RU:ts

Student Linguistics Journal
Radboud University Nijmegen

Edition 4, June 2024

RU:ts

http://rutsjournal.nl

Colofon

Editor-in-chief

Michelle Suijkerbuijk

Editors

Nivedha Narayanaswamy

Andreas Wölfle

Shravani Patil

Lay-out editor

Dalí Chirino

PR officer

Shravani Patil

Treasurer

Michelle Suijkerbuijk

Special thanks to

Dr. Nelleke Oostdijk

Department of Language and Communication

And, most importantly, our authors and reviewers

Introduction

Dear readers,

In this issue of RU:ts Student Linguistics Journal, you will find excellent work by Radboud master's students about their research projects in all fields concerning language and linguistics. Most of the papers in this issue cover a topic in the field of bilingualism: there are papers investigating the effect of language distance, of having older siblings, or of the gender of the parents and their input at home, and there is a paper looking at code-switching patterns in Yakut-Russian speakers. Two other papers are about completely different topics, namely the usage of verbal fluency tests in the research of Parkinson's disease and the attitudes of teachers and students towards English as a medium of instruction in education. This issue thus has something for everyone!

The papers are available online for free, as they are published under Open Access Creative Commons Licence: the authors receive the rights to their papers, and can (re)publish them anywhere as long as they mention RU:ts in some way.

Before starting on this fourth edition, I first had to put together an entirely new team. Luckily, a very talented and hard-working group of students were interested in helping make this fourth edition a success. In the process, we learned each other's strengths and how we could work together, we said goodbye to some of the team members, but also welcomed new team members. We all worked very hard to present the fourth edition of RU:ts and are very proud of the results. I want to thank the whole RU:ts team for all of their hard work.

We also want to give special thanks to the people who helped us during the process of publishing the fourth edition, especially dr. Nelleke Oostdijk and the department of Language and Communication. Finally, we would also like to thank all the reviewers and authors who made this fourth edition possible!

Best wishes,
Michelle Suijkerbuijk

June, 2024

Page index

The Usage of Verbal Fluency Tests in the Research of Parkinson's Disease and the Potential Implications for Diagnosis: Systematic Literature Review

Literature review

Elizaveta Vilenchik

Research Master *Linguistics and Communication Sciences*, Radboud University Nijmegen

Abstract

Parkinson's disease (PD) is a neurodegenerative disorder that primarily impacts motor system. However, it also influences language system, and verbal fluency tests can greatly help in the research of PD in this field. 17 papers were chosen with the primary eligibility criteria being the presence of verbal fluency testing in PD population and potential implications for diagnosis. Several conclusions were drawn. (1) Verbal fluency testing has been used for a long time and continues to be of importance. (2) When it comes to the procedure, most verbal fluency tests are used similarly. (3) Verbal fluency tests have downsides (e.g., difficulties with differentiating language testing and executive function). (4) The points of possible improvement are standardizing the categories in the tests and exploring less common types of tests (e.g., action fluency testing) further.

Keywords: Parkinson's disease, verbal fluency, diagnosis, literature review

1. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder of the central nervous system, with its main symptoms being the ones related to the motor system. According to DeMaagd & Philip (2015), the main pathological sign is the death of dopaminergic neurons in the part of the brain called basal ganglia. Most prominent signs of PD include tremor, rigidity and slowness of movement. Currently, there is no definite cure for PD, and treatment involves mitigating symptoms. Moreover, anxiety and depression are prevalent in persons with PD. Other issues are also frequent, for example, problems with one's senses and sleep. These changes may appear even before the onset of motor decline (DeMaagd & Philip, 2015).Treatment for PD may include using drugs like levodopa at first. Levodopa helps with increasing dopamine concentrations in the brain as it is able to cross the blood-brain barrier which is resistant to dopamine itself. As PD progresses, medication can become less effective, so subthalamic nucleus deep brain stimulation (STN-DBS) can be considered. DBS is a neurosurgical procedure which involves microelectrodes placed under the skull to send signals to specific brain areas. DBS can reduce the symptoms of PD and facilitate the decrease in medications such as levodopa (Beitz, 2014).

PD is usually diagnosed via the assessment of motor functions, even though according to some scholars, speech of persons with PD can also be used in diagnosis as it has features that are not present in the speech of people without PD (Hireš et al., 2022). In general, the speech in PD is characterized by slurring of words, stuttering and mumbling. Persons with PD may experience issues with expressing emotions and having the right pitch and volume of the voice. Dysphagia (problems with swallowing) and aspiration are also the issues that persons with PD struggle with. Speech therapy can help with these problems (Parkinson's Foundation, n.d.). There are pieces of advice that can help – for example, using short phrases, communicating in non-noisy spaces, and planning the periods of vocal rest. Unfortunately, as with the treatment of PD overall, there is no immediate cure or definite solution to speech problems (Parkinson's Foundation, n.d.). Therefore, research is being conducted to counteract speech problems in PD.

In addition, both motor-speech and language-cognition characteristics can alter throughout the course of the disease. The latter is often studied with the help of verbal fluency tests as deficits in verbal fluency are considered existent in PD. Verbal fluency in PD may be impacted by many factors, for example, the presence of dementia, the age of the patient, the number of years since the onset of PD, etc. (Smith & Caplan, 2018). Verbal fluency testing has been described extensively in the literature, but it is still under development. The verbal fluency tests that are most commonly used are the phonemic (or letter) fluency and semantic fluency tests. Phonemic fluency involves participants naming as many words as they can that start with a specific letter within a prespecified amount of time. This type of fluency is thought to rely more on the frontal areas of the brain. Semantic fluency involves naming words from the same semantic category. Unlike phonemic fluency, semantic fluency uses neural networks in the temporal regions (Henry & Crawford, 2004).

Other fluency tests include alternating verbal fluency, alternating phonemic fluency and alternating semantic fluency. The first one is a test which consists of switching between a semantic category and a letter (e.g., *animals* and *T*). Alternating phonemic fluency involves switching between two letters during the naming process (e.g., *B* and *F*). Finally, alternating semantic fluency is a task where a participant switches between two semantic categories (e.g., *animals* and *professions*). Both the ability to switch from one category to another one effectively and the ability to generate items in the same category (so-called cluster size) are areas of increasing interest in the research of speech in PD. These two abilities in patients with PD may have differences when compared to the typical population (Smith & Caplan, 2018). Additionally, it may be the case that there are also differences in patients on or off STN-DBS and with or without dementia.

In summary, verbal fluency testing has been explored in the literature, and various directions have been taken. The aim of this literature review is to find the tests that are the most relevant for possible implications in diagnosis.

2. Research Questions

Four research questions are presented that this systematic literature review strives to provide answers to.

> ***RQ 1****: What is the current state of using verbal fluency tests in patients with Parkinson's disease?*
>
> ***RQ 2****: How are different types of verbal fluency tests used?*
>
> ***RQ 3****: What are the main conclusions regarding Parkinson's disease that are made thanks to verbal fluency tests?*
>
> ***RQ 4****: What can be improved in testing of verbal fluency?*

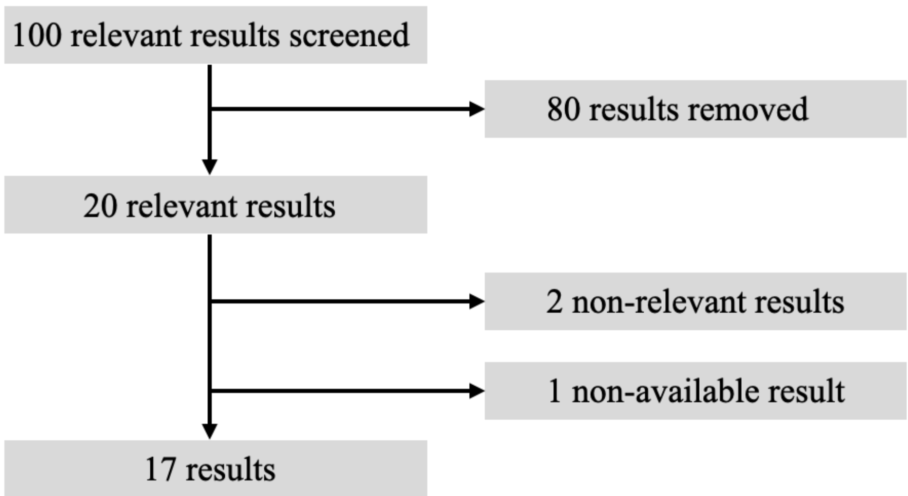
3. Methods

3.1 Search strategy and eligibility criteria

The main stage of searching for the papers involved conducting a search with the help of *Web of Science* (Web of Science, n.d.) which is a research tool to look for relevant research literature. The key concepts in searching were (1) Parkinson's disease (PD), (2) fluency, and (3) language or speech. The following two inquiries were used: (1) *ALL=(("parkinson's disease" OR pd) AND fluency AND (language OR speech))* which produced 321 results, and (2) *ALL=(("parkinson's disease" OR pd) AND "verbal fluency" AND (language OR speech))* which produced 186 results. The second search query appeared more convenient as it was more refined and produced a smaller number of results, however, there was a chance that some papers could have been missed with this search. Therefore, the first inquiry was chosen as the main one. The main eligibility criteria are described below (Table 1). In the end, 17 papers were chosen. They cover the field of verbal fluency testing with the implications for the diagnosis well and can lead to significant conclusions. Out of the first 100 results of the search *ALL=(("parkinson's disease" OR pd) AND fluency AND (language OR speech))* which produced 321 results, the reasons for the papers not meeting the criteria were that they: (1) were not concerned with verbal fluency at all (n=36), (2) were not considered to have major possible implications for diagnosis (n=24), (3) described fluency of speech or other types of fluency (e.g., reading fluency) without testing (usually with analysing recordings and speech samples) (n=19), (4) described PD as in phonological disorder instead of Parkinson's disease (n=1). The list of the papers can be found in the Supplementary Materials.[1]

Figure 1

Summary of the Search Process to Select the Studies (Related to the First 100 Results Screened)



[1] Supplementary materials are accessible at https://www.rutsjournal.nl.

