatási Hivatala és az Európa Tanács Modern Nyelvek Szekciója, Budapest.