COMPARING THE VOCABULARY OF CHILDREN FROM MARGINALISED ROMA COMMUNITIES AND BILINGUAL AND MONOLINGUAL CHILDREN IN SLOVAKIA BASED ON THE SHORT VERSION OF THE SLOVAK CDI-II

SVETLANA KAPALKOVÁ[1*], DANIELA SLANČOVÁ[2], IVANA VOLÁKOVÁ[3]

[1] Department of Communication Disorders at Faculty of Education, Comenius University

[2] Faculty of Arts, Prešov University of Prešov

[3] SCPP Institute of Child Language in Bratislava

Corresponding author: Svetlana Kapalková, e-mail: kapalkova@fedu.uniba.sk

Received: 31.10.2025
Accepted: 26.01.2026

ORIGINAL RESEARCH ARTICLE
UDK: 81'23:[811.214.58+811.162.4]
81'23:81'246.2
DOI: https://doi.org/10.31299/hrri.62.si.4

Abstract*: The main research question of this pilot study is to show how the vocabulary assessed by the Short TEKOS II (Slovak version of CDI) questionnaire differs children from marginalized Roma communities (CMRC), monolingual Slovak-speaking (MC) and bilingual Slovak-speaking (BC) children during the 17- to 36-month age range. CMCR are children living in extreme poverty in socially excluded communities. Thus, also the Romani translation of Short TEKOS II was used. The results in both, receptive and productive vocabulary, confirm our main assumption that the socio-economic status of the families, the lack of stimuli, together with lower education of mothers significantly affect the children´s vocabulary.*

Keywords*: vocabulary, monolinguals, bilinguals, children from marginalized Roma communities*

INTRODUCTION

The main research question of this pilot study concerns how vocabulary assessed using the Short TEKOS II questionnaire (a shortened version of the Slovak CDI: W&S) can differ in children from marginalised Roma communities (CMRC), as well as monolingual Slovak-speaking children (MC) and bilingual Slovak-speaking children (BC) between the ages of 17 and 36 months.

- As part of meeting the minimum standards for primary paediatrics, a psychomotor development screening (S-PMV questionnaire) is implemented for children up to the age of three years in order to identify developmental risks among children in Slovakia (Matušková et al., 2021). If a general risk is detected, paediatricians also consider a specific risk assessment in the area of autism (using Modified Checklist for Autism in Toddlers (M-CHAT); Hnilicová & Ostatníková, 2018) and developmental language disorders (using the Short TEKOS II)[1].
- Within the population of children in Slovakia, 18% of bilingual children are exposed to languages other than Slovak at an early age, and this number is increasing (Kapalková et al., 2010). Additionally, 12% of the adult Slovak population claims to have a mother tongue other than Slovak (Štatistický úrad Slovenskej republiky, 2021).
- There is also a relatively large group of bilingual children in Slovakia being raised

[1] The Short TEKOS II was adapted by Kapalková & Kaletová (2020) and a reference value was calculated for Slovak-speaking children. The norm study is currently under preparation.

in socially excluded Roma communities, who are affected by generational poverty and, as a result, tend to be marginalised.

The Romani population represents the third largest national minority group among the thirteen groups officially recognised in Slovakia. According to the last census conducted in 2021, more than 150,000 people registered it as either their primary (67,179) or secondary (88,985) nationality (Census, 2021). However, according to the Atlas of Romani Communities (2019), the Slovak Romani population includes 402,840 people, which is around 7.45% of Slovakia's total population. Approximately half of this population is marginalised (Atlas of Romani Communities, 2019).

The socio-cultural position of the Romani minority is quite complicated. Approximately half of the Romani population living in Slovakia were able to adopt to the Slovak way of life and live among the majority. The rest of them live within municipalities but live in clusters with other individuals from their community, particularly on the outskirts. The remaining population (more than 18%) lives in segregated and remote settlements. These communities are characterised by social and ethnic isolation, very high unemployment rates, poverty, and social exclusion.

Considering the linguistic situation, approximately 60% of the Romani population speaks Romani as either their first or second language (Štatistický úrad Slovenskej republiky, 2021). They are considered to be bilingual, mostly in Romani-Slovak and Romani-Hungarian, or tri-lingual (Romani-Slovak Hungarian), using the given languages in their spoken vernacular form. The Romani language was standardised in 2008, mainly based on its East-Slovak variant, and remains under the influence of Slovak and its dialects (Rácová & Samko, 2017).

The research into language acquisition in marginalised Roma communities in Slovakia is quite limited. For example, Kubaník (2012) analysed child-directed speech in Romani. Samko et al. (2021) analysed the context and relationship between first language acquisition by monolingual Slovak-speaking children and Roma-Slovak bilingual children, whose first language consisted of either Slovak or Romani in their first year of schooling: the progress in first language acquisition by Roma-Slovak bilingual children is determined by the type of Roma community in which the child lives, with the lowest success rate achieved by those from marginalised communities. The project, Roma Research on Early Childhood in marginalised Roma communities (Roma REACH), which focuses on exploring the complex mechanisms influencing psychomotor development (including language) in the first three years of a child's life in marginalised Roma communities, is still in its early stages (Chovan et al., 2024)[2]. However, as shown by other studies (Law, Reilly & McKean, 2022), we can assume that the lack of stimuli, low socio-economic status, and the lack of education among mothers can significantly impact the language skills of the population.

METHOD

The main research group in our study was a group of children from the marginalised Roma communities (CMRC) in Slovakia. We compared their receptive and productive vocabulary to two control groups of children with a middle-class upbringing - monolingual children (MC) and bilingual children (BC). The children were matched in age and gender (details in Table 1 and Fig. 1). The sample was categorised into groups based on data from a parental questionnaire mapping the child's personal and family history. Socio-economic status was assigned by parental education, profession, unemployment rate, and place of residence. The CMRC group lived in segregated settlements; their parents had completed a basic education, the mothers did not work, and the fathers were mostly unemployed. A child was characterised as having a bilingual background based on their daily exposure to both languages.

[2] One of the hypotheses presented is that the differences in psychomotor development between children from marginalised Roma communities and the middle-class majority control group increases between the ages of 15- 18 months and 30-36 months (Chovan et al., 2024).

Table 1. *Characteristics of study sample based on age and gender*

Group	N (girls/boys)	Min.	Max.	Average	Median	Standard deviation	Age subgroups (number of children)
CMRC	95 (50/45)	17	36	27.26	27	6.16	17-22 (31) 23-29 (23) 30-36 (41)
MC	95 (50/45)	17	36	27.26	27	6.16	17-22 (31) 23-29 (23) 30-36 (41)
BC	92 (48/44)	17	36	27.03	27	6.08	17-22 (30) 23-29 (24) 30-36 (38)

BC, bilingual children; CMRC, children from marginalised Roma communities; MC, monolingual children
Age is listed in months.

The BC group was homogeneous in terms of the type of simultaneous bilingualism they exhibited. However, it was heterogeneous in terms of language combinations: Slovak + Czech, English, Polish, Hungarian, German, Croatian, French, Serbian, Bulgarian, Spanish, Arabic, Romanian, Romani, Ruthenian, Italian, Latin, Russian, Chinese, and Dutch.

Kyuchukov, de Villiers & Takahesu Tabori (2017) stated the absence of diagnostic tools in the mother tongue of CMRC and recommended the need to conduct assessments of these children in their mother tongue. Therefore, we decided to translate the Slovak version of the Short TEKOS II (both lexical and grammatical part) into Romani. The translation process was organised with the help of mentors, mostly social workers, who collaborated with Romani mothers from the *Omama* programme established by the Civil Organisation *Cesta von* (Way Out)[3]. The lexical and grammatical terms were translated based on shared online tables with translations of the items from the Slovak Short TEKOS II into standard and vernacular Romani by fourteen members of the *Omama* programme who were based in 11 different locations across East and Central Slovakia. Subsequently, the most frequently occurring variants of words were selected as the official Romani translation. Both the Slovak and Romani versions of the TEKOS II questionnaire contain the same 81 vocabulary terms. For this research project, only the lexical part of the Short TEKOS II were used. All the words that the child understood and/or was able to express, regardless of language, were assessed. The Slovak questionnaire was administered to the MC and BC groups in both paper form and online. Parents of the CMRC group could choose the language version that suited them best. In total, 73 questionnaires were completed in Romani and 22 in Slovak. In the CMRC group, the midwives and mentors working in the *Omama* programme asked parents of the children for individual items and filled in their answers in the online questionnaire. According to them, the process of filling out the questionnaire was time-consuming for the parents, who would not have been able to complete the questionnaires without assistance**.**

RESULTS

Based on a statistical comparison using the Kruskal-Wallis test, we found that children's vocabulary differs significantly across the groups, both in terms of receptive ($H(2) = 51.11$, $p < .001$) and productive vocabulary ($H(2) = 12.19$, $p < .002$). Mann-Whitney U tests were used to confirm this finding. Multiple comparisons with Bonferroni corrections ($p = .025$) showed that the CMRC group has a significantly lower level in both receptive (Median = 60 (51 - 68); $U = 1715.5$, $r = - .54$) and productive vocabularies (Median = 42 (15 - 55); $U = 2899.5$, $r = - .31$) compared to the MC group. Similar results were observed in receptive ($U = 2349.5$, $r = - .40$) and

[3] The *Omama* programme brings together Romani mothers who are capable, responsible, hardworking, and well-respected in their communities. Under the mentorship, supervision, and practical guidance of several professionals, they help support small children within the Romani communities.

productive vocabularies (U = 3421, r = - .19) when the CMRC group was compared to the BC group. However, there were no significant differences in the receptive or productive vocabularies ($p > 0.05$) of children from the BC group (Median = 71.50 (60.25 - 80) and those from the MC group (Median = 77 (67 - 80); see Fig. 1).

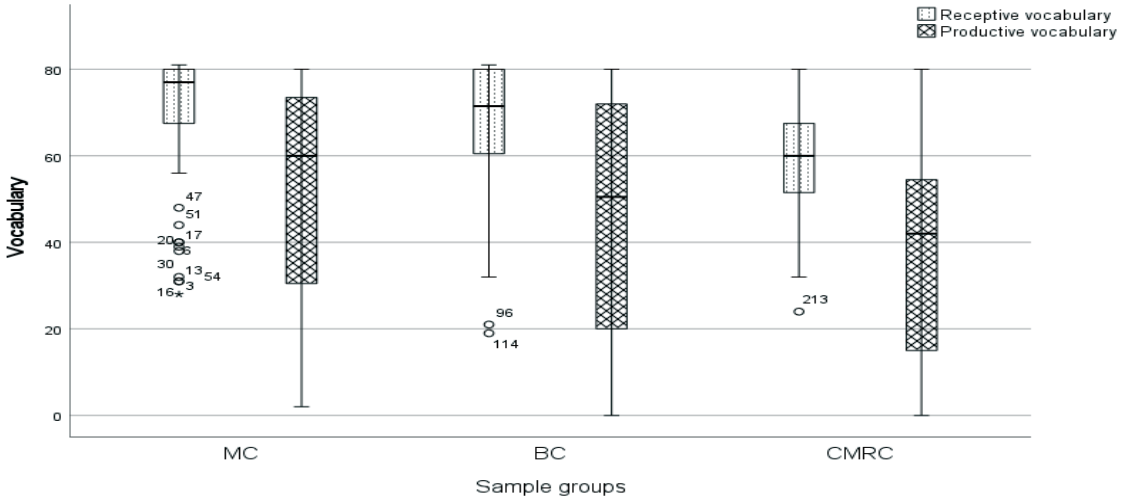

Figure 1. *Comparison of the vocabulary of monolinguals (MC), bilinguals (BC), and children from the marginalised Roma communities (CMRC)*

Subsequently, vocabulary growth from a temporal perspective was examined in more detail. Therefore, we categorised the CMRC target group and both control groups (MC and BC) into three age groups: youngest children (17-22 months), older children (23-29 months), and the oldest ones (30-36 months). Figures 2 and 3 illustrate the dynamics of children's receptive vocabulary and production growth.

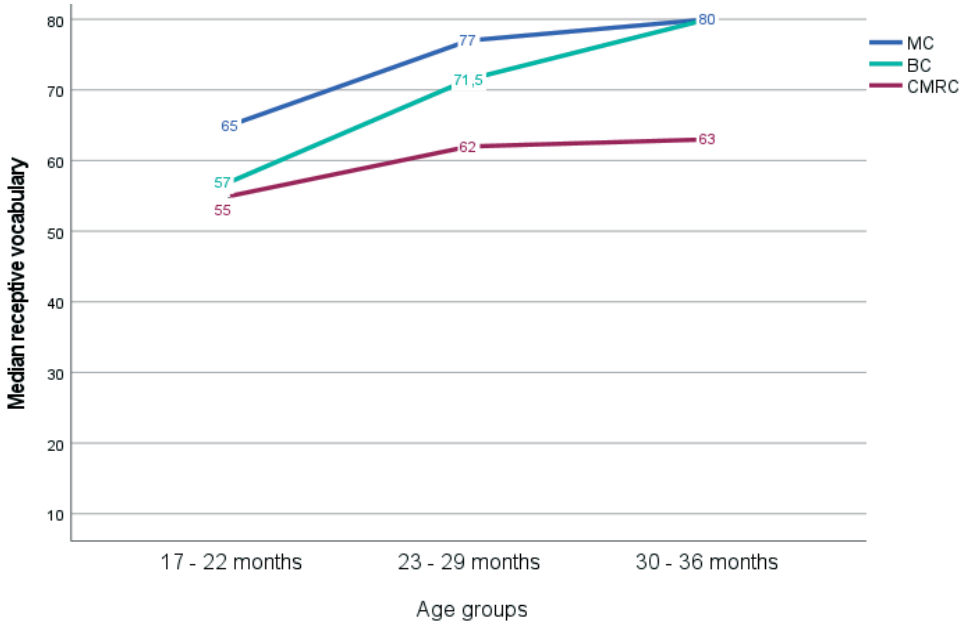

Figure 2. *Receptive vocabulary growth across the study sample (MC, BC, CMRC) from the perspective of the children's age*

In the youngest age group, the receptive vocabulary of CMRC does not differ significantly from BC (U = 418.50, z = - .671, p = ns). However, the receptive vocabulary of the CMRC differs from that of the MC (U = 290, z = 2.68, p = .007). On the other hand, in the oldest age group, we can see the opposite pattern: BC does not differ from MC in receptive vocabulary (U = 1160, z = - 1.829, p = ns), while the receptive vocabulary in the CMRC group stagnates. Furthermore, it shows a significant difference compared to both control groups: for BC (U = 1010, z = - 2.75, p < .01) and for MC (U = 757.5, z = - 4.31, p < .001).

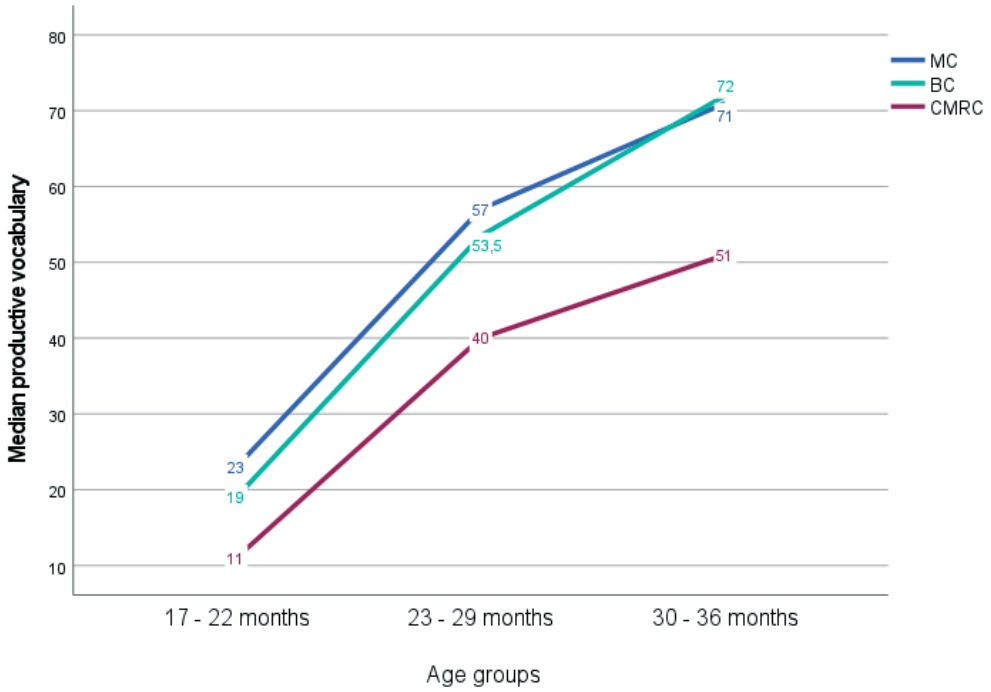

Figure 3. *Production vocabulary growth across the study sample from the perspective of the children's age*

The productive vocabulary of CMRC and BC do not differ in the youngest age group (U = 379.5, z = 1.24, p = ns). However, CMRC and MC differ in the size of the productive vocabulary (U = 335.5, z = 2.04, p < .05). At the end of the observed age range, the pattern is reversed: BC overtake the MC (U = 1322, z = 836, p = ns), and the CMRC stagnate in comparison to the MC (U = 1684, z = - 3.63, p < .001). The gap becomes more pronounces, similar to the pattern observed in receptive vocabulary.

CONCLUSION

Population-level screenings for Slovak children at risk for delayed psychomotor development and autism, as well as developmental language disorders, is a beneficial and promising method that facilitates early interventions in young children. The Pan American Health Organisation states that, despite the lack of reliable evidence on the benefits and disadvantages of screening for speech and language disorders at an early age, several screening tools are being used in clinical practice. However, the procedures used cannot reliably identify only those children who remain at risk even in adulthood, and they often capture children who "seem" to be cured (US Preventive Services Task Force, 2024). Therefore, in practice, we lack the necessary evidence on the predictive validity of such screenings. However, these screenings also raise many questions regarding the tools used and the interpretation of results in the case of a heterogeneous group of children from multilingual and/or marginalised

backgrounds. Our pilot study aims to contribute to this question by examining the factors affecting language development in three groups of typically developing children living in Slovakia (monolingual, bilingual, and children from marginalised Roma communities).

Our findings show that both the bilingual groups – bilingual Slovak children and children from the marginalised Romani communities - do not show statistically significant differences in the onset of language development by 22 months. After the second year, however, children with the same socio-economic status start to diverge in terms of performance. In contrast, children from marginalised communities begin to stagnate in both vocabulary production and comprehension. By the third year, bilingual children have the biggest advantage over monolinguals. In contrast, marginalised bilingual children stagnate significantly, and their profile matches that of children at risk in terms of referential values. Based on a comparison of the results in our sample, we assume that this is not a case of multilingualism. However, the socio-economic status of the family has a more significant impact on vocabulary size. These results are consistent with studies that confirm the relationship between socio-economic factors and children's language abilities in England (Law, Reilly & McKean, 2022).

Since 2019, population-level screening has been provided for Slovak children by paediatricians. The present study offers a preliminary comparison of results for children with different backgrounds relative to the norming study, and it highlights the potential limitations of applying a single set of norms to all Slovak children. Our findings suggest that such an approach may lead to unreliable interpretations, especially in linguistically and socio-culturally diverse populations. Taking these results into consideration, future studies should focus on re-analysing the available data, with particular attention to whether the reference scores appropriately identify children at risk.

Acknowledgment

This work was supported by Grant APVV-20-0126 from the Slovak Research and Development Agency (APVV) and by Grant VEGA 1/0364/26.

REFERENCES

Atlas rómskych komunít (2019). https://www.romovia.vlada.gov.sk/atlas-romskych-komunit/atlas-romskych-komunit-2019/?csrt=14815749339318409230

Chovan, S., Fiľakovská Bobáková, D., Madarasová Gecková, A., Hubková, B., Štrkolcová, G., Reijneveld, S. A., & de Kroon, M. L. A. (2024). Biological and contextual determinants of early development in marginalized Roma communities: A research protocol of the RomaREACH study. *International Journal for Equity in Health*, *23*(1), 200. https://doi.org/10.1186/s12939-024-02287-0

Hnilicová, S., & Ostatníková, D. (2018). Poruchy autistického spektra – včasná diagnostika a skríning. *Pediatria pre prax*, *19*(2), 52–56.

Kapalková, S., & Kaletová, I. (2020). Krátka slovenská verzia testu komunikačného správania. *Logopaedica*, *22*(1-2), 5–13.

Kapalková, S, Slančová, D., Bónová, I., Kesselová, J., & Mikulajová, M. (2010). *Hodnotenie komunikačných schopností detí v ranom veku*. Slovenská asociácia logopédov.

Kubaník, P. (2012). Dětský registr romštiny. *Romano džaniben*, 19(1), 61–80.

Kyuchukov, H., de Villiers, J., & Takahesu Tabori, A. (2017). Why Roma children need language assessments in Romani. *Psychology of Language and Communication, 21*(1), 215–243. https://doi.org/ 10.1515/plc-2017-0011

Law, J., Reilly, S., & McKean, C. (2022). *Language development. Individual Differences in a Social Context*. Cambridge University Press.

Matušková, O., Prokopová, E., Rajkovičová, H., Jurišová, E., Vodičková, B. & Gondec, M. (2021). *Štandard vyšetrení psychomotorického vývinu detí pri 2.– 11. preventívnej prehliadke v primárnej starostlivosti* –1. revízia (2021). https://www.health.gov.sk

Rácová, A., & Samko, M. (2017). On the Vitality and Endangerment of the Romani Language in Slovakia. *Asian and African Studies*, *26*(2), 185–208.

Samko, M., Čerešník, M., & Čerešníková, M. (2021). First Language Acquisition by Roma and Slovak Children. *East European Journal of Psycholinguistics*. https://doi.org/10.29038/eejpl.2021.8.2.sam

Štatistický úrad Slovenskej republiky (2021). *Sčítanie obyvateľov, domov a bytov 2021*. Štatistický úrad Slovenskej republiky. https://www.scitanie.sk/en

US Preventive Services Task Force. (2024). Screening for Speech and Language Delay and Disorders in Children: US Preventive Services Task Force Recommendation Statement. *JAMA*, *331*(4), 329–334. https://doi.org/10.1001/jama.2023.26952