Table 1
Inclusion and Exclusion Criteria

	Inclusion Criteria	Exclusion Criteria
Aim	Testing verbal fluency in PD with possible implications for diagnosis	Other aims
Methods	Using verbal fluency tests	Testing fluency but not using verbal fluency tests (e.g., automatic speech, fluency of speech, etc.)
Publication Type	Empirical research papers	Book chapters, reviews, manuals, etc.
Publication Language	English	Any other languages

4. Results

4.1 Overview of Main Themes

The main condition for a paper to be eligible for this systematic literature review was to have the contents that revolve around verbal fluency tasks, hence the main theme in these papers is studying verbal fluency in the population of persons with PD. However, there are more narrow themes discussed in these papers that are relevant for this literature review.

Several papers focused on cognitive deficits and different stages of PD. These include not only dementia (Azuma et al., 1997; Koerts et al., 2013; Piatt et al.,1999; Signorini & Volpato, 2006; Suhr & Jones, 1998) but also mild cognitive impairment (MCI) (Hamada et al., 2021; Yang et al., 2022) which can indicate a future decline to dementia. Verbal fluency is important when studying cognitive decline as the results of the tests might point at the stages of deterioration (e.g., Hamada et al., 2021).

Other papers concerned the relationship between PD and different aspects that are relative for patients with PD. Considering executive function and its connection to verbal fluency is a prominent direction of research when it comes to several papers (Barbosa et al., 2017; Hedman et al., 2022). Differentiating the deficits of executive function from the deficits of language can be difficult, and this is why research in this direction is of significance. Medication (Herrera et al., 2012) and DBS (Romann et al., 2017) are other factors potentially influencing verbal fluency in PD. Persons with PD on and off dopamine medication have different outcomes in verbal fluency tests, therefore, describing these differences is essential. As for STN-DBS, there have been multiple studies on its impact on patients' wellbeing and the rate of cognitive deterioration. However, the relationship between DBS and verbal fluency still requires a lot of research (Beitz, 2014).

A few papers were primarily concerned with action fluency (Bayram & Akbostanci, 2018; Rodrigues et al., 2015; Signorini & Volpato, 2006). The tendency to look closely at testing verbs can also be observed in other fields of studying language in clinical populations. For instance, action naming has become more common in preoperative and intraoperative language mapping during awake brain surgery in order to diminish the possible aphasia outcomes (Rofes & Miceli, 2014). More attention is paid to verbs and action language when it comes to studying clinical populations, and understandably, persons with PD should also be looked at closely.

In several papers, neuroscientific methods were used to study verbal fluency. Functional imaging helps to learn more about verbal fluency abilities of patients with PD on a neuroimaging level (Ellfolk et al., 2014; Hamada et al., 2021). This can be a major strand for future research as neuroscientific methods can uncover the areas of brain involved in different processes relating to verbal fluency.

4.2 Employed methods

In this part of the literature review, the methods used in the chosen papers are presented.

4.2.1 Types of tests

The tests that were used in the papers selected for this systematic literature review most often were semantic and phonemic (letter) fluency tests. Semantic fluency testing was present in 16 out of 17 papers, and phonemic fluency was in 14 out of 17 papers. When semantic fluency is tested, the most frequently used semantic category is *animals*. It is considered standard and has many benefits, for instance, examining clustering and switching between subcategories can be achieved quite easily for this category. When phonemic fluency is tested, the most frequently used letters are *F*, *A* and *S* which are also considered standard for English. In some papers, only one semantic category or letter was tested but often two or more groups in both fluency tests are used.

There are verbal fluency tests that are less common than phonemic and semantic fluency tests but are employed nonetheless. Name fluency is a rather uncommon task which involves generating names starting with a specific letter. It is a task with an additional constraint that is used in the study by Azuma (1997). In an initial-letter-specific name fluency task the participants produce names beginning with a letter *M* (e.g. Mila, Mary, etc.). This task is used to understand how PD patients perform in tasks that are more effortful than semantic or phonemic fluency tests.

Action fluency is a test that is gaining popularity in the 2020s. The task for the participants is to name as many verbs as they can. The verbs should be infinitives and the same word forms should not be named (e.g., *go, gone, going*). It is interesting that action naming in verbal fluency testing is becoming more common now, as there is also a tendency to use it more often in language testing during awake brain surgery (Rofes & Miceli, 2014).

Regarding semantic fluency, in the experiment by Azuma (1997), the tasks were divided into sets, and the semantic categories were *animals* and *fruits* (Set A), and *vegetables* and *colours* (Set B). Before the main part of the task participants were given a category to practice with. For the practice session for semantic category, *transportation* was chosen. Both sets were given to all participants. In the other studies, when one category was tested, it was usually the *animals* category (Barbosa et al., 2017; Bayram & Akbostanci, 2018). When there were multiple categories in an experiment, there was a wider range of categories, for example, categories like *fish, flowers, weapons, fruits, birds, clothing, insects, sports* (Gurd, 2000). It can be noticed that the *animals* category is not used here but it is rather separated into subcategories (*fish, birds, insects*). In one of the papers, there were three unusual semantic fluency tasks present. They had two conditions, for example, the participants had to name things that are both square-shaped and hard. This was a practice task, after which feedback on the performance was provided by the examinator. Following this, the participants had to name things that are round and flat, and then things that are long and sharp (Hedman et al., 2022). It is also important not to confuse *name fluency* with the *boys' names* category which is a semantic category also used in the studies (Yang et al., 2022; Zec et al., 1999). Suhr and Jones (1998) used double categories, for instance, *fruits/vegetables* and *tools/kitchen utensils*.

As for the phonemic fluency, in the experiment by Azuma (1997), where the tasks were divided into sets, the letters for the phonemic fluency were *A* and *S* (Set A), and *F* and *D* (Set B). The letter for the practice category was *T*. In general, the choice of the letters is dependent on the language. For instance, in a study on the Turkish language, the English letters for F, A and S, correspond to the Turkish letters K, A and

S, respectively. As a result, a slight modification was done (Bayram & Akbostanci, 2018). In Japanese, the letter system is completely different, therefore the participants had a task to generate as many words as possible, beginning with the Japanese letter that is pronounced as /ka/ (Hamada et al., 2021). In Dutch, the letters used were *D*, *A* and *T* (Koerts et al., 2013). In Portuguese, the letters were *P*, *M* and *R*. They were selected instead of more popular *F*, *A* and *S* as they were a better fit for Portuguese based on their orthographic transparency and the number of words in a Portuguese dictionary beginning with these letters (Rodrigues et al., 2015). However, in a study on Brazilian Portuguese by Romann et al. (2017), the letters were of the standard FAS version. Overall, the *F*, *A* and *S* letters are used most often (e.g., Signorini & Volpato, 2006; Yang et al., 2022; Zec et al., 1999). Another version that is also used is the CFL test which is found to be more difficult (e.g., Suhr & Jones, 1998).

As for the other types of verbal fluency tests, in the alternating semantic fluency tasks, *animals* and *furniture* were chosen as the categories by Ellfolk et al. (2014). *Fruits* and *furniture* were agreed by Yang et al. (2022), while Zec et al. (1999) picked *colours* and *occupations*, and *animals* and *states*. For alternating phonemic fluency, letters *C* and *P* were chosen (Zec et al., 1999). In the name experiment by Azuma (1997), names that participants had to say needed to start with the letters *M* and *J* (Set A), and *L* and *P* (Set B). The letter for the practice category was *B*.

4.2.2 Carrying out testing

The standard time of testing is one minute. However, it is not the case in every study. In older papers, the time could be longer, for instance, five minutes for the phonemic fluency task and four minutes for the semantic fluency task (Flowers et al., 1995). However, in their study, Hedman et al. (2022) opted for the longer time for the tasks as well. The semantic fluency task had two constraints, and since it was a more difficult task than a regular semantic fluency task, two minutes were allotted.

The instructions for participants need to be clear in order to be understandable. This is how Flowers et al. (1995, p. 35) describe the instructions given in the experiment: “Tell me as many words as you can think of beginning with the letter S; keep going until I tell you to stop”. In the next task (for semantic fluency), it was indicated to the participants that they no longer had to rely on the letter S and only the semantic category of words was relevant.

