

**A mixed-methods study on verbal abilities and cognitive flexibility of Hungarian learners in CLC and general language programmes**

**Review of the PhD Thesis by Ágnes Sántha-Malomsoki**

The doctoral dissertation describes a study on how cognitive abilities differ in primary school children participating either in CLIL or in general language programmes. The study uses several research methods to compare the performance of CLIL and general language learner students from various perspectives. Results in general argue for better L2 levels as well as greater cognitive flexibility in CLIL students.

The introduction is composed of two halves. The first part provides a nice overview of the current state of Hungarian language education, the second part provides an overview of how bilinguals and monolinguals differ from each other in executive functions. The introduction covers a very broad range of literature, explaining and discussing some of the most important concepts. A shortcoming of the introduction is that the two halves are not clearly integrated. I believe that the Author should draw the conclusions herself, and should not let the reader link the phenomena to each other. A more thorough integration would provide a stronger motivation to the hypotheses articulated on page 61. **Could the Author summarize how general language programme participants and CLIL participants could be integrated to the monolingual-bilingual continuum?**

The methods section is described in detail. The research methods are well-chosen to answer the research question. The description of the tasks is slightly unbalanced. Some tasks are explained in more details, while in the case of others, the reader is only provided tables, which require a lot of effort to interpret. It was interesting to read about the differences in the educational background of the families (pages 65-66). There are certain studies showing that executive functions are affected by socio-economic status (e.g. papers by Pascale Engel de Abreu). **Would the Author expect that controlling for parental education/SES modifies the results of the current study?**

Throughout the description of the LEAP-Q test and its results, the Author describes the question of accents. It is certainly a central phenomenon in English L2 instruction, as well as in English dialectology. At the same time, the topic of accents is also introduced for Hungarian (page 86). While English (as well as German, etc) is a pluricentric language, this might not apply to Hungarian. Also, I believe that it is rather infrequent to meet individuals with Hungarian as their L2. **So my question is how the Author would characterise the nature of accents/dialects in Hungarian?**

Reading the results of the d2-R task, I was struck by the fact that the general learning programme group processed more stimuli, but made more omission errors as well as incorrect markings. That is, maybe the general learning programme children had a criterion to process as many stimuli as possible, whereas the CLIL students' aim was not to make any misses. The use of  $d'$  (signal detection theory) may be suitable to control for such differences in criterion, and would provide a better insight to group-based differences. **Would the Author expect any changes in the pattern of results when criteria are controlled for?**

I had difficulties in interpreting the results of the fluency tests. The Author provided the sum of answers group-wise. I believe that the use of the sum of answers is misleading, due to not completely identical sample sizes. For example, the following statement is made on page 94: "The total number of words generated in the English phonemic fluency tests by the CLIL group is 1681, and 1326 in case of the control group. The CLIL group participants generated 1.27 times as many words as the control group participants. " Since the CLIL group was slightly smaller, the difference should be greater in this case. The paragraph is continued as so: "To reveal whether this number is significantly different Mann-Whitney test was run. It can be concluded from the results of the Mann-Whitney test that there is a significant difference ( $p < .05$ ) between the two groups for all three test conditions (F:  $p = .004$ ; A:  $p = .000$ ; S:  $p = .000$ )" It is not clear what data was analysed here. **Was it group-wise data or the observed mean of individual data? Could the Author clarify data treatment in general?** The question of group-wise versus individual presentation is even more misleading in the case of word classes (Table 32., page 104). Almost all the values are higher for the CLIL group, which is not surprising, since they provided more answers. But the proportion of the different word classes could be compared, and would be more meaningful.

The study incorporates a principle component analysis for both the CLIL and general language programme groups. While the prerequisites of the PCA are described in detail, the acquired principle components are only hinted upon (pages 79-84). **Could the Author detail the results of the PCAs?** E.g. how each variable loads on the principle components.

And a minor, but important suggestion to the end. While the Author has done a great job in applying a variety of the statistical analyses, and describing and interpreting their outcomes, some of the analyses are only referred to with a p-value. I would have loved to read the full description of the statistical analyses including effect sizes. It would have helped me to assume the importance of the individual results.

Graz, 18. 08. 2021



Dr Ferenc Kemény  
Senior Scientist  
Institute of Psychology  
University of Graz