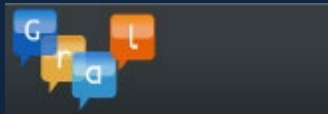# The identification of learner profiles and the role of sound-symbol correspondence

Mª del Mar Suárez
mmsuarez@ub.edu
Universitat de Barcelona

UNIVERSITAT DE BARCELONA

WESTMINSTER
INTERNATIONAL UNIVERSITY IN TASHKENT
An Accredited Institution of the University of Westminster (UK)

**USING
CLUSTER
ANALYSIS**

Using cluster
analysis

Research
questions

Methodology

Results

Discussion

Further research

**LITERATURE
REVIEW**

RESEARCH QUESTIONS

Using cluster
analysis

Literature review

Research
questions

Results

Discussion

Further research

METHODOLOGY

RESULTS

Using cluster analysis

Literature review

Research questions

Methodology

Results

Further research

DISCUSSION

Using cluster analysis
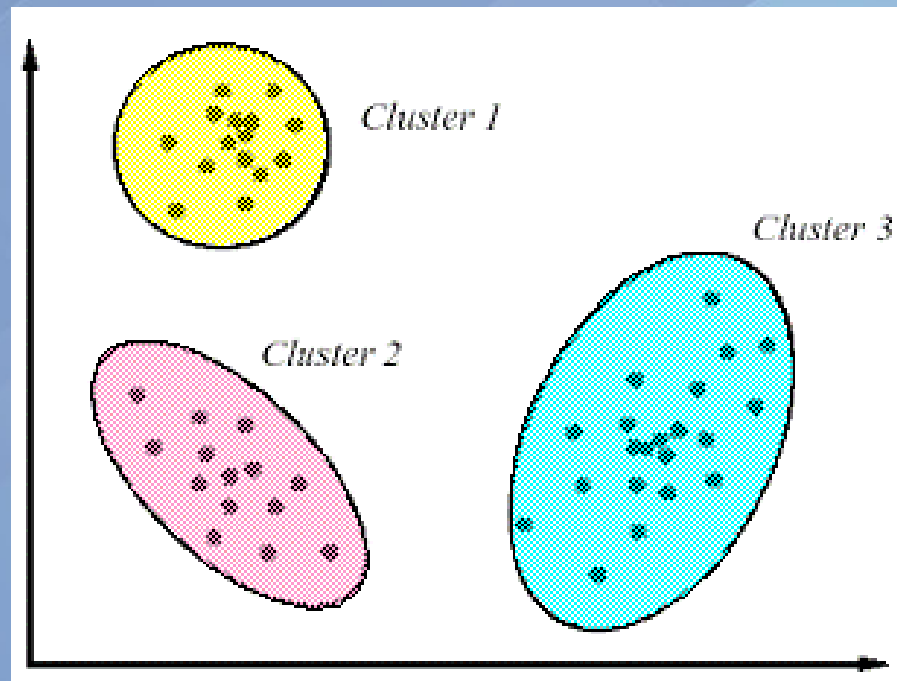
Literature review

Research questions

Methodology

Results

Discussion

FURTHER RESEARCH

Using cluster analysis

Multivariate exploratory technique used to "subdivide a set of objects into homogeneous subgroups or into a hierarchical arrangement of homogeneous subgroups" (Lorr, 1983)
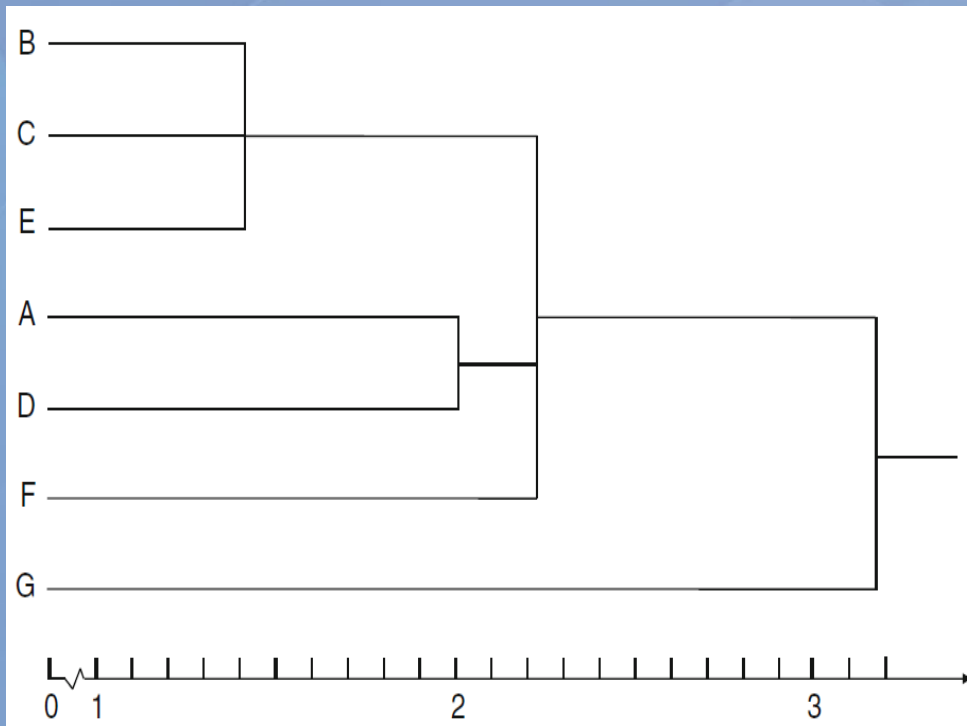
1

Using cluster analysis

2

 Multivariate exploratory technique used to "subdivide a set of objects into homogeneous subgroups or into a hierarchical arrangement of homogeneous subgroups" (Lorr, 1983)

Hierarchical →Dendrogram

Using cluster analysis

Multivariate exploratory technique used to "subdivide a set of objects into homogeneous subgroups or into a hierarchical arrangement of homogeneous subgroups" (Lorr, 1983)

2

K- means (non-hierarchical method):

1. Identify k clusters to assess how distinct our clusters are

2. Using the cluster centers identified before

3. F- values → how well the dimension discriminates

Aims of cluster analysis

1   Identify **natural** clusters within a mixture of entities → several distinguishable populations

2   Construct a useful **conceptua**

- Outliers
- Not normally distributed data
BUT
- representativeness of the sample
- multicollinearity

3   Generate hypotheses to be te unsuspected clusters

4   Identify homogeneous subgroups → patterns useful for prediction

Protocol

Why use cluster analysis in aptitude research?

FL aptitude: one of the best predictors of language learning success
(Abrahamsson & Hyltenstam, 2008; Dörnyei, 2005; Sawyer & Ranta, 2001; Skehan, 1998)

Correlational studies

- Total score as best predictor
- Norming studies
- Different aptitude tests, f.ex. MLAT & MLAT-E
  (Harper & Kieser, 1977; Hauptman, 1971; Kiss, 2009; Kiss & Nikolov, 2005; Suárez, 2010)
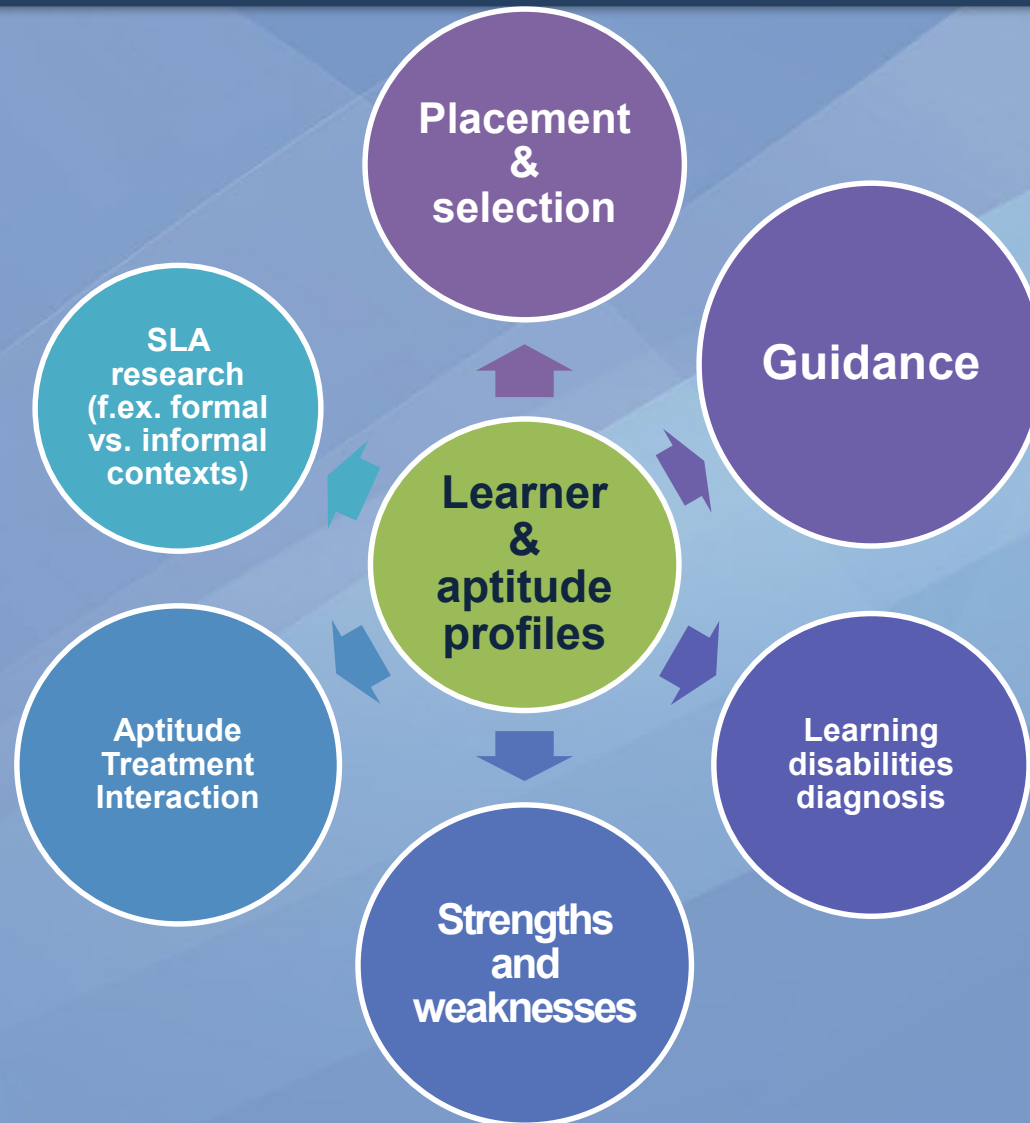
Multicomponential

- Test batteries tapping different abilities

Learner profiles vs. aptitude profiles

- Including or not proficiency measures

Why use cluster analysis in aptitude research?

Literature review

1 ## Language aptitude components (Carroll, 1981)

Phonemic coding ability

- the ability to identify and memorize new sounds or strings of sounds

Grammatical sensitivity

- the ability to understand how words function grammatically in sentences

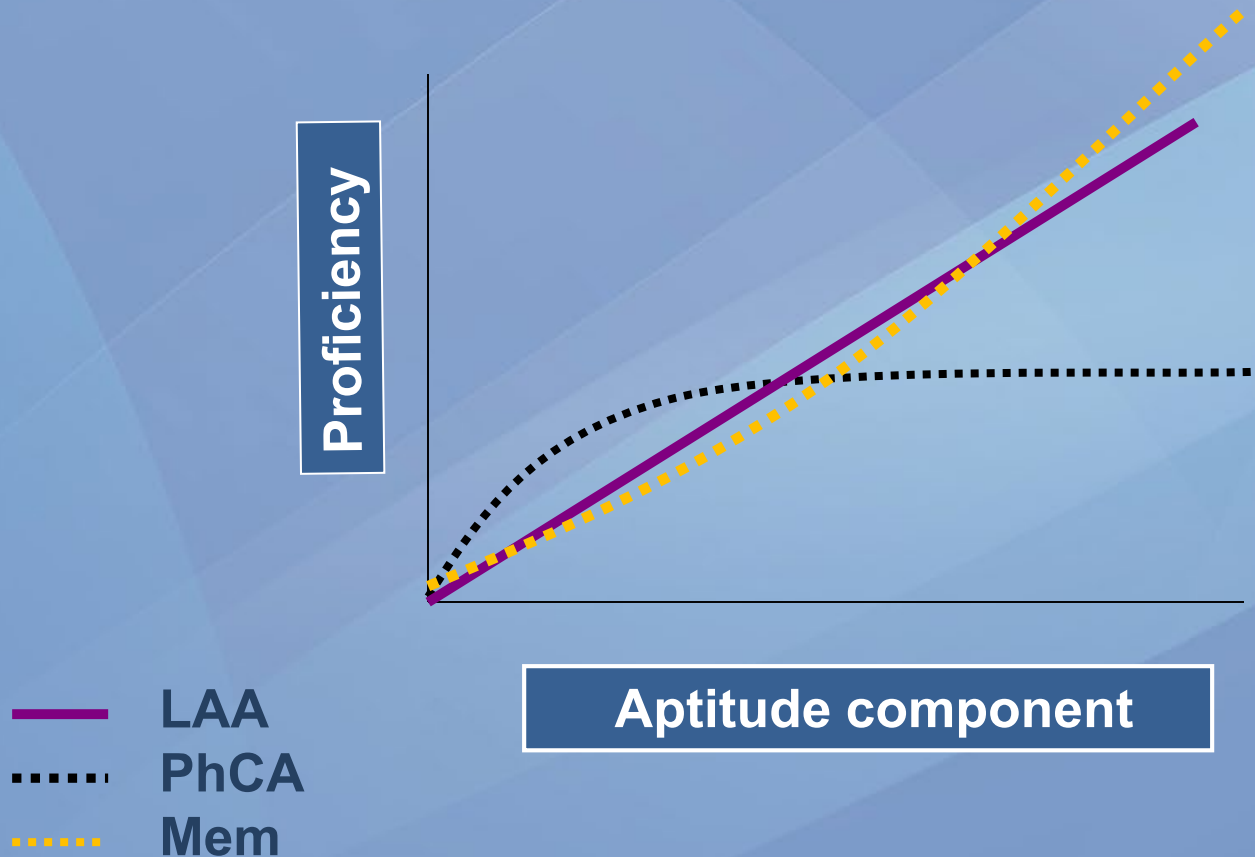### Inductive language learning ability

- the ability to infer grammatical rules from language samples

Rote learning ability for FL materials

- the ability to learn a large number of semantic-symbol and/or sound-symbol associations in a short period of time

Literature review

2 Aptitude components & proficiency

Literature review

3   Language aptitude components (Carroll, 1967)

MLAT-E (grades 3 to 6)	MLAT (adults)	Construct
Hidden words	Spelling clues	- English vocabulary - sound-symbol association
Matching words	Words in Sentences	- grammatical sensitivity
Finding rhymes	-	- hear and make distinctions between speech sounds
Number learning	Number learning	- rote memory - aural comprehension

Literature review

4a MLAT-E parts (Carroll, 1967)

MLAT-E Part 1 – Hidden words

| 1. wntr | A champion | B season | C liquid | D happy |
| 4. snak | A hard wood | B tease | C reptile | D type of shoe |

MLAT-E Part 2 – Matching sentences

1. Yesterday, Mary caught a FISH at the lake.
    Cindy cut a cake with a knife.
    ☐ ☐ ☐ ☐ ☐ ☐ ☐

2. Amy SANG a pretty song to her class.
    James throws big rocks into the lake.
    ☐   ☐ ☐ ☐ ☐ ☐ ☐ ☐

Literature review

4b **MLAT-E parts** (Carroll, 1967)

MLAT-E Part 3 – Finding Rhymes

1. TIME ……☐ tame ….☐ tide ….☐dime ….☐shin

2. RAIN ….. ☐vine …..☐cane …☐keen ….☐fine

Literature review

4c **MLAT-E parts** (Carroll, 1967)

MLAT-E Part 4 – Number Learning

Units
• ba = one
• baba = two
• dee = three

Tens
• tu = twenty
• ti = thirty

ti 30 **+** ba 1 **=** thirty – one 31

Literature review

5a SLA research using cluster analysis

Proficiency as criterion variable

- Motivation profiles (Dörnyei et al., 2006)
- Vocabulary learner strategies profiles (Kojic-Sabo & Lightbown, 1999)

Proficiency included in the cluster analysis

- LAA in learner profiles → Metalinguistic awareness in L1 & L2 and L2 proficiency over time (Ranta, 2002)

- Learner differences in strategy use, will to learn and achievement over time (Yamamori et al., 2003)

- Learner cognitive profiles including MLAT-4 Words in Sentences, age and criterion test score (Skehan, 1986)

Literature review

5b | **Learner profiles as of cluster analysis**

High achievers

- Strong on LAA (Ranta, 2002)
- Strong on both LAA and memory OR strong on either LAA (*syntactically oriented students*) or memory (*lexically-oriented students*) (Skehan, 1986, 1998, 2002)
- Auditory abilities more relevant in younger students (Skehan, 1986)

Low achievers

- Average or weak on LAA (Ranta, 2002)
- Very poor memory except associative memory, average language ability (Skehan, 1986)
- Intelligent but with poor language ability (Skehan, 1986)
- High linguistic ability and memory but average IQ (Skehan, 1986)

Literature review

5b **How about young(er) learners (grades 3-4, and beyond?**

1. **Concrete operational thinking stage vs formal operational thinking (Piaget)**

1. **First stages of literacy development**

Research questions

1 Do language learner profiles change over time in young learners in grades from 3 to 7?

2 What learner profiles do younger high achievers and low ~~ers~~ have?

3 What determines achievement in younger learners?

Methodology / Protocol

1 **Selection of a representative and adequately large sample of subjects**

▶ **Context: 6 schools in Barcelona and Lleida**

▶ **Participants: bilingual Catalan-Spanish**

▶ **English as a FL at school**

◗ **Participants' mean age and aptitude measures**

	Grade	3	4	5	6	7	All
	Mean age	8.8	9.9	10.9	11.7	12.9	
Group 1	N MLAT-ES/EC	43	57	57	58	65	280
Group 2	N MLAT-EC/ES	55	62	30	60	62	269
	Total N	98	119	87	118	127	549

Methodology / Protocol

2 **Selection of set of variables from a domain of similarity**

❯**MLAT-EC Part scores**

❯ **Proficiency measures: cloze passage measure (different versions for different grades) → integrative measure**

3 **Description and measurement of each entity**

MLAT- EC Parts and constructs

Parts	MLAT-EC (grades 3 to 7)	MLAT (adults)	Construct
1	Paraules ocultes	Spelling Clues	- vocabulary **- sound-symbol association**
2	Paraules que es corresponen	Words in Sentences	- grammatical sensitivity
3	Paraules que rimen	-	**- hear and make distinctions between speech sounds**
4	Aprenguem números	Number Learning	- rote memory - aural comprehension

MLAT-EC Part 1

15. bakka   A és dolça   B part de la cara   **C fa llet**   D amb arrugues

MLAT-EC Part 2

6. Em vaig tallar el DIT amb un ganivet.
   El meu germà s'oblidà les claus a casa.
   ☐   ☐   ☐   ☐   ☐

MLAT- EC Part 3

45. FLABIOL   A pèsol   **B Oriol**   C flascó   D avió

MLAT-EC Part 4

Units	Tens
• co = u (1) • vein = dos (2) • ras = tres (3)	• silca = deu (10) • vinca = vint (20) • rasca = trenta (30)

rasca 30 ➕ **ras 3** 🟰 **trenta-tres 33**

Items removed after item analysis

MLAT-EC > 13 3 > 122
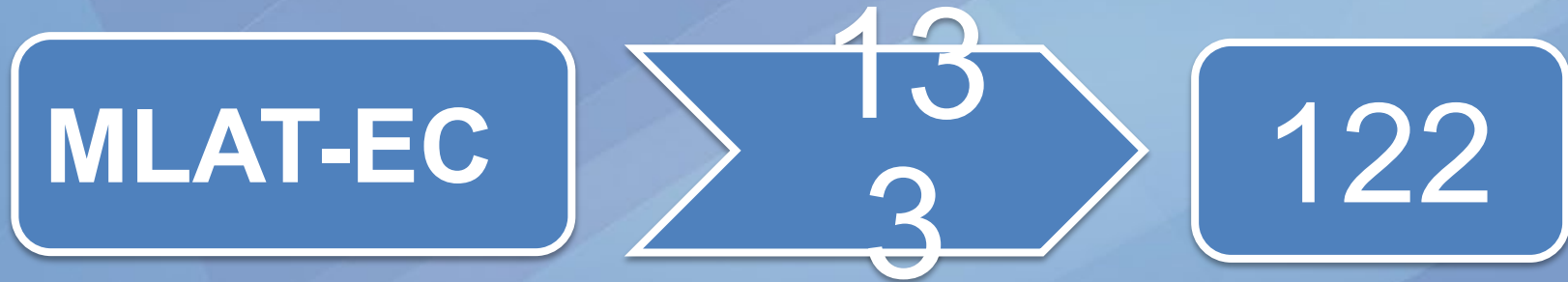
Reliability

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7
Part 1	.934	.900	.871	.878	.873
Part 2	.856	.897	.903	.910	.915
Part 3	.934	.939	.909	.919	.913
Part 4	.936	.909	.922	.893	.941
Total	.957	.960	.950	.944	.961

Mean *p*-values

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7
Part 1	.75	.81	.83	.87	.89
Hidden words	easy	v e r y e a s y			
Part 2	.32	.50	.65	.67	.74
Matching words	difficult	mid-difficult	e a s y		easy
Part 3	.55	.72	.78	.83	.83
Rhyming words	e a s y		v e r y e a s y		
Part 4	.62	.82	.88	.85	.91
Number learning	e a s y	v e r y e a s y			

Results MLAT-EC

Mean raw part scores

Part 1
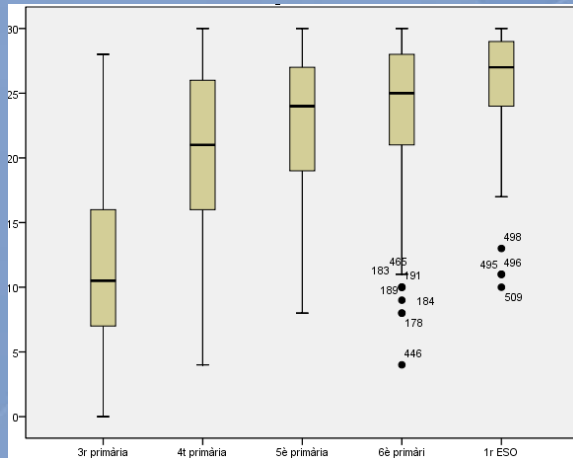
Part 2
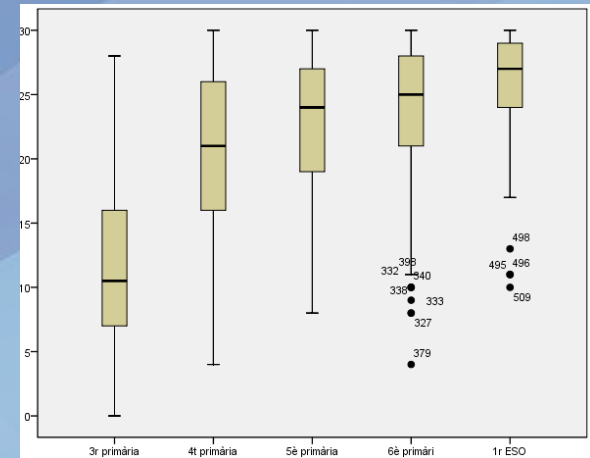
Part 3
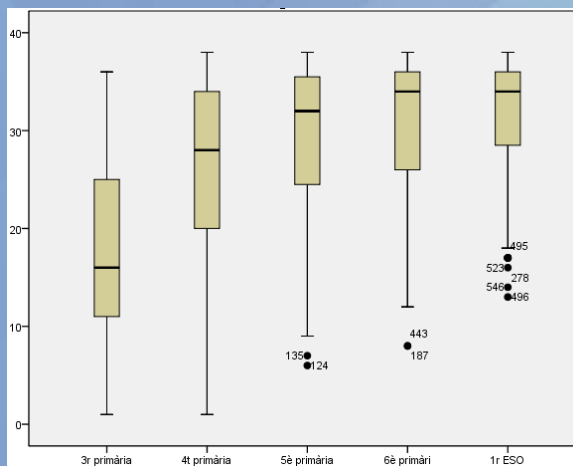
Part 4
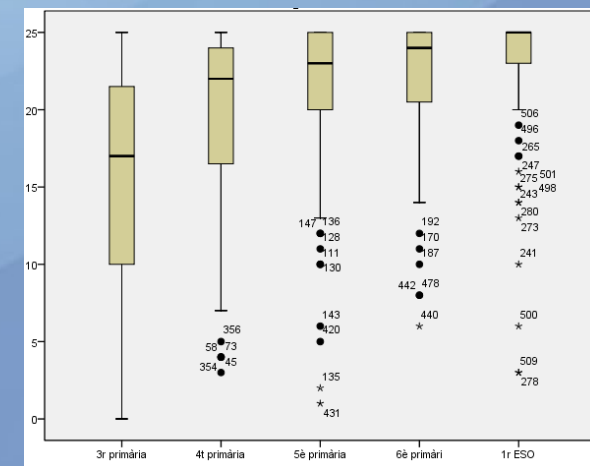
Results MLAT-ES and MLAT-EC

Mean raw total scores

MLAT-EC x/122

Chart data points:
- gr. 3: 55.62
- gr. 3–4: 83.67
- gr. 4: 93.17
- gr. 5: 94.20
- gr. 6: 101.47

Legend: MLAT-EC x/122
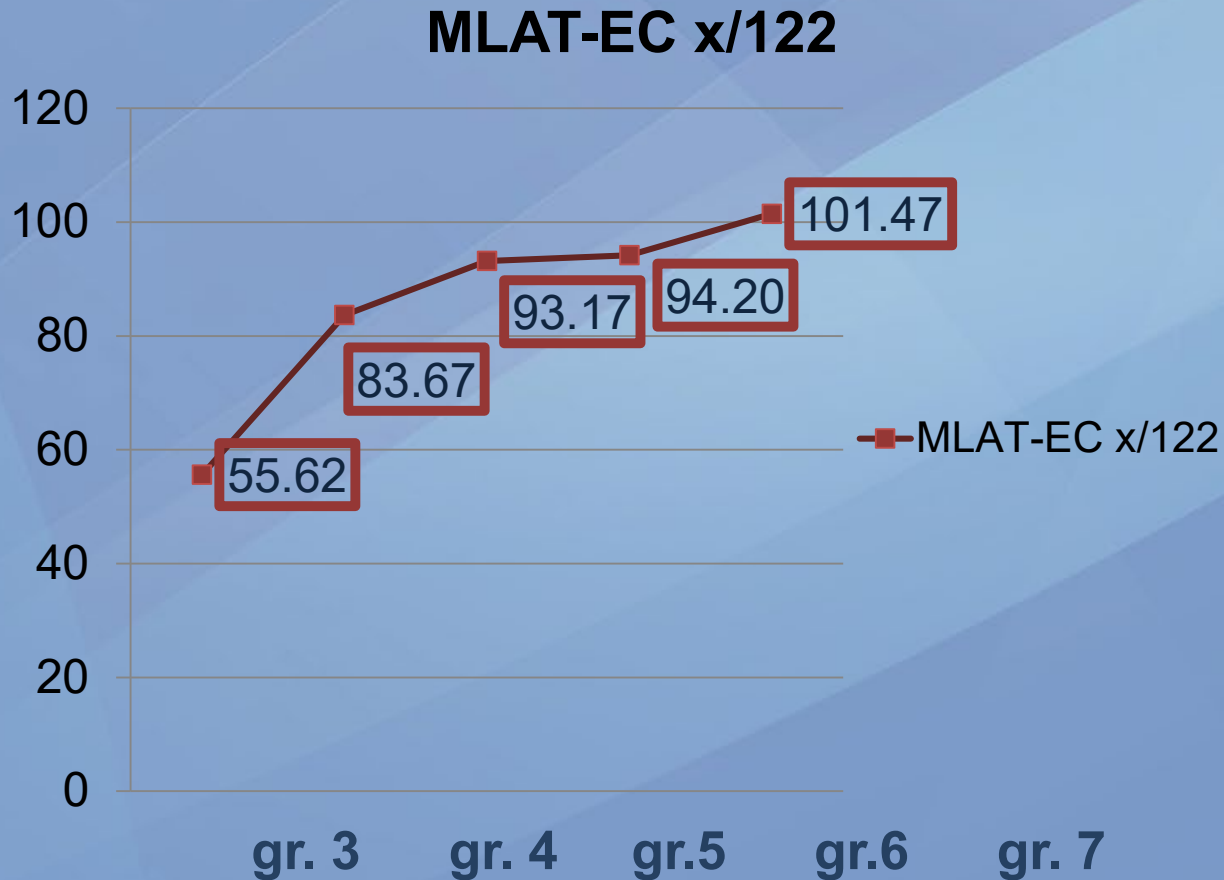
Proficiency measures

Grade 3
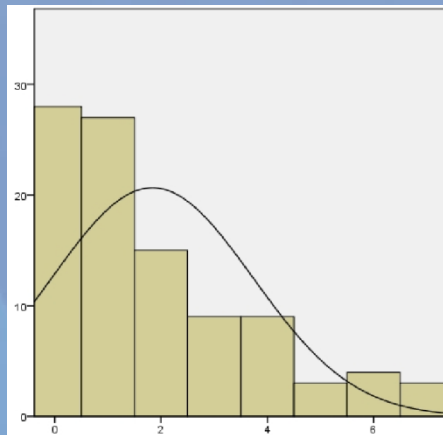
M=1.84 SD=1.89

Grade 4
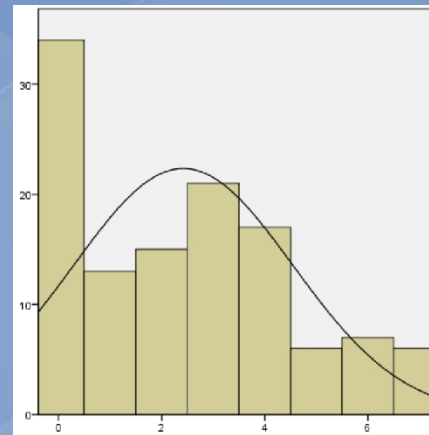
M= 2.42 SD=2.13

Grade 5
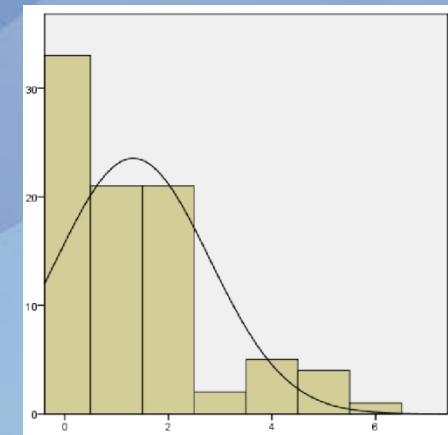
M= 1.32 SD=1.48

Grade 6
M=4.92   SD=4.96
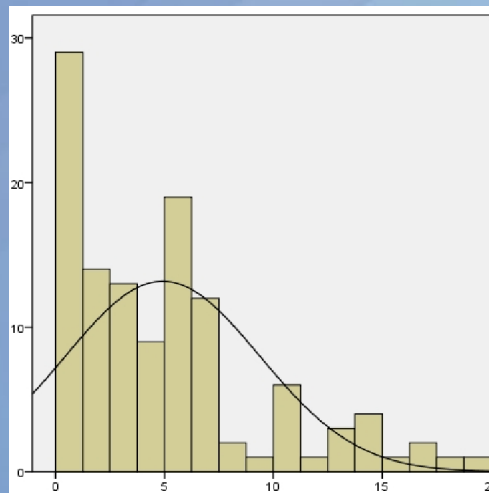
Grade 7
M=6.47   SD=4.70
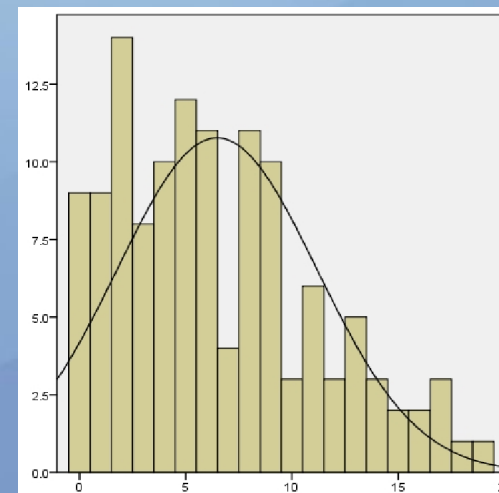
Reliability >.750 in all grades

Factor analysis if numerous variables

4 **Intercorrelation of parts**

	Part 1	Part 2	Part 3	Part 4
Part 2	.616**			
Part 3	.694**	.616**		
Part 4	.513**	.476**	.472**	

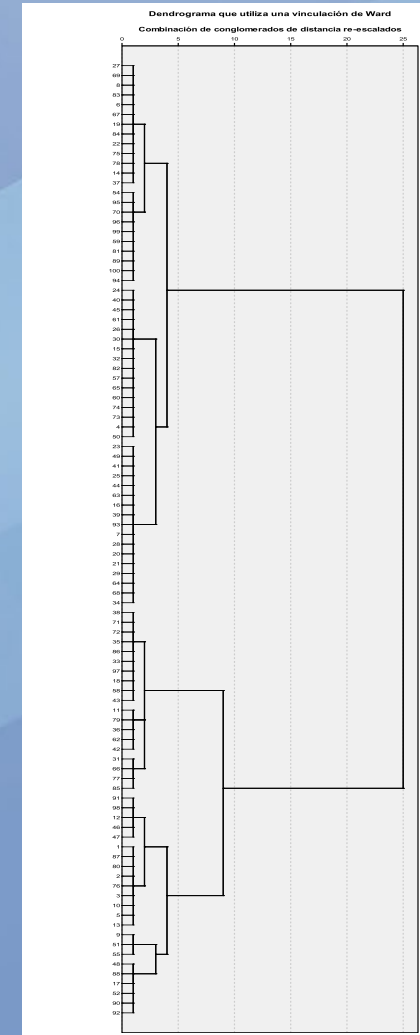****Significant <.01 (two-tailed)**
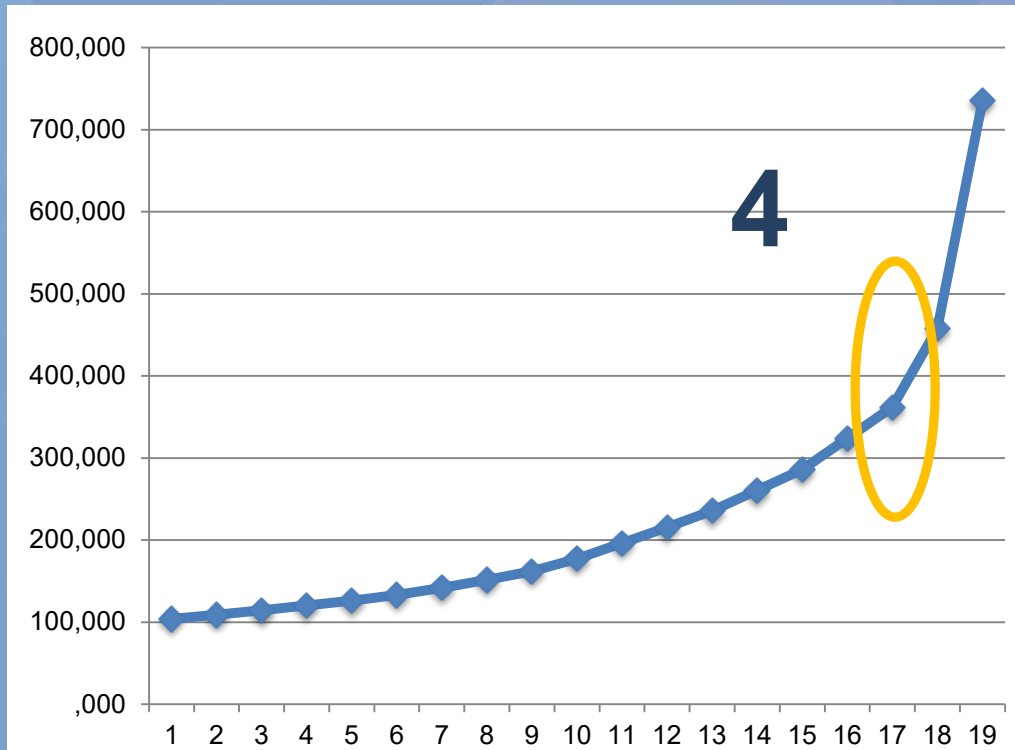
5 **Standardization of measures**

6 **Selection of proximity measure and clustering algorithm**

▶ **Hierarchic: Ward's method, squared Euclidean distance & detection of outliers**
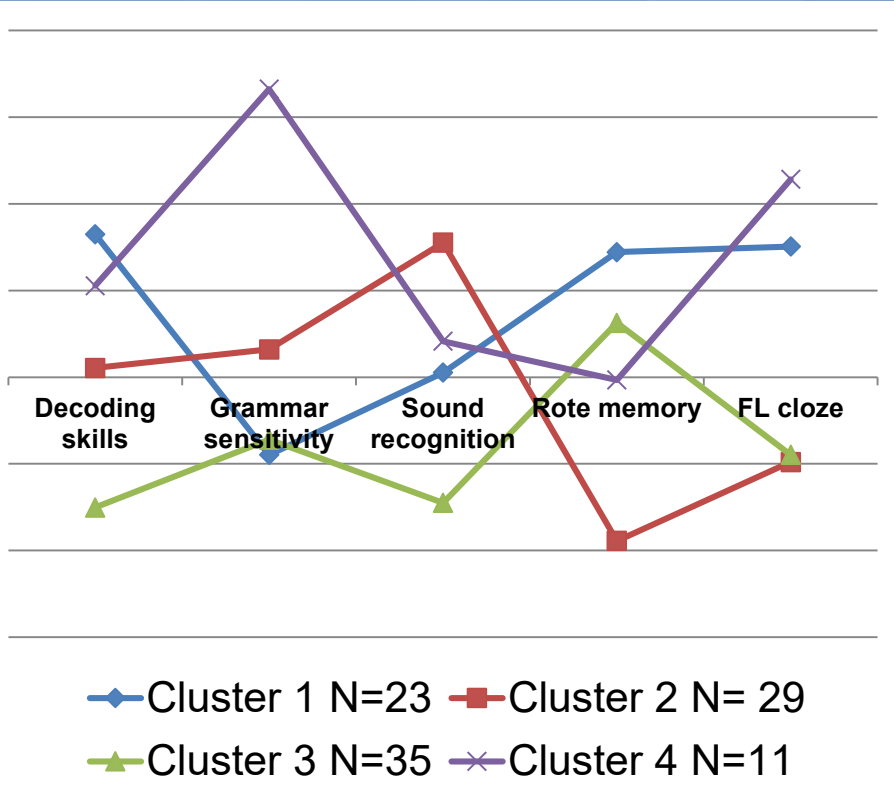
▶ **Non-hierarchical *k* – means cluster**

Results RQ1: Change of profiles over time

Grade 3

Results RQ 1 & 2: profiles over time + high vs. low achievers

Grade 3

Cluster 1 N=23   Cluster 2 N= 29
Cluster 3 N=35   Cluster 4 N=11
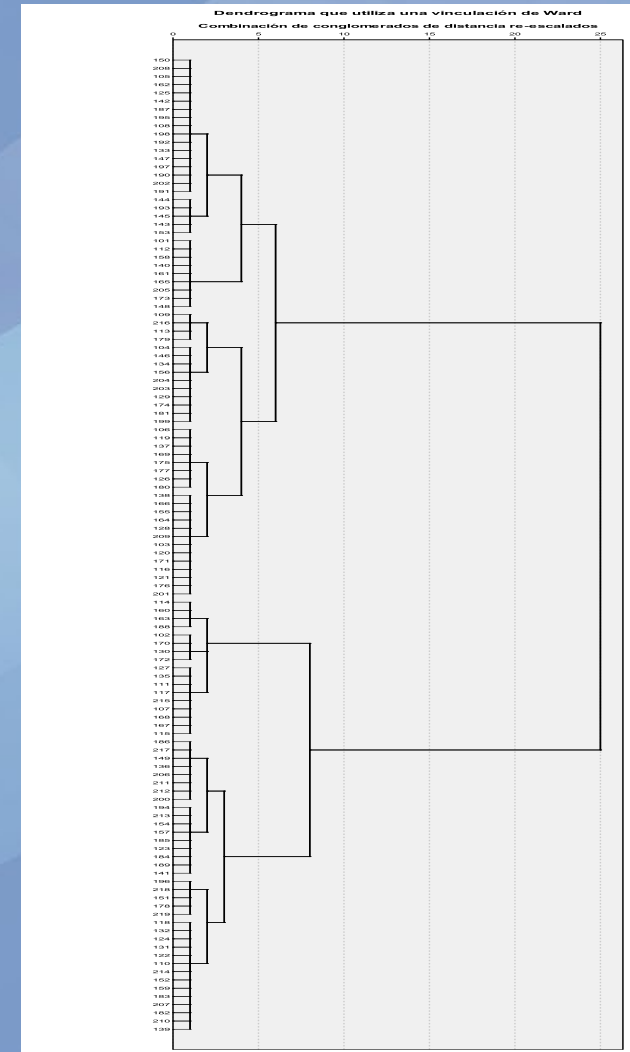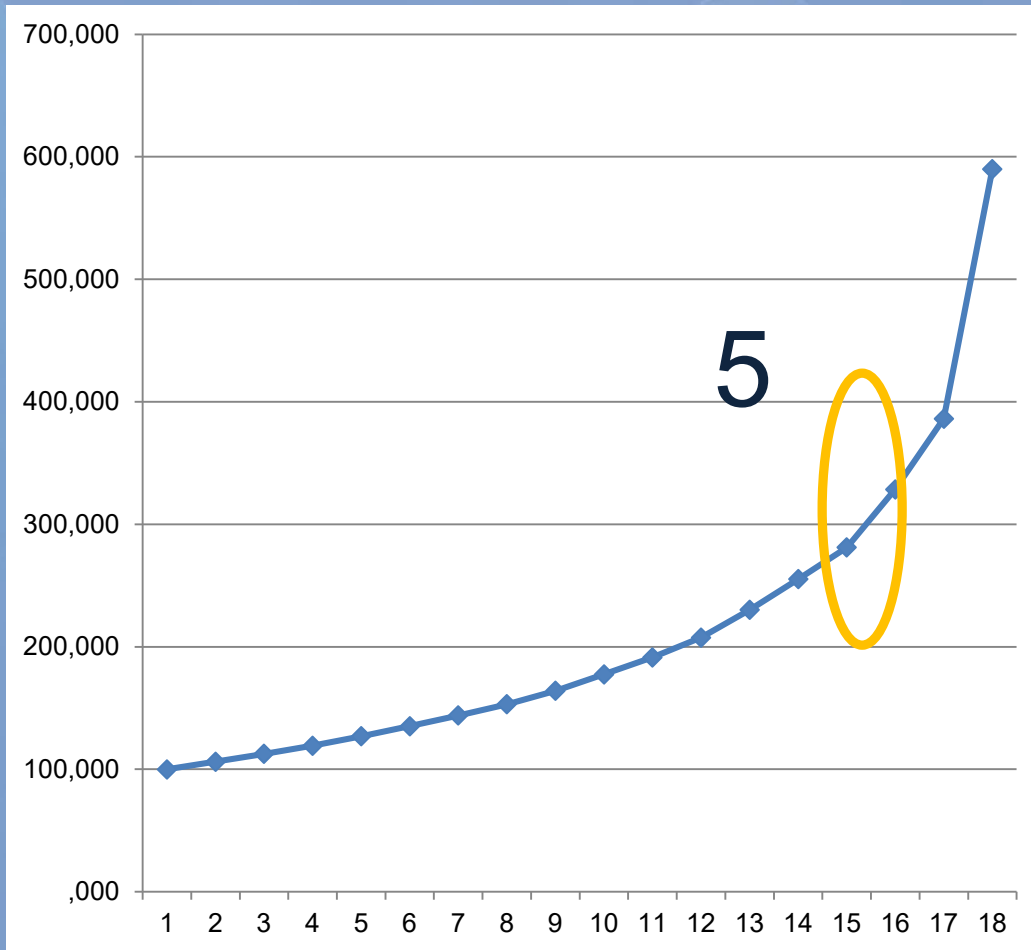
-<u>High decoding skills</u> and average/high sound recognition ability
-<u>EITHER</u> high <u>grammatical sensitivity</u> OR <u>high memory</u> + <u>decoding skills</u> for high achievement
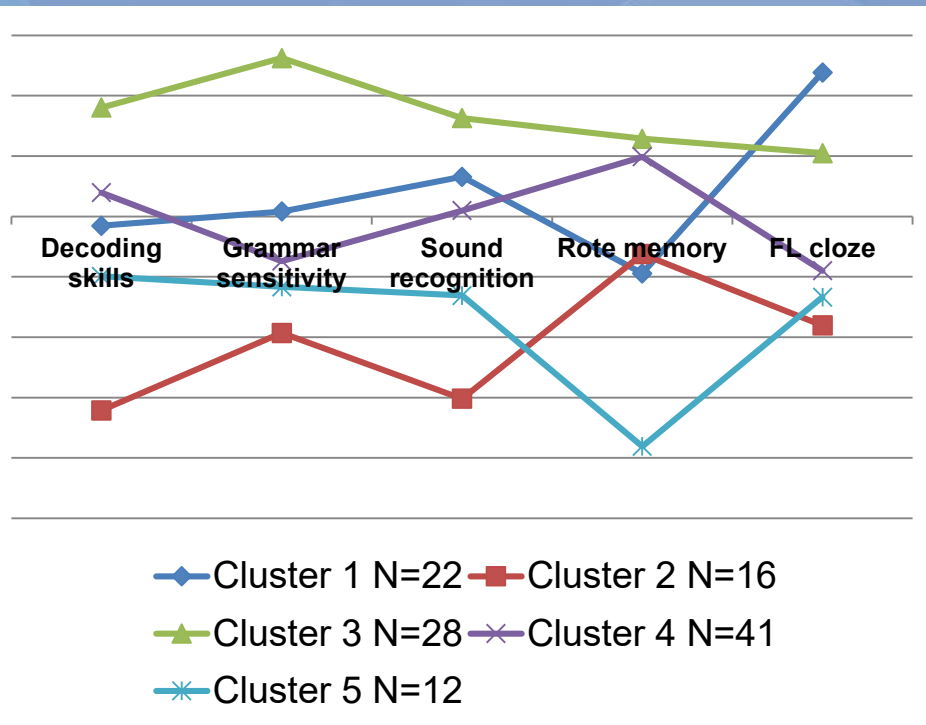- Sound recognition alone, no effect

Results RQ1: Change of profiles over time

Grade 4

5

Dendrograma que utiliza una vinculación de Ward
Combinación de conglomerados de distancia re-escalados

Results RQ 1 & 2: profiles over time + high vs. low achievers

Grade 4

Chart x-axis labels: Decoding skills, Grammar sensitivity, Sound recognition, Rote memory, FL cloze

Legend:
- Cluster 1 N=22
- Cluster 2 N=16
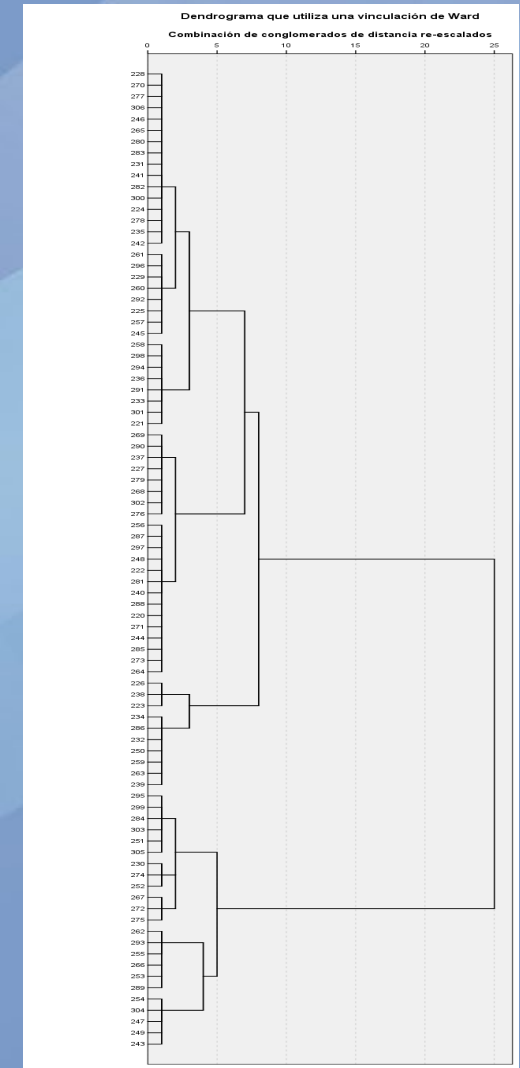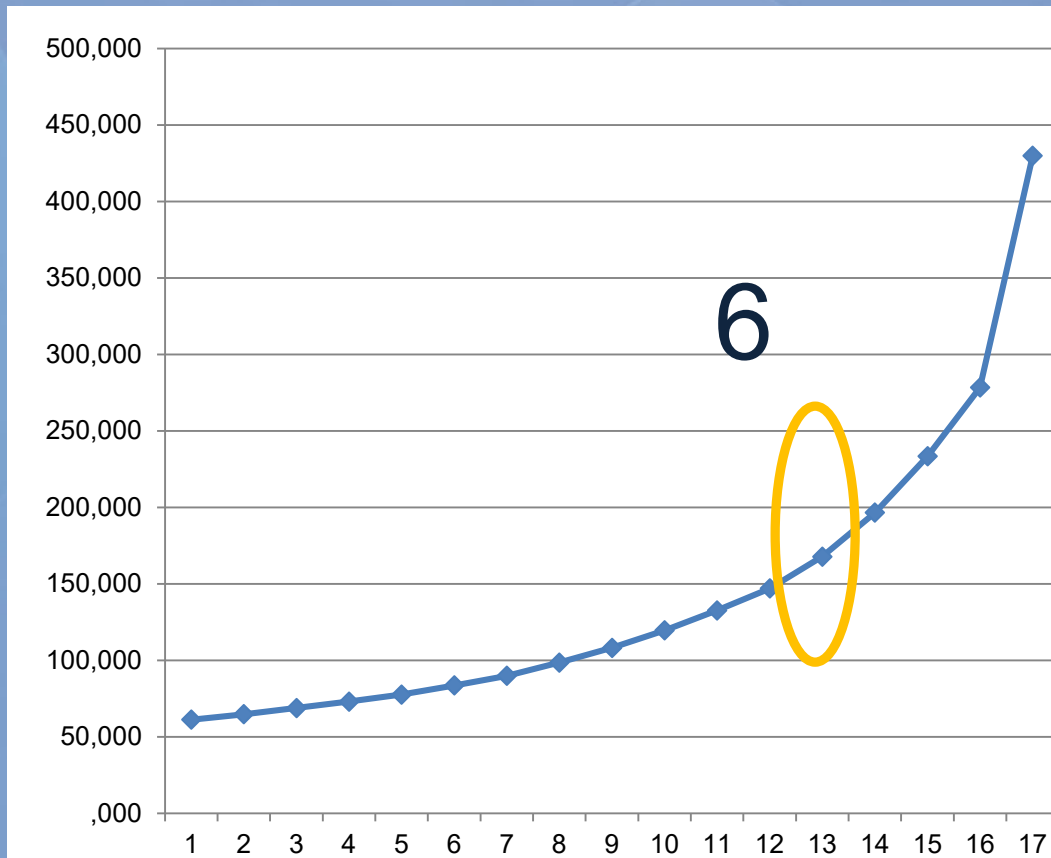- Cluster 3 N=28
- Cluster 4 N=41
- Cluster 5 N=12

- **Decoding skills** needed for average and high achievement

- **Grammatical sensitivity** relevant in high overall aptitude profile and in high achievement

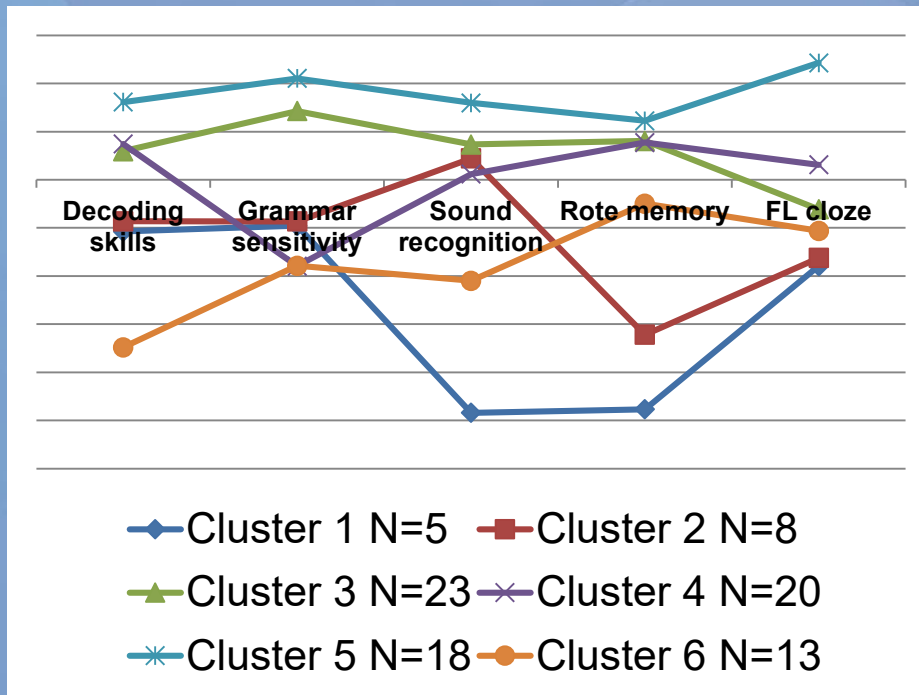- High and average **memory** for average/ high performance

Results RQ1: Change of profiles over time

Grade 5

6

Results RQ 1 & 2: profiles over time + high vs. low achievers

Grade 5

Chart with axis labels: Decoding skills, Grammar sensitivity, Sound recognition, Rote memory, FL cloze

Legend:
- Cluster 1 N=5
- Cluster 2 N=8
- Cluster 3 N=23
- Cluster 4 N=20
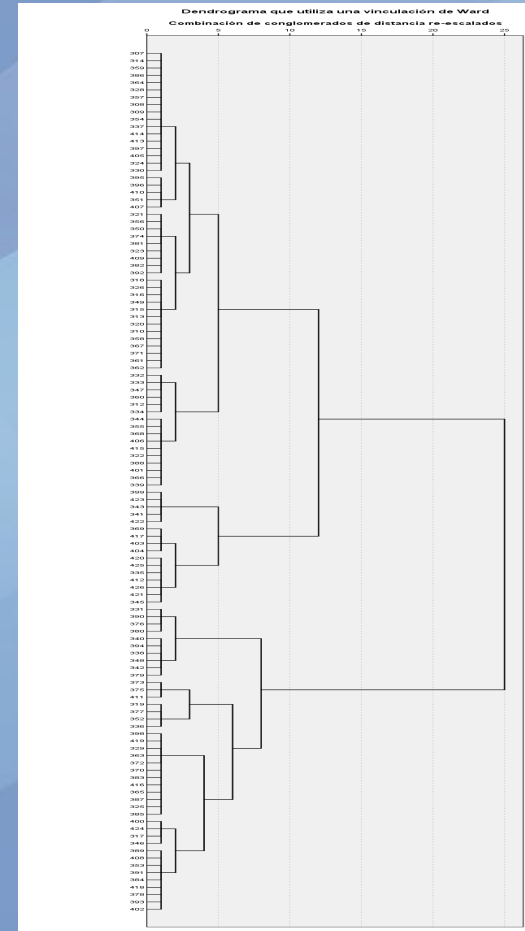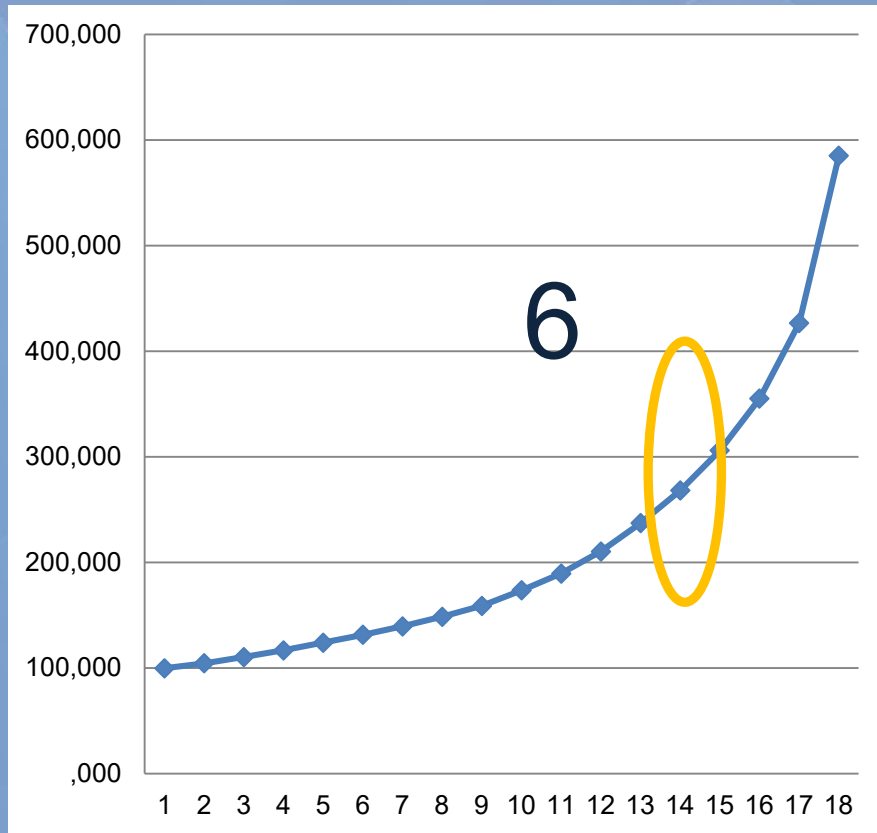- Cluster 5 N=18
- Cluster 6 N=13

- 2 high achiever profiles with high overall aptitude and with high grammar sensitivity

- **Grammar – oriented** vs. **Memory oriented** learner

- Low achievers have poor decoding skills, poor sound recognition and average / poor memory
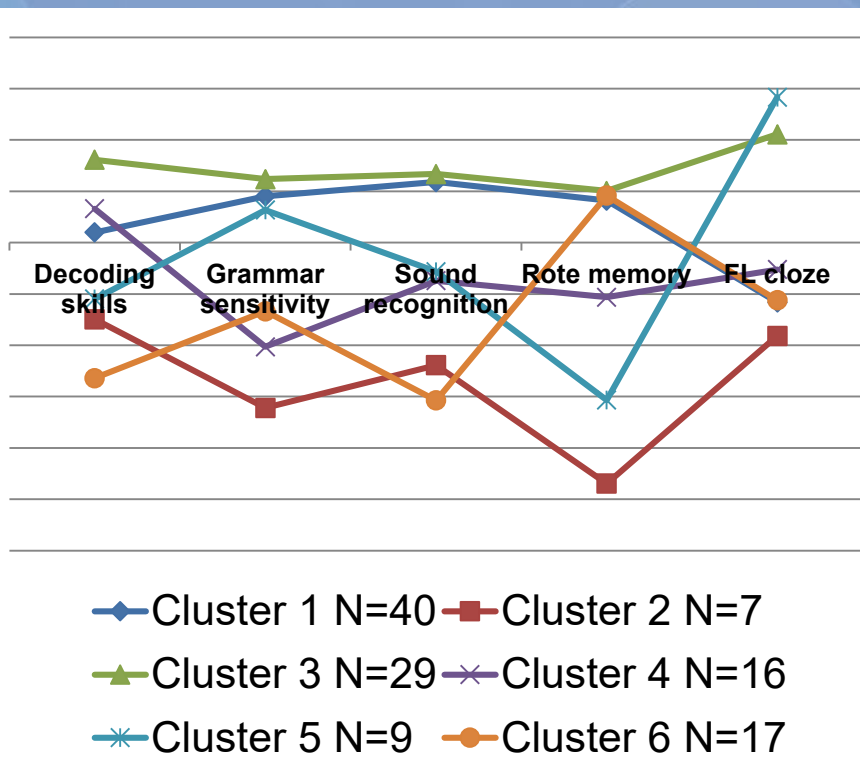
Results RQ1: Change of profiles over time

Grade 6

Results RQ 1 & 2: profiles over time + high vs. low achievers

Grade 6

Chart with axis labels: Decoding skills, Grammar sensitivity, Sound recognition, Rote memory, FL cloze

Legend:
- Cluster 1 N=40
- Cluster 2 N=7
- Cluster 3 N=29
- Cluster 4 N=16
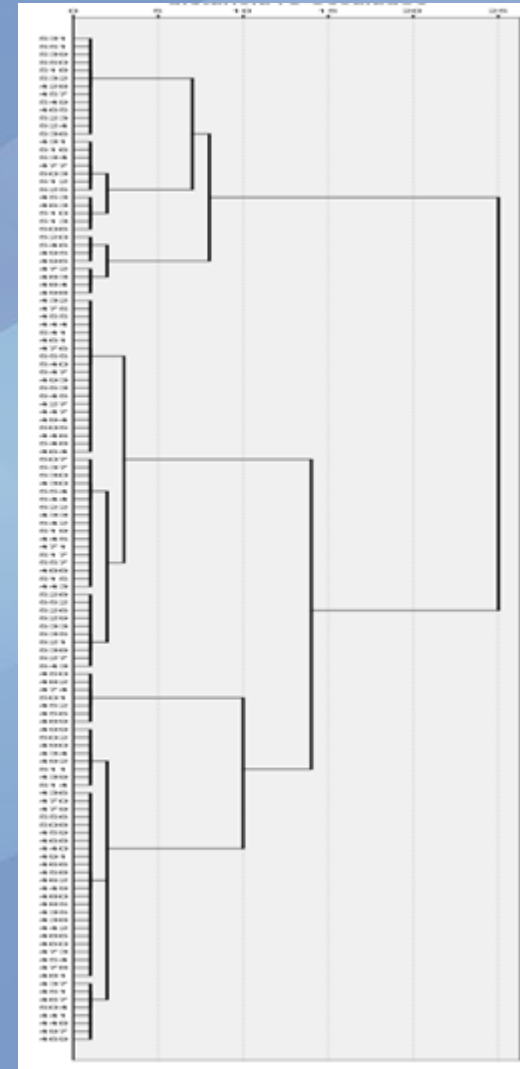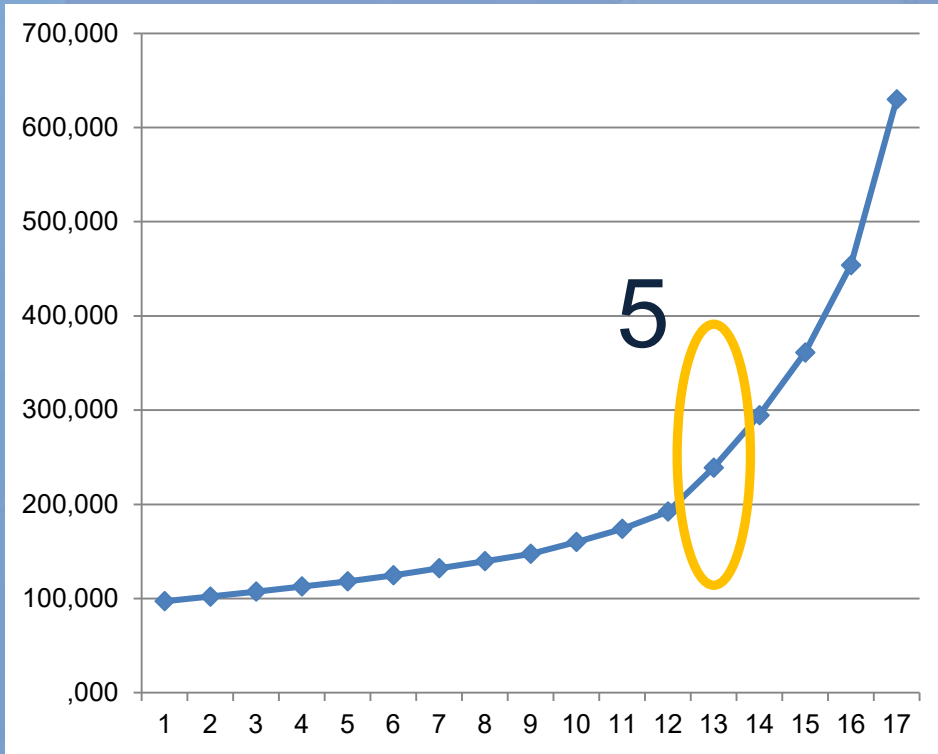- Cluster 5 N=9
- Cluster 6 N=17

-2 high achiever profiles: with high aptitude, or with high grammar sensitivity in spite of lower memory

- **Memory oriented** learners who are average in the rest are average achievers

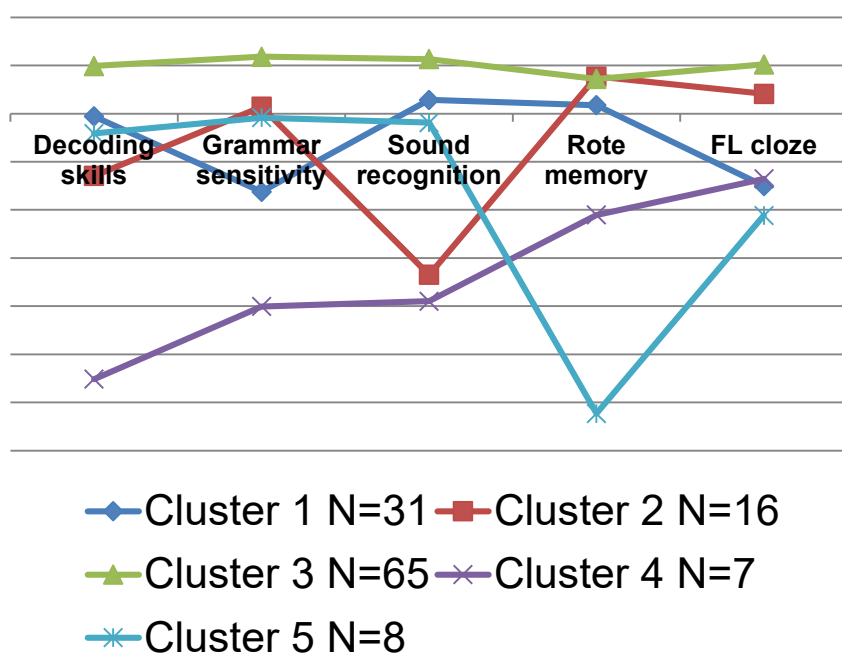- **Low/average achievers have average or poor memory and are poor in the other abilities**

Results RQ1: Change of profiles over time

Grade 7

Results RQ 1 & 2: profiles over time + high vs. low achievers

Grade 7

Chart axis labels: Decoding skills, Grammar sensitivity, Sound recognition, Rote memory, FL cloze

Legend:
- Cluster 1 N=31
- Cluster 2 N=16
- Cluster 3 N=65
- Cluster 4 N=7
- Cluster 5 N=8

-2 high achiever profiles: with high aptitude or with both high grammar sensitivity and high memory but low sound recognition

-<u>Low memory</u>, low achiever

-Average sound recognition and spelling not crucial to have average achievement

ANOVAs F-values

Memory as the dimension to distinguish clusters overall, with literacy skills coming second at lower levels

Grade	3	4	5	6	7
Part 1	**21.067** (2nd)	43.031 (2nd)	39.852 (3rd)	21.468 (4th)	46.721 (3rd)
Part 2	17.641	27.589	**45.896**	25.613	39.749
Part 3	18.898	34.819	32.327	48.318	85.473
Part 4	**21.361**	**54.467**	**43.925**	**52.375**	**99.537**

Discussion

Components & acquisition processes
(Skehan 1998, 2002; Dörnyei & Skehan 2003)

Input
Phonemic coding ability

- **Noticing**

- **Attention**

- **Related to phonological STM**

Central processing
Language analytic abilities

- **Pattern identification**

- **Restructuration of the IL system**

- **Carroll's grammatical sensitivity + inductive language learning ability**

Output
Memory (WM, LTM)

- **Retrieval of info processed**

- **More important than LAA**

- **Salient in talented language learners**

Discussion

RQ1   **Changes in language learner profile over time**

❱ **Yes, regarding decoding skills**
❱ **Yes, no linear high-aptitude profiles in the lower grades**
❱ **BUT memory-oriented vs grammar-oriented learners were high-achievers in the lower grades and memory-oriented were average or low-achievers from grade 4 on, not on grade 3, where decoding skills prevail.**

RQ2 **Profiles in high-achievers vs low-achievers**

❱ **high-aptitude even profiles**
❱ **either high grammar sensitivity or memory-oriented for average and high achievement at higher levels, not in grade 3 & 4**
❱ **low overall aptitude in low achievers**

RQ3 **Dimension to differentiate clusters**

❱ **decoding skills for younger learners → applicable to all alphabets?**

Further research

▶**Similar profiles with different proficiency results**

▶**The role of sound recognition at lower levels (over decoding skills?)**

▶ **Memory as differentiating dimension except in grade 5, followed by decoding skills at lower grades: LCDH in L1 and FL, but in all alphabets?**

▶**Use of different proficiency measures tapping different abilities**

▶**Aptitude profiles using proficiency as a criterion variable**

Thank you