When giving scores to the participants, different strategies were used in different studies. Usually, one point was given for each correct example of a word. The points were not given when, for instance, in the semantic fluency task, participants named general category exemplars (like *birds* for the *animals* category, even though the *birds* group has many other exemplars inside it), or when they named words that began with the same phoneme but a different letter (*cement* for a condition of letter *S*) in the phonemic fluency task. When it came to the name fluency, a name reference book and the researchers' common knowledge were taken into account (Azuma et al., 1997). Repeated words were counted only once, and derived words were not counted (Barbosa et al., 2017). Additionally, in some studies some metrics were calculated. The percentage of intrusion (words that did not suffice the condition of the task) and repetition (any repeated word that was already given as a response before) errors were calculated in an experiment by Suhr and Jones (1998). Multiple measures were calculated by Koerts et al. (2013, p. 404–405): (1) the number of correct exemplars for a category, (2) the number of clusters (for semantic fluency – groups of at least two words in the same subcategory; for phonemic fluency – words beginning with the same two letters, rhymes, etc.), (3) the size of the largest observed cluster, (4) the number of extra-dimensional shifts (that is the shifts between different clusters, like from *pets* to *insects*), (5) the number of intra-dimensional shifts (that is the shifts within one cluster, like *European foreign animals* and *Asian foreign animals*).

In most papers, it was described or presumed that testing was conducted in a quiet and calm setting with the researchers both writing down and recording the responses given. All errors and repetitions were written down as well.

5. Discussion

In this systematic literature review, the papers that focused on verbal fluency as a method and that could give a new outlook on applying verbal fluency tests for diagnosis were summarized. The review provides a brief summary of the previous research on the topic and clarifies how it can be used in the direction of diagnosis. Several research questions were stated in the introduction of the paper that this review aims to answer, and they are going to be addressed in this section.

Verbal fluency tests are commonly used in research on PD but are not prevalent enough when it comes to diagnosis, even though their diagnostic value is quite high, particularly when it comes to cognitive deterioration in PD (e.g., Hamada et al., 2021). Areas that have been developing and are continuing to develop are the ones that are associated with action fluency (Piatt et al., 1999; Signorini & Volpato, 2006; Rodrigues et al., 2015), subcategorization as a strategy in verbal fluency tasks (e.g., Azuma et al., 1997), dementia (Piatt et al., 1999; Signorini & Volpato, 2006), and MCI (Hamada et al., 2021; Yang et al., 2022).

The first research question concerned the current state of using verbal fluency tests in patients with Parkinson's disease. The systematic literature review revealed that even though these tests have been around for quite a long time, they still have a lot of potential. The first paper included in this review was published in 1995 (Flowers et al., 1995), and the most recent papers are dated 2022 (Hedman et al., 2022; Yang et al., 2022) which indicates that fluency tests are continuing to stay relevant. Phonemic and semantic fluency tests are used most often, but other tests like the alternating verbal fluency and action fluency tests are also used in research and can be used in diagnosis.

As for the second question, different types of verbal fluency tests are used in a similar manner. The examinators present them one by one, with similar instructions and usually the same time is given to the participants. In the majority of the papers described in this review, the time for verbal fluency tasks is one minute. However, sometimes it can be extended, for instance, when a task is not an ordinary semantic fluency task but a semantic fluency task with two conditions-constraints (Hedman et al., 2022). Such a task is more difficult and therefore requires more time. The instructions are not described in all papers, but typically the task is stated, and in the case of a phonemic fluency task preceding a semantic fluency task, an additional explanation can be added (as explained in the Testing section of the review). The instructions need to be specified for the alternating verbal fluency and name fluency tasks as there are additional constraints for the items that can be said by the participants. The instructions for the action fluency task are quite vague compared to the ones for the other tasks (e.g., infinitive form of verbs, ʻʻthings you can doʼʼ – Herrera et al., 2012, p. 3637). Sometimes, even an example was asked from a participant before the time started and scoring began. For example, the question after the instructions could be: "…Can you give me an example of something that people do?" – then an acceptable response had to be generated by a participant to continue with an experiment, and then an examinator clearly stated that this is the right idea for this task and asked to name as many items like this as possible in one minute (Rodrigues et al., 2015, p. 521).

Thirdly, verbal fluency testing is a convenient method in PD. The downside is that it can be unclear what exactly is tested with verbal fluency tests. It can be argued that only executive functions are tested but it can also be claimed that there are more factors influencing performance at verbal fluency tasks besides executive function (e.g., Rodrigues et al., 2015).

There are several main conclusions regarding PD that can be drawn from the literature on verbal fluency tests. The first one is that the relationships of verbal fluency tasks with a variety of different phenomena can be studied. Dementia has attracted a lot of scientific interest both in the past and today, but nowadays other stages of cognitive decline such as MCI are being explored. The second conclusion is that persons with PD tend to experience more difficulties with verbal fluency tests. Some of these difficulties can be attributed to bradykinesia, however, others are more than just motor impairments slowing down a patient. The third conclusion is that verbal fluency tests are important for neuroimaging studies. rs-fMRI analysis showed that there is significantly reduced connectivity in the MCI group in many areas compared to the control group. What is particularly interesting is that some connectivity values were observed to correlate with the number of switches in semantic fluency task (Hamada et al., 2021). Impaired verbal fluency, in particular in semantic switching, is associated with the increased activity in the right angular gyrus in PD-MCI. A follow-up analysis for imaging results in the study showed that the right angular gyrus was revealed to be upregulated in PD-MCI patients compared to the other groups. It can be stated that the more correct responses are given in each task, the less activity can be found in the right angular gyrus (Yang et al., 2022).

Finally, based on my conclusions derived from the papers, there are multiple things that could be done to improve verbal fluency testing. It would be beneficial to have standard semantic categories for semantic fluency tasks. For the phonemic fluency task, there are two versions that are used most commonly, i.e., the FAS and the CFL versions. They can get replaced in languages other than English based on the phonology of a different language, but it is beneficial to have a version to refer to in need. Using the same letters for testing can also help when the researchers would want to compare results with other studies (however, this would only work for comparison on the material of one language). This way the impact of using different letters does not have to be considered as an influencing factor. For the semantic fluency task, the *animals* category is often used, however, in the papers there are multiple other options, for example, *boys' names, professions,* etc. The *animals* category is standard, but it would be favorable to have other standard semantic categories for the usage in the semantic fluency task. *Animals* are a good fit as clustering and switching between clusters can be analyzed quite easily. A comparable category is *professions*, as used in the study by Rodrigues et al. (2015). The researchers defined nine possible clusters (healthcare, agriculture, etc.), and the analysis of these semantic categories was also rather easy. On the contrary, when it comes to *boys' names* and *states* as used in the experiment by Zec et al. (1999), studying clustering and switching would be difficult with such categories, and it was already determined in this literature review that these performance characteristics in the semantic fluency task should be explored extensively.

Furthermore, more attention needs to be paid to other types of verbal fluency tests besides phonemic and semantic fluency tests. Action fluency is being studied as was shown in this review, but semantic fluency with additional constraints and alternating verbal fluency need to be examined further. Tasks like these are more demanding and require more effort and are on the other side of the spectrum from the action fluency task (that does not have any constraints), so it might be appropriate to explore them further.

6. Conclusion

When talking about the current stage of the art of verbal fluency tests in PD, it can be pointed out that verbal fluency testing still has a lot of potential which is why it continues to be in use. Tests are generally used in a similar manner with the same number of minutes and similar instructions given to participants. Thanks to these tests important conclusions about PD are made, for instance, the ones related to the difficulties that persons with PD have when it comes to such tests and how it is connected to other disorders such as dementia. As for future improvement, standardization of existing tests and exploration of more complex and difficult versions should be done. In summary, there are limitations in the research to date that can be changed in the future direction of the field.

References

Azuma, T., Bayles, K. A., Cruz, R. F., Tomoeda, C. K., Wood, J. A., McGeagh, A., & Montgomery, E. B., Jr. (1997). Comparing the difficulty of letter, semantic, and name fluency tasks for normal elderly and patients with Parkinson's disease. *Neuropsychology, 11*(4), 488–497. https://doi.org/10.1037/0894-4105.11.4.488

Barbosa, A. F., Voos, M. C., Chen, J., Francato, D. C. V., Souza, C. O., Barbosa, E. R., Chien, H. F., & Mansur, L. L. (2017). Cognitive or Cognitive-Motor Executive Function Tasks? Evaluating Verbal Fluency Measures in People with Parkinson's Disease. *BioMed research international, 2017*, 7893975. https://doi.org/10.1155/2017/7893975

Bayram, E., & Akbostanci, M. C. (2018). Verb naming fluency in hypokinetic and hyperkinetic movement disorders. *Cortex; a journal devoted to the study of the nervous system and behavior, 100*, 21–31. https://doi.org/10.1016/j.cortex.2017.06.014

Beitz, J. M. (2014). Parkinson's disease: a review. *Frontiers in bioscience (Scholar edition)*, *6*(1), 65–74. https://doi.org/10.2741/s415

DeMaagd, G., & Philip, A. (2015). Parkinson's Disease and Its Management: Part 1: Disease Entity, Risk Factors, Pathophysiology, Clinical Presentation, and Diagnosis. *P & T : a peer-reviewed journal for formulary management*, *40*(8), 504–532.

