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

Mª del Mar Suárez
mmsuarez@ub.edu
Universitat de Barcelona
<table>
<thead>
<tr>
<th>Literature review</th>
<th>Research questions</th>
<th>Methodology</th>
<th>Results</th>
<th>Discussion</th>
<th>Further research</th>
</tr>
</thead>
</table>

**USING CLUSTER ANALYSIS**
<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature review</td>
</tr>
<tr>
<td>Research questions</td>
</tr>
<tr>
<td>Results</td>
</tr>
<tr>
<td>Discussion</td>
</tr>
<tr>
<td>Further research</td>
</tr>
</tbody>
</table>

**METHODOLOGY**
DISCUSSION
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)
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)
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)

**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 conceptual scheme**
   - Outliers
   - Not normally distributed data
   - Representativeness of the sample
   - Multicollinearity
3. **Generate hypotheses to be tested** → 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

### Multicomponentional
- Test batteries tapping different abilities

### Learner profiles vs. aptitude profiles
- Including or not proficiency measures
Why use cluster analysis in aptitude research?

- Placement & selection
- Guidance
- Learner & aptitude profiles
- Learning disabilities diagnosis
- Strengths and weaknesses
- Aptitude Treatment Interaction
- SLA research (f.ex. formal vs. informal contexts)

Learner & aptitude profiles

- Placement & selection
- Guidance
- Learning disabilities diagnosis
- Strengths and weaknesses
- Aptitude Treatment Interaction
- SLA research (f.ex. formal vs. informal contexts)
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
Aptitude components & proficiency

2

Literature review

Aptitude component

Proficiency

LAA
PhCA
Mem

Aptitude component
## Language aptitude components (Carroll, 1967)

<table>
<thead>
<tr>
<th>MLAT-E (grades 3 to 6)</th>
<th>MLAT (adults)</th>
<th>Construct</th>
</tr>
</thead>
</table>
| 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 |
4a **MLAT-E parts** (Carroll, 1967)

### MLAT-E Part 1 – Hidden words

<table>
<thead>
<tr>
<th></th>
<th>wntr</th>
<th>A champion</th>
<th>B season</th>
<th>C liquid</th>
<th>D happy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>snak</td>
<td>A hard wood</td>
<td>B tease</td>
<td>C reptile</td>
<td>D type of shoe</td>
</tr>
</tbody>
</table>

### 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.
MLAT-E Part 3 – Finding Rhymes

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

2. RAIN …… □ vine …… □ cane …… □ keen …… □ fine
MLAT-E Part 4 – Number Learning

**Units**
- ba = one
- baba = two
- dee = three

**Tens**
- tu = twenty
- ti = thirty

\[ \text{ti} \quad 30 \quad + \quad \text{ba} \quad 1 \quad = \quad \text{thirty-one} \quad 31 \]
<table>
<thead>
<tr>
<th>Proficiency as criterion variable</th>
<th>Proficiency included in the cluster analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Motivation profiles (Dörnyei et al., 2006)</td>
<td>• LAA in learner profiles → Metalinguistic awareness in L1 &amp; L2 and L2 proficiency over time (Ranta, 2002)</td>
</tr>
<tr>
<td>• Vocabulary learner strategies profiles (Kojic-Sabo &amp; Lightbown, 1999)</td>
<td>• Learner differences in strategy use, will to learn and achievement over time (Yamamori et al., 2003)</td>
</tr>
<tr>
<td></td>
<td>• Learner cognitive profiles including MLAT-4 Words in Sentences, age and criterion test score (Skehan, 1986)</td>
</tr>
</tbody>
</table>
Learner profiles as of cluster analysis

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

Low achievers
- Average or weak on LAA \((\text{Ranta, 2002})\)
- Very poor memory except associative memory, average language ability \((\text{Skehan, 1986})\)
- Intelligent but with poor language ability \((\text{Skehan, 1986})\)
- High linguistic ability and memory but average IQ \((\text{Skehan, 1986})\)
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
Do language learner profiles change over time in young learners in grades from 3 to 7?

What learner profiles do younger high achievers and low achievers have?

What determines achievement in younger learners?
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

<table>
<thead>
<tr>
<th>Grade</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age</td>
<td>8.8</td>
<td>9.9</td>
<td>10.9</td>
<td>11.7</td>
<td>12.9</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N MLAT-ES/EC</td>
<td>N MLAT-EC/ES</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>55</td>
</tr>
<tr>
<td>Total N</td>
<td>280</td>
<td>269</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N MLAT-EC/ES</th>
<th>57</th>
<th>58</th>
<th>65</th>
<th>62</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total N</td>
<td>269</td>
<td>280</td>
<td>549</td>
<td></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Parts</th>
<th>MLAT-EC (grades 3 to 7)</th>
<th>MLAT (adults)</th>
<th>Construct</th>
</tr>
</thead>
</table>
| 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 |
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.
Items removed after item analysis

MLAT-EC 13 3 122
## Reliability

<table>
<thead>
<tr>
<th></th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
<th>Grade 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part 1</td>
<td>.934</td>
<td>.900</td>
<td>.871</td>
<td>.878</td>
<td>.873</td>
</tr>
<tr>
<td>Part 2</td>
<td>.856</td>
<td>.897</td>
<td>.903</td>
<td>.910</td>
<td>.915</td>
</tr>
<tr>
<td>Part 3</td>
<td>.934</td>
<td>.939</td>
<td>.909</td>
<td>.919</td>
<td>.913</td>
</tr>
<tr>
<td>Part 4</td>
<td>.936</td>
<td>.909</td>
<td>.922</td>
<td>.893</td>
<td>.941</td>
</tr>
<tr>
<td>Total</td>
<td>.957</td>
<td>.960</td>
<td>.950</td>
<td>.944</td>
<td>.961</td>
</tr>
</tbody>
</table>
## Mean $p$-values

<table>
<thead>
<tr>
<th></th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
<th>Grade 7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Part 1</strong></td>
<td>.75</td>
<td>.81</td>
<td>.83</td>
<td>.87</td>
<td>.89</td>
</tr>
<tr>
<td>Hidden words</td>
<td>easy</td>
<td>very</td>
<td>easy</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Part 2</strong></td>
<td>.32</td>
<td>.50</td>
<td>.65</td>
<td>.67</td>
<td>.74</td>
</tr>
<tr>
<td>Matching words</td>
<td>difficult</td>
<td>mid-difficult</td>
<td>easy</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Part 3</strong></td>
<td>.55</td>
<td>.72</td>
<td>.78</td>
<td>.83</td>
<td>.83</td>
</tr>
<tr>
<td>Rhyming words</td>
<td>easy</td>
<td>very</td>
<td>easy</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Part 4</strong></td>
<td>.62</td>
<td>.82</td>
<td>.88</td>
<td>.85</td>
<td>.91</td>
</tr>
<tr>
<td>Number learning</td>
<td>easy</td>
<td>very</td>
<td>easy</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Grade</th>
<th>Mean Raw Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>gr. 3</td>
<td>55.62</td>
</tr>
<tr>
<td>gr. 4</td>
<td>83.67</td>
</tr>
<tr>
<td>gr. 5</td>
<td>93.17</td>
</tr>
<tr>
<td>gr. 6</td>
<td>94.20</td>
</tr>
<tr>
<td>gr. 7</td>
<td>101.47</td>
</tr>
</tbody>
</table>

Note: The x/122 indicates that the scores are normalized per 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

### Intercorrelation of parts

<table>
<thead>
<tr>
<th></th>
<th>Part 1</th>
<th>Part 2</th>
<th>Part 3</th>
<th>Part 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part 2</td>
<td>.616**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part 3</td>
<td>.694**</td>
<td>.616**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part 4</td>
<td>.513**</td>
<td>.476**</td>
<td>.472**</td>
<td></td>
</tr>
</tbody>
</table>

**Significant <.01 (two-tailed)**

### Standardization of measures

### 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

![Graph showing the change of profiles over time with a focus on grade 3.](image-url)
Results RQ 1 & 2: profiles over time + high vs. low achievers

Grade 3

- **High decoding skills** and average/high sound recognition ability
- **EITHER** high grammatical sensitivity **OR** high memory + decoding skills for high achievement
- Sound recognition alone, no effect
Results RQ1: Change of profiles over time

Grade 4

<table>
<thead>
<tr>
<th>Grade 4</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>600,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>700,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Grade 4

- **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

![Graph showing the change of profiles over time for Grade 5. The graph includes a line graph with data points and a dendrogram. The x-axis represents time from 2000 to 2017, and the y-axis represents the number of students. The graph highlights a significant increase in the student profile starting from year 12.]
Grade 5

- 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

Grade 6
Grade 6

- 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

Grade 7
Grade 7

- 2 high achiever profiles: with high aptitude or with both high grammar sensitivity and high memory but low sound recognition
- Low memory, low achiever
- Average sound recognition and spelling not crucial to have average achievement
Memory as the dimension to distinguish clusters overall, with literacy skills coming second at lower levels

<table>
<thead>
<tr>
<th>Grade</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part 2</td>
<td>17.641</td>
<td>27.589</td>
<td>45.896</td>
<td>25.613</td>
<td>39.749</td>
</tr>
<tr>
<td>Part 3</td>
<td>18.898</td>
<td>34.819</td>
<td>32.327</td>
<td>48.318</td>
<td>85.473</td>
</tr>
<tr>
<td>Part 4</td>
<td>21.361</td>
<td>54.467</td>
<td>43.925</td>
<td>52.375</td>
<td>99.537</td>
</tr>
</tbody>
</table>
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
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