Ellfolk, U., Joutsa, J., Rinne, J. O., Parkkola, R., Jokinen, P., & Karrasch, M. (2014). Striatal volume is related to phonemic verbal fluency but not to semantic or alternating verbal fluency in early Parkinson's disease. *Journal of neural transmission (Vienna, Austria : 1996), 121*(1), 33–40. https://doi.org/10.1007/s00702-013-1073-2

Flowers, K. A., Robertson, C., & Sheridan, M. R. (1995). Some characteristics of word fluency in Parkinson's disease. *Journal of Neurolinguistics, 9*, 33-46.

Gianelli, C., Maiocchi, C., & Canessa, N. (2021). Action Fluency in Parkinson's Disease: A Mini-Review and Viewpoint. *Frontiers in aging neuroscience*, *13*, 778429. https://doi.org/10.3389/fnagi.2021.778429

Gurd, J. M. (2000). Verbal fluency deficits in Parkinson's disease: Individual differences in underlying cognitive mechanisms. *Journal of Neurolinguistics, 13*(1), 47–55. https://doi.org/10.1016/S0911-6044(99)00011-1

Hamada, T., Higashiyama, Y., Saito, A., Morihara, K., Landin-Romero, R., Okamoto, M., Kimura, K., Miyaji, Y., Joki, H., Kishida, H., Doi, H., Ueda, N., Takeuchi, H., & Tanaka, F. (2021). Qualitative Deficits in Verbal Fluency in Parkinson's Disease with Mild Cognitive Impairment: A Clinical and Neuroimaging Study. *Journal of Parkinson's disease, 11*(4), 2005–2016. https://doi.org/10.3233/JPD-202473

Hedman, E., Hartelius, L., & Saldert, C. (2022). Word-finding difficulties in Parkinson's disease: Complex verbal fluency, executive functions and other influencing factors. *International journal of language & communication disorders, 57*(3), 565–577. https://doi.org/10.1111/1460-6984.12707

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology*, *18*(2), 284–295. https://doi.org/10.1037/0894-4105.18.2.284

Herrera, E., Cuetos, F., & Ribacoba, R. (2012). Verbal fluency in Parkinson's disease patients on/off dopamine medication. *Neuropsychologia, 50*(14), 3636–3640. https://doi.org/10.1016/j.neuropsychologia.2012.09.016

Hireš, M., Gazda, M., Drotár, P., Pah, N. D., Motin, M. A., & Kumar, D. K. (2022). Convolutional neural network ensemble for Parkinson's disease detection from voice recordings. *Computers in Biology and Medicine, 141*, 105021. https://doi.org/10.1016/j.compbiomed.2021.10502

Koerts, J., Meijer, H. A., Colman, K. S., Tucha, L., Lange, K. W., & Tucha, O. (2013). What is measured with verbal fluency tests in Parkinson's disease patients at different stages of the disease?. *Journal of neural transmission (Vienna, Austria : 1996), 120*(3), 403–411. https://doi.org/10.1007/s00702-012-0885-9

Piatt, A. L., Fields, J. A., Paolo, A. M., Koller, W. C., & Tröster, A. I. (1999). Lexical, semantic, and action verbal fluency in Parkinson's disease with and without dementia. *Journal of* *clinical and experimental neuropsychology, 21*(4), 435–443. https://doi.org/10.1076/jcen.21.4.435.885

Rodrigues, I. T., Ferreira, J. J., Coelho, M., Rosa, M. M., & Castro-Caldas, A. (2015). Action verbal fluency in Parkinson's patients. *Arquivos de neuro-psiquiatria, 73*(6), 520–525. https://doi.org/10.1590/0004-282X20150056

Rofes, A., & Miceli, G. (2014). Language mapping with verbs and sentences in awake surgery: a review. *Neuropsychology review*, *24*(2), 185–199. https://doi.org/10.1007/s11065-014-9258-5

Romann, A. J., Beber, B. C., Olchik, M. R., & Rieder, C. R. M. (2017). Different outcomes of phonemic verbal fluency in Parkinson's disease patients with subthalamic nucleus deep brain stimulation. *Arquivos de neuro-psiquiatria, 75*(4), 216–220. https://doi.org/10.1590/0004-282X20170024

Signorini, M., & Volpato, C. (2006). Action fluency in Parkinson's disease: a follow-up study. *Movement disorders : official journal of the Movement Disorder Society, 21*(4), 467–472. https://doi.org/10.1002/mds.20718

Smith, K. M., & Caplan, D. N. (2018). Communication impairment in Parkinson's disease: Impact of motor and cognitive symptoms on speech and language. *Brain and language*, *185*, 38–46. https://doi.org/10.1016/j.bandl.2018.08.002

Suhr, J. A., & Jones, R. D. (1998). Letter and semantic fluency in Alzheimer's, Huntington's, and Parkinson's dementias. *Archives of Clinical Neuropsychology, 13*(5), 447–454. https://doi.org/10.1016/S0887-6177(97)00040-1

Tröster, A. I., Fields, J. A., Testa, J. A., Paul, R. H., Blanco, C. R., Hames, K. A., Salmon, D. P., & Beatty, W. W. (1998). Cortical and subcortical influences on clustering and switching in the performance of verbal fluency tasks. *Neuropsychologia*, *36*(4), 295–304. https://doi.org/10.1016/s0028-3932(97)00153-x

Troyer, A. K., Moscovitch, M., Winocur, G., Leach, L., & Freedman, M. (1998). Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease. *Journal of the International Neuropsychological Society : JINS*, *4*(2), 137–143. https://doi.org/10.1017/s1355617798001374

Yang, J., McMahon, K. L., Copland, D. A., Pourzinal, D., Byrne, G. J., Angwin, A. J., O'Sullivan, J. D., & Dissanayaka, N. N. (2022). Semantic fluency deficits and associated brain activity in Parkinson's disease with mild cognitive impairment. *Brain imaging and behavior, 16*(6), 2445–2456. https://doi.org/10.1007/s11682-022-00698-7

Zec, R. F., Landreth, E. S., Fritz, S., Grames, E., Hasara, A., Fraizer, W., Belman, J., Wainman, S., McCool, M., O'Connell, C., Harris, R., Robbs, R., Elble, R., & Manyam, B. (1999). A comparison of phonemic, semantic, and alternating word fluency in Parkinson's disease. *Archives of Clinical Neuropsychology, 14*(3), 255–264. https://doi.org/10.1016/S0887-6177(98)00008-0

Understanding Parkinson's. (n.d.) Parkinson's Foundation. Retrieved January 10, 2023, from https://www.parkinson.org/understanding-parkinsons

Web of Science. (n.d.) Web of Science. https://www.webofscience.com

Language Distance and Language Production Accuracy in Dutch Bilingual Children

Research report

Ilse Hoeben[1], Anna Petukhova[2], Willeke van der Varst[3]

[1]Master *Language Variation and Multilingualism* Radboud University Nijmegen

[2]Research Master *Linguistics and Communication Sciences*, Radboud University Nijmegen

[3]Master *Artificial Intelligence - Cognitive Computing*, Radboud University Nijmegen

Abstract

When investigating bilingualism, a typological distance between languages is an important factor that might contribute to language acquisition in bilingual children. The present study investigates the role of language distance on children's language production using the 2in1 project dataset collected to explore the influence of acquired languages on each other. We used the data of 151 Dutch-English, Dutch-German, Dutch-Greek and Dutch-Turkish bilingual children from 4 to 10 years old to predict their accuracy in the sentence repetition task (SRT) by the distance between the languages in the language pair. Multiple linear regression analysis showed that the typological distance was indeed a significant predictor of the language performance: the closer the languages were, the higher the scores in the SRT were. These findings demonstrate that language distance plays a crucial role in child language development, especially with regard to productive vocabulary.

Keywords: child bilingualism, language production, language distance, cross-linguistic influence, vocabulary production

1. Background

Many studies have compared bilingual children's vocabulary with that of monolingual children. Bilingual children have a smaller vocabulary in one language than monolinguals, but their total vocabulary size (both languages combined) was similar compared to their monolingual peers (Bialystok, Luk, Peets & Yang, 2010; Cote & Bornstein, 2014). One factor that might affect bilingual children's language development is language distance, which affects all levels of the language: phonology, lexicon, syntax and pragmatics (e.g. Asli-Badarneh et al., 2023). For example, one can imagine that Dutch-German bilinguals will have more overlaps in their vocabulary due to the high number of cognates in both languages. On the other hand, Dutch-Turkish speakers are more likely to have two less overlapping lexicons. Will language distance facilitate or interfere with language production? The present study examines the role of language distance on language production accuracy of bilingual children.

Bialystok et al. (2010) examined the receptive vocabulary skills of monolingual and bilingual children, analysing 1,738 children between 3 and 10 years old using the Peabody Picture Vocabulary Test. Additionally, the study explored the impact of language distance, aiming to identify any correlation between vocabulary score differences and language pairs: bilingual children speaking English and an East Asian language were compared with those speaking English and a non-East Asian language. Interestingly, no significant effect of language distance on the vocabulary skills of bilingual children was observed, even in the case of the simplified separation of languages (East Asian and non-East Asian).

Floccia and colleagues (2018) also were interested in the relationship between language distance and vocabulary knowledge of bilingual children. The participants of that study were 2-year-old bilingual toddlers, and all were learners of British English and one of 13 additional languages. Floccia et al. (2018) measured language distance through phonological similarity, morphological complexity, and word order typology. That study found an effect of language distance on the vocabulary skills of bilingual children: children exposed to two languages with a closer linguistic distance demonstrated larger receptive vocabulary in each language. Moreover, not only phonological, but also morphological and syntactic knowledge were all transferred across the two languages.

Because of these contradictory results, Blom et al. (2020) reconsidered the role of language distance by investigating the receptive vocabulary of monolingual Dutch and bilingual (Turkish-Dutch, Moroccan-Dutch, Frisian-Dutch) children. Their receptive vocabulary was tested with the Peabody Picture Vocabulary Test. Participants were divided into three different groups: (1) bilingual group with a small between-language distance, (2) bilingual group with a large between-language distance, and (3) monolinguals (control group). The results of Blom et al. (2020) showed that language distance significantly influenced the development of receptive vocabulary in bilingual children: bilingual children with a smaller language distance had receptive vocabulary scores similar to monolingual children, whereas those with a larger language distance showed significantly lower receptive vocabulary scores. Overall, the study highlights the significance of linguistic distance as a crucial factor in individual differences and underscores the necessity for a nuanced perspective on bilingualism. Whereas most studies focused on receptive vocabulary, this study aims to investigate the role of language distance and productive skills, more specifically, productive vocabulary. Moreover, the combination of language distance and bilingual children's vocabulary is not well-researched in general. A recent paper by Dixon et al. (2022) dedicated to cross-dataset comparison (143 languages) of vocabulary development between children learning English and their monolingual peers revealed that English learners catch up with monolinguals in receptive vocabulary but not in expressive vocabulary. This discrepancy suggests that there might be different patterns of productive and receptive skills development.

2. Research Question

In our study, we will investigate to what extent language distance influences bilingual children's productive language skills measured by a sentence repetition task (SRT; Marinis & Armon-Lotem, 2015) in Dutch-English, Dutch-German, Dutch-Greek and Dutch-Turkish bilinguals. Our hypothesis is that the closer the languages are, the higher the scores in SRT will be due to a possible facilitation effect of overlapping vocabulary and syntactic structures in two languages. For example, as Dutch and German both belong to a Germanic branch of the Indo-European family, Dutch-German bilinguals are expected to score higher in the SRT in both languages than Dutch-Turkish bilinguals, because Turkish is a member of a different language family (Glottolog 4.8).

3. Participants

The data for this study is a subset of a larger dataset created by the 2in1 project (https://www.ru.nl/cls/our-research/research-groups/cognitive-developmental-aspects-multilingualism/2in1-project-nl/) that investigates language interaction in simultaneous or early sequential bilingual children. To cover the largest number of different language pairs represented in the projects, data of 151 Dutch-English, Dutch-German, Dutch-Greek and Dutch-Turkish bilingual children living in the Netherlands from 4 to 10 years old (Table 1) were selected for the study.

Table 1
Participants Overview

Dataset	Language pair	Age range, years	Number
cvd3	Dutch-German	7-10	31
cvd3	Dutch-Turkish	7-10	23
gjk1	Dutch-German	7-10	35
gjk1	Dutch-English	7-10	36
ek1	Dutch-Greek	4-9	26

4. Methods

The 2in1 dataset consists of a parental questionnaire as a measure of language exposure BiLEC (Bilingual Language Experience Calculator; Unsworth, 2013), SRT scores and productive vocabulary task scores as measures of children's language production abilities. In accordance with our research question, the selection of the datasets was motivated by the diversity of represented languages (we aimed at including various language pairs with the dataset we had access to) to provide for different levels of language relatedness.

Among all BiLEC variables, our choice fell on the age of testing (age_testing), the percentage of the use of Dutch during activities (nld_richness) and the amount of Dutch used by children at home (nld_output_home). This selection was motivated by the intent to narrow down the model in the interest of time, focusing on language input and output. Both language exposure (Thordardottir, 2019) and use (Ribot et al., 2018) were reported to have influence on productive vocabulary skills.

Age as a parameter was included to control for the overall amount of the time the language is used.

SRT was chosen as a measure of productive skills as the data on this task was provided for the majority of the language pairs in the dataset. VerbatimPropCorr and GrammScore.StrictPropCorr variables for the task in Dutch and another language were selected for the model since they were present across all datasets. VerbatimPropCorr stands for the accuracy of repeated sentences, 1 being 'the sentence was repeated entirely verbatim' (Marinis & Armon-Lotem, 2015, p. 23) and 0 being 'there were one or more changes in the children's response' (Marinis & Armon-Lotem, 2015, p. 23). GrammScore.StrictPropCorr is a score of grammaticality of the sentence varying from 0, there was a grammatical error, to 1, there were no errors. To the dataset, a column was added for each participant, detailing how closely related the languages they speak are. This new variable was called language_distance and was coded as a continuous variable with 1 being the most closely related and 4 being the least closely related. Dutch-German was given the value 1, Dutch-English the value 2, Dutch-Greek the value 3 and Dutch-Turkish the value 4.

After this, linear regression models were fitted in R (R Core Team, 2022) to see whether this language_distance variable is an important predictor for SRT-scores. 4 different types of base models were made, according to the type of SRT-score that was to be predicted:

- Verbatim model: predicts VerbatimPropCorr scores.
- OL Verbatim model: predicts OL_VerbatimPropCorr scores.
- GrammScore model: predicts GrammScore.StrictPropCorr scores.
- OL GrammScore model: predicts OL_GrammScore.StrictPropCorr scores.

All these base models predicted scores based on age_testing, nld_output_home and nld_richness. After all base models were run, language_distance was added to them all to verify the effect of language_distance on the predicted scores. For the Verbatim models, data was available from all languages (Dutch-German, Dutch-English, Dutch-Greek and Dutch-Turkish), but for the GrammScore models Dutch-Turkish data was not available.

5. Results

Multiple Linear regressions were calculated to predict SRT-test scores based on age_testing, nld_output_home and nld_richness. In all cases, a significant regression equation was found (see Table 2).

For all model types, inclusion of the language_distance variable, improved the model's AIC-score (see Table 3). Lower AIC-scores indicate better predictor models, so inclusion of language_distance improves the model.

Table 2

Regression Analysis Results

Model	R2-value	F-value	Degrees of Freedom	p-value
Verbatim	0.40	25.66	3, 143	<0.005
OL Verbatim	0.58	50.82	3, 143	<0.005
GrammScore	0.52	18.92	3, 63	<0.005
OL GrammScore	0.64	30.25	3, 63	<0.005

Table 3

AIC-scores for all Models (Base and With Inclusion of language_distance)

Model	AIC-score Base	AIC-score with language_distance
Verbatim	-68.24	-113.71
OL Verbatim	62.06	-12.26
GrammScore	-78.30	-79.29
OL GrammScore	1.47	-16.82

For each model with language_distance included, it was checked which variables were significant predictors, to check if language_distance is indeed an important predictor variable (see Table 4). The b-value indicates the magnitude and direction of the influence of the variable to the score. A negative b-value indicates that the lower the value of the variable is, the higher the predicted score is and vice versa.

This means that for all models, except the Grammar Score, language_distance is a significant predictor, and the direction of the effect is always negative: the closer the languages are, the higher the SRT-score is. The distribution of SRT-scores, based on language distance, can be seen in the Figures 1-4.

Table 4

Significant Predictors for Each Model and Their b-coefficient, Standard Deviation (std), t-value and p-value

Model	Significant predictors	b-coefficient	Std	t-value	p-value
Verbatim	age_testing	0.05	0.01	5.34	<0.005
	language_distance	-0.10	0.01	-7.37	<0.005
OL Verbatim	age_testing	0.06	0.01	4.29	<0.005
	nld_output_home	-0.23	0.07	-3.38	<0.005
	nld_richness	-0.39	0.11	-3.45	<0.005
	language_distance	-0.18	0.02	-9.84	<0.005
GrammScore	age_testing	0.07	0.01	6.27	<0.005
	language_distance	-0.05	0.03	-1.80	0.07
OL GrammScore	age_testing	0.06	0.02	3.76	<0.005
	nld_output_home	-0.17	0.09	-1.94	0.06
	nld_richness	-0.43	0.15	-2.91	0.01
	language_distance	-0.20	0.04	-4.73	<0.005

Figure 1

Predicted Verbatim Scores per Language Distance Group

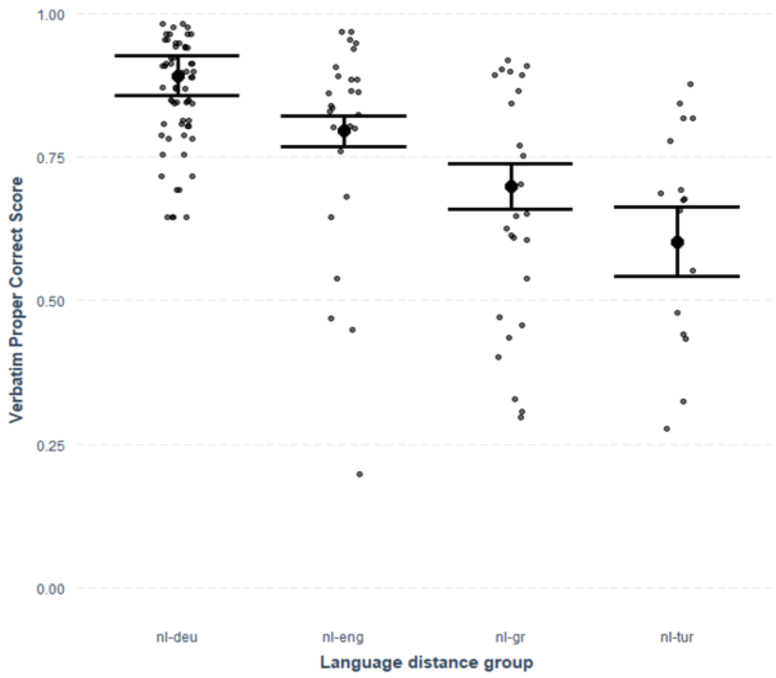


Figure 2

Predicted OL Verbatim Scores per Language Distance Group

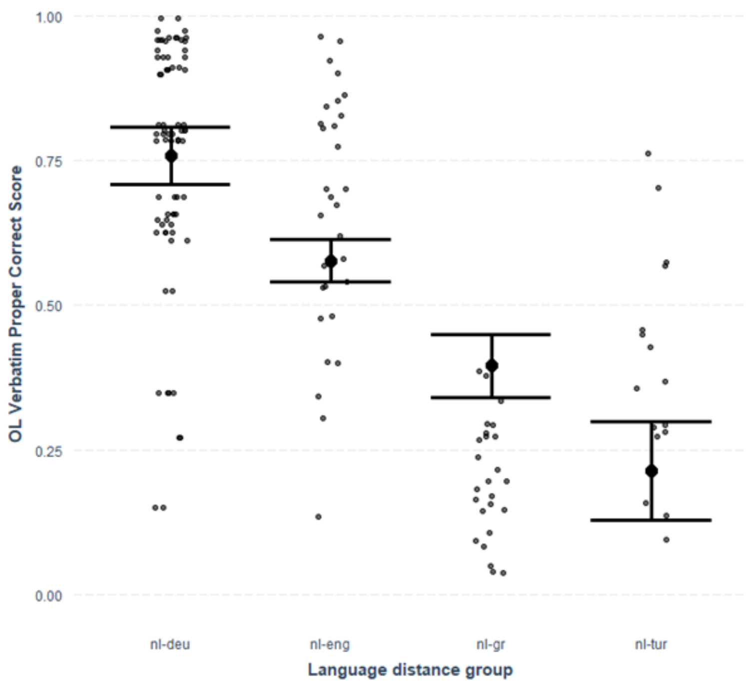


Figure 3

Predicted Grammatical Scores per Language Distance Group

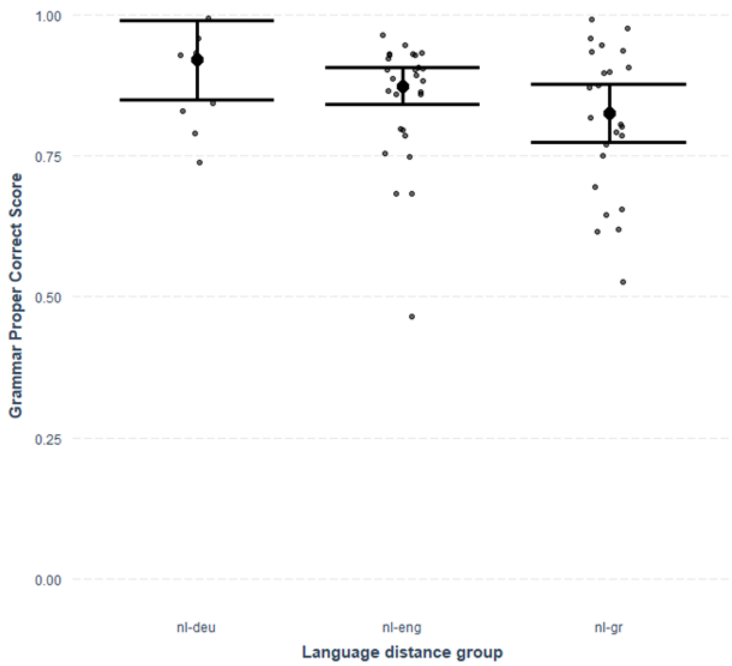
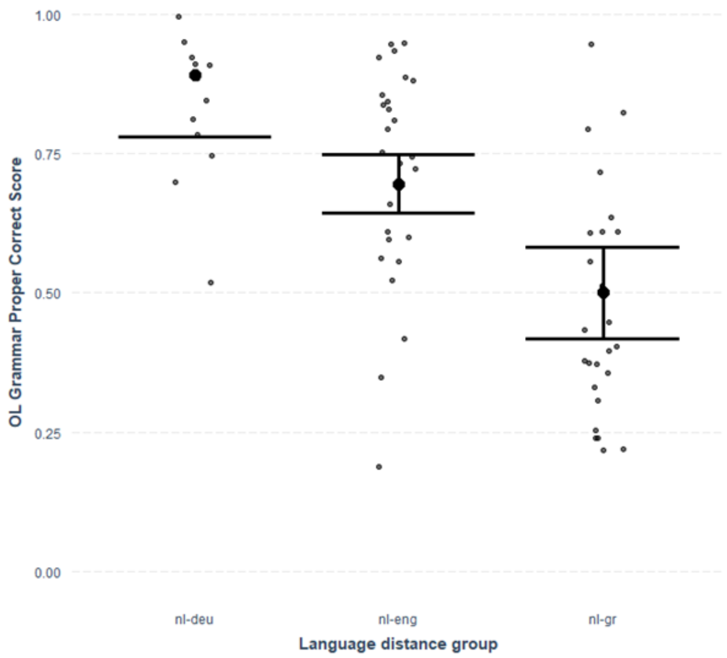


Figure 4

Predicted OL Grammatical Scores per Language Distance Group



6. Discussion

Inclusion of language distance as a predictor in the regression models improved the models: AIC-scores were lower over the base models which did not include language_distance, meaning that language distance is an significant predictor of SRT-scores. Moreover, the effect of language_distance on the SRT-scores was always negative and significant, suggesting that the closer the languages are (lower language distance), the higher the SRT-scores are. These results are in agreement with our hypothesis: children speaking more typologically similar languages score higher in the productive vocabulary task due to a possible facilitation effect of overlapping vocabulary and syntactic structures in two languages. The results are also in accordance with the described literature (e,g, Blom et al., 2020), supporting the claim that if languages a person learns are more closely related, they perform better on tests in both languages. Compared to the literature mentioned above that only looked at receptive vocabulary, our study addressed productive vocabulary. However, as mentioned in the background section, the productive skills of bilingual children can be quite different from their receptive skills, thus research into their productive skills should not be ignored.

Some interesting findings were the following: The largest influence of language distance was observed for the verbatim component of the SRT-tasks in both languages, which can be indicated by the AIC-score improvements. While the grammar scores did also improve, the AIC-value difference was smaller than for the verbatim results. This could be due to the difficulty of both tasks: in the verbatim task children have to repeat exactly what was being said, while in the grammar task only the sentence structure needs to be correct. Thus, the verbatim task may be harder and more proficiency in the language may be needed. This would explain the difference in AIC-scores between both tasks, as the harder task might benefit more from language similarities.

There is quite a large variation between the AIC-score differences between the GrammScore and the OL GrammScore models, with the AIC-score difference in the GrammScore model being quite small. This could be, because exposure to the main language (in this case Dutch) was a lot higher for most children in this study than their other language, indicating some sort of the threshold effect. This means that the higher exposure there is to a language, the less influence of language distance there is on certain tasks. This effect was not visible for the OL GrammScores, because exposure to the other language was not high enough, thus the threshold for less language distance effects was not present.

7. Conclusion

We found that language distance is a significant predictor of SRT-scores. Moreover, language distance negatively influences SRT-scores, meaning that the closer the two languages are related, the higher the test-scores are. These results are in line with our hypothesis and previous literature. Future research could address some shortcomings in our study:

– In our regression models, we only included predictors related to language use and output, while not including predictors related to socioeconomic status, family composition, neighbourhood, etc. due to time constraints. In a future study, these variables could be added to more accurately verify the influence of language distance on test scores.
– While 151 participants is not a low number, regression models become more accurate the more data you have, so gathering data from more participants could yield more accurate results. Moreover, we mainly tested children from age 7-10, so expanding this age range could also give more insights into the effects of language distance.
– In this study, we only looked at one task (SRT) as a measure of how well children are able to speak a language. Inclusion of a broader spectrum of language tasks could shed more light on the effects of language distance while learning two languages.

These outcomes provide a better understanding of child bilingualism, and they demonstrate that language distance plays a role in bilingual children's productive vocabulary skills.

References

Asli-Badarneh, A., Hipfner-Boucher, K., Bumgardner, X. C., AlJanaideh, R., & Saiegh Haddad, E. (2023). Narrative microstructure and macrostructure skills in Arabic diglossia: The case of Arab immigrant children in Canada. *International Journal of Bilingualism, 27*(3), 349-373. https://doi-org.ru.idm.oclc.org/10.1177/13670069221077306.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism* (Cambridge, England), *13*(4), 525–531. https://doi.org/10.1017/S1366728909990423.

Blom, E., Boerma, T., Bosma, E., Cornips, L., van den Heuij, K., & Timmermeister, M. (2020). Cross-language distance influences receptive vocabulary outcomes of bilingual children. *First Language*, *40*(2), 151–171.https://doi.org/10.1177/0142723719892794.

Cote, L. R., & Bornstein, M. H. (2014). Productive vocabulary among three groups of bilingual American children: Comparison and prediction. *First Language*, *34*(6), 467-485. https://doi.org/10.1177/0142723714560178.

Dixon, C., Hessel, A., Smith, N., Nielsen, D., Wesierska, M., & Oxley, E. (2022). Receptive and expressive vocabulary development in children learning English as an additional language: Converging evidence from multiple datasets. *Journal of Child Language*, 1-22. https://doi.org/10.1017/S0305000922000071.

Floccia, C., Sambrook, T.D., Luche, C.D., Kwok, R.K., Goslin, J., White, L., Cattani, A., Sullivan, E., Abbot-Smith, K., Krott, A., Mills, D., Rowland, C.F., Gervain, J., & Plunkett, K. (2018). Vocabulary of 2-Year-Olds Learning English and an Additional Language: Norms and Effects of Linguistic Distance. *Monographs of the Society for Research in Child Development*, *83*(1), 7–29.

Glottolog 4.8. (2023). English (en). Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://glottolog.org, Accessed on 2023-12-07.)

Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. *Assessing multilingual children: Disentangling bilingualism from language impairment*, 95–124.

Ribot, K. M., Hoff, E., & Burridge, A. (2018). Language use contributes to expressive language growth: Evidence from bilingual children. *Child Development, 89*(3), 929-940. https://doi.org/10.1111/cdev.12770.

Thordardottir, E. (2019). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism, 23*(1), 236-255. https://doi.org/10.1177/1367006917722418.

Unsworth, S. (2013). Assessing the role of amount and timing of exposure in simultaneous bilingual development and ultimate attainment. *Bilingualism: Language and Cognition, 16*, 86–110.

Code-Switching Production Patterns Predicting Comprehension in Yakut-Russian Bilinguals

Research proposal

Anna Petukhova

Research Master *Linguistics and Communication Sciences*, Radboud University Nijmegen

Abstract

Switching between languages is a distinctive feature of being bilingual. Investigating code-switching in psycholinguistics can reveal the mechanisms behind language interaction in a bilingual mind since two language systems are activated. In this study, I will be testing a hypothesis introduced in Tamargo et al. (2016) that code-switching comprehension reflects production patterns of code-switching by using the data from eye movements of young adult Yakut-Russian bilinguals during a reading task. The prediction is that switches at more frequently witnessed syntactic sites (verb) in speech corpora would result in shorter reading times than switches that occur at less often observed boundaries (e.g., with a modal particle). In addition, participants will be divided into two groups based on their code-switching behavior as code-switching processing may differ depending on the frequency of being involved in such practice. The findings will contribute to the research of contact-induced phenomena in typologically distinct languages and the investigation of the relationship between production and comprehension.

Keywords: bilingualism, code-switching, language comprehension, eye-tracking

1. Introduction

Code-switching, 'the use of two language varieties in the same conversation' (Myers-Scotton, 2005, p. 239), is a widespread language phenomenon among bilinguals. There are two types of code-switching being distinguished: intersentential, code switches that happen between clauses, and intrasentential, code switches that occur within the clause (Poplack, 2013).

From a psycholinguistic perspective, code-switching draws attention as a window to the cognitive mechanisms behind the competition between languages as both languages remain activated (Traxler, 2011, p. 416). One prominent topic in the research of code-switching is switching costs which is a difference in comprehension time when encountering a switch or a non-switch. It is usually asymmetric, especially in unbalanced bilinguals: switching to a dominant language takes longer time than switching to a non-dominant one (Costa & Santesteban, 2004).

Experimental paradigms in the research on code-switching include corpus (e.g., Fricke & Kootstra, 2016), elicitation (e.g., Kootstra et al., 2020), self-paced reading (Adler et al., 2020), rapid serial visual presentation, and eye-tracking (e.g., Lipski, 2020) studies. The advantages of the latter method in code-switching processing research, according to Kroff et al. (2018), overcome such methodological challenges as the artificialness of the task and stimuli, allowing higher ecological validity in the studies of reading comprehension. In addition, using corpora data based on spontaneous speech as experimental material complements ecological validity even further.

Code-switching research based on corpus data has a number of advantages. First, it allows for high generalizability of the findings as they can be based on rich data. Second, corpora data, which is usually based on hundreds of hours of natural speech, allows for high ecological validity. Computer linguistics tools enable relatively fast and easy data extraction of any language phenomena of interest if it is annotated in the corpus.

One study that implemented both corpus and eye-tracking methods is a paper by Tamargo and colleagues (2016). It tested a hypothesis that distributional patterns of code-switching influence code-switching comprehension by implementing statistical knowledge to predict cues that will be followed by a code-switch. First, the authors made an exploratory spontaneous speech corpus analysis to determine frequent constructions where code-switch happens. They found out that there were more instances of code-switching at present participle or its auxiliary than at perfect participle and its auxiliary when switching from Spanish to English, which allowed them to build their hypothesis further, stating that they would expect higher processing costs for the code switches at perfect participle. Next, Tamargo et al. (2016) performed an eye-tracking experiment with Spanish-English bilinguals with early (*N*=42) and late exposure to code-switching (*N*=27). Participants' first-pass reading time and total time were measured while they were reading sentences in four conditions: 1) a code switch at the auxiliary (present), 2) a code switch at the participle (present), 3) a code switch at the auxiliary (perfect) and 4) a code switch at the participle (perfect). Linear mixed effects models analysis showed that both groups' comprehension facilitated from the production patterns found in corpora: code switches at the perfect auxiliary appeared to be more costly than code switches at present auxiliary, confirming the hypothesis.

The current study aims to provide more evidence for the patterns of code-switch comprehension for the pair of understudied and unrelated languages, Yakut (>Turkic) and Russian (>Indo-European). Yakut language is predominantly spoken on the territory of Sakha Republic, Russia, has a governmental status, and is compulsorily taught in schools (as L2 in Russian-speaking schools). The majority of the Yakut population speaks Russian which is the national language in all subjects of the Russian Federation. The use of Russian, however, is more predominant in towns, while Yakut is more frequently spoken in rural areas. For most families, Yakut is mainly spoken at home, as Russian remains to be a primary language of education. Hence, it is quite common to be a balanced Yakut-Russian bilingual in terms of the frequency of use, although the domains might not overlap.

The study follows the research line by Tamargo and colleagues (2016) and will investigate the influence of code-switching patterns in spoken production on reading comprehension. Even though production and comprehension modalities differ, Kroff and colleagues (2018) argue the patterns of code-switching are equivalent in written and spoken language. Narrowing the research question down, I will look at the presence of switching costs, reflected in the length of the reading time in an eye-tracking experiment, at the verb at the ending position when switching from Russian to Yakut. It has been attested in a Yakut-Russian spontaneous speech corpora (Petukhova & Sokur, 2021) that intrasentential switches occur predominantly when switching from Yakut to Russian, and the most frequent syntactic site that a code switch happens at is a verb/verb phrase. Thus, I hypothesize that code switches at the boundary with a verb will be read faster than switches with a boundary on other syntactic phrases. Furthermore, participants who code-switch regularly in their daily life are expected to read code switches quicker than the participant who rarely participate in such a language behaviour. In addition, the study can also pave the way for other research in Russian and a minority language pair. There are more than a hundred languages spoken in different regions of Russia that belong to various language families, especially with a lot of variety in the East Caucasus area.

2. Methods

2.1 Participants

Balanced/early young adult Yakut-Russian bilinguals will participate in the study, as the corpus I base our hypothesis on represents speech of that sample of a population. Participants will fill in a language background questionnaire (Anderson et al. 2018), including questions about code-switching habits (Rodriguez-Fornells et al., 2012), and complete language proficiency tasks in both languages. Participants whose scores in both tasks would differ in more than two standard deviations will be excluded from the analysis. Next, participants will be divided into the groups of frequent and non-frequent code switchers based on the questionnaire scores: participants with a mean score of up to 2 will be considered non-frequent switchers, participants with a mean score of 4 to 5 will be considered frequent switchers, the data of the rest will not be included into the analysis (5-point Likert scale used for the evaluation). Previous research showed that production can influence comprehension of code-switches (e.g., Beatty-Martínez & Dussias, 2017), so the differentiation based on the frequency of code switching might reveal differences in code-switch processing as well. Participants will be matched on age, socio-economic status, and place of residence to exclude the influence of dialect differences to avoid possible confounding variables and identify the specific linguistic factors that influence code-switching. Participants will be recruited via advertisement in social media, messengers, and word of mouth and will be compensated for their time.

The number of participants will be calculated using the sjstats package (Lüdecke, 2021) in R (R Core Team, 2022) with a large effect size determined by Cohen’s d (0.8).

2.2 Materials

Following Keating's (2014) suggestion, 12 experimental stimuli will be used per condition, resulting in 24 items: condition 1 is a switch at the verb at the end of the sentence (Example 1), and condition 2 is a switch at the last word of the sentence which is a modal particle (Example 2). The latter was chosen to be a control condition, as it has been attested to be less common than a code switch at the verb in the corpus. In addition, 48 fillers will be embedded into the final list of stimuli, consisting of sentences with one-word long switches at different positions in a sentence, except for the final one. All stimuli will represent switches from Russian to Yakut according to the switching frequency asymmetry reported in the corpus. To prevent spill-over effects, each stimulus will consist of two sentences: the first sentence will be a sentence starting in Russian with a code-switch to Yakut at the end of a sentence, and the second sentence will be a monolingual sentence in Russian (Russian and Yakut share the writing system). Regions of interest (ROIs) will be set at the first two words of the first sentence to analyze the factors influencing comprehension before encountering a code -switch; a critical region of a code switch, and two words at the beginning of the second sentence to analyze a spill-over effect of a code switch processing.

(1a)	*Представляешь,*	*сегодня*	*я*	*успел*	*сделать*
	imagine.IPF.PRS.2.SG	today	1.SG.NOM	manage.PRF.PST.M.SG	do.PRF.INF

все	*дела,*	*полезный*	*день*	*ааспыт.*
all.PL.ACC	deed.PL.ACC	efficient.M.SG.NOM	day.SG.ACC	pass.PRF.PST.M/N.SG

'Imagine, today I have managed to complete all my tasks, it was an efficient day.'

(1b)	*Надеюсь,*	*завтра*	*я*	*наконец-то*	*доберусь*
	hope.IPF.PRS.1.SG	tomorrow	1.SG.NOM	finally	get.PRF.FUT.1.SG

до	*бассейна,*	*давно*	*там*	*не*	*был.*
to	pool.SG.GEN	long.ago	there	not	be.IPF.PST.M.SG

'Hope that tomorrow I will finally get to the pool, I haven't been there for a long time.'

(2a)	*Ты*	*не*	*купила*	*стиральный*	*порошок,*
	2.SG.NOM	not	buy.PRF.PST.F.SG	washing.M.SG.ACC	powder.SG.ACC

как	*я*	*тебя*	*просил,*	*да?*
how	1.SG.NOM	2.SG.ACC	ask.IPF.PST.M.SG	MOD

'You haven't bought the washing powder as I asked you?'

(2b)	*Тогда*	*мы*	*перенесем*	*стирку*	*на*	*завтра,*
	then	1.PL.NOM	postpone.PRF.FUT.1.PL	wash.SG.ACC	to	tomorrow

когда	*я*	*схожу*	*в*	*магазин.*
when	1.SG.NOM	go.PRF.FUT.1.SG	to	store.SG.ACC

'Then we'll postpone the laundry for tomorrow, when I go to the store.'

All sentences will be constructed to match in length (8-10 words, following corpus data). The words at the code-switch will be matched in length and frequency to meet the criteria of equal predictability. Moreover, as Yakut script has additional characters not present in Russian script, critical words will be chosen in such a way that no characters absent in Russian will be in a word to keep spelling uniform.

Stimuli will be followed by comprehension questions in Russian to ensure that participants pay attention and read sentences carefully for better reflection of natural reading patterns. Experimental items will comprise two lists, as there are two experimental conditions, and participants will be assigned the lists in a random order. Within each list stimuli will be presented in a pseudorandomized order, so no stimulus type (code switch at a verb, code switch at a modal particle, filler) will be encountered by a participant more than three times in a row.

2.3 Procedure

Participants will be sat in a noise-isolated booth 70 cm away from a computer monitor using a chin rest to prevent head movement. Eye movements will be recorded using a desktop Eyelink 1000 system (SR Research, Mississauga, Ontario, Canada) with the camera located below the screen. Monocular eye movements of a dominant eye will be recorded at a sampling rate of 1000 Hz given the size of the stimuli (text) and ROIs. Text will be presented in a Courier New font with 14 pt size and double line spacing. At the beginning of the experiment and after each break (two breaks in total, after 24 stimuli) eye tracker will be calibrated with the accuracy rate set to 0.3°. Participants will perform a nine-point calibration and validation before each trial. It will start with a calibration point appearing on the left side of the screen at the position of the first word in the sentence.

Participants will be asked to read sentences as they would usually read them, followed by a comprehension question regarding the content of the sentences to which they would need to answer 'yes' or 'no' by pressing Q and P buttons on the keyboard respectively. Prior to the experiment, participants will fill out the language background questionnaire and complete language proficiency tasks. At the beginning of the experiment, participants will perform a practice trial consisting of six stimuli to get acquainted with the procedure.

3. Data analysis

In the analysis, I will use first-pass reading time (a sum of fixation durations between first entering an ROI and first leaving it), and total time (a sum of all fixation duration in an ROI), following the experimental setting in Tamargo et al. (2016). As suggested by Conklin et al. (2017), inaccurate trials (trials where there was a wrong answer to a comprehension question), first fixation, and fixations with a duration of less than 100 ms. or two standard deviations above and below the participant's mean will be excluded from the analysis.

Additionally, linear mixed-effects models will be used via the lme4 package (Bates et al., 2015) in R (R Core Team, 2022), with a switch type (verb or modal particle) and participant group (frequent or non-frequent switchers) and their interaction as fixed factors and stimuli and participants as random factors.

References

Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods, 50*(1), 250–263. https://doi.org/10.3758/s13428-017-0867-9.

Adler, R. M., Valdés Kroff, J. R., & Novick, J. M. (2020). Does integrating a code-switch during comprehension engage cognitive control? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(4), 741. https://doi.org/10.1037/xlm0000755.

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language, 95*, 173-189. https://doi.org/10.1016/j.jml.2017.04.002.

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research.* Cambridge University Press.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50*(4), 491-511. https://doi.org/10.1016/j.jml.2004.02.002.

Fricke, M., & Kootstra, G. J. (2016). Primed codeswitching in spontaneous bilingual dialogue. *Journal of Memory and Language, 91*, 181-201. https://doi.org/10.1016/j.jml.2016.04.003.

Keating, G. D. (2014). Eye-tracking with text. *Research Methods in Second Language Psycholinguistics*, 69-92.

Kootstra, G. J., Dijkstra, T., & Van Hell, J. G. (2020). Interactive alignment and lexical triggering of code-switching in bilingual dialogue. *Frontiers in Psychology*, 1747. https://doi.org/10.3389/fpsyg.2020.01747.

Kroff, J. R. V., Tamargo, R. E. G., & Dussias, P. E. (2018). Experimental contributions of eye-tracking to the understanding of comprehension processes while hearing and reading code-switches. *Linguistic Approaches to Bilingualism, 8*(1), 98-133. https://doi.org/10.3389/fpsyg.2020.01747.

Lipski, J. M. (2020). Pronouns, Interrogatives, and (Quichua-Media Lengua) Code-Switching: The Eyes Have It. *Languages, 5*(2), 11. https://doi.org/10.3390/languages5020011.

Lüdecke, D. (2021). *sjstats: Statistical Functions for Regression Models (Version 0.18.1).*

Myers-Scotton, C. (2005). *Multiple voices: An introduction to bilingualism.* Wiley-Blackwell

Petukhova, A. A., & Sokur, E. O. (2021). *Yakut-Russian Corpus of Code-Switching*, Moscow: International Laboratory of Language Convergence, Higher School of Economics. Available online at: http://lingconlab.ru/cs_yakut, accessed on 06.04.2022.

Poplack, S. (2013). "Sometimes i'll start a sentence in spanish y termino en español": toward a typology of code-switching. *Linguistics, 51*(S1), 11–14. https://doi.org/10.1515/ling-2013-0039.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from https://www.R-project.org/.

Rodriguez-Fornells, A., Kramer, U., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology, 2*, 388. https://doi.org/10.3389/fpsyg.2011.00388.

Tamargo, R. E. G., Kroff, J. R. V., & Dussias, P. E. (2016). Examining the relationship between comprehension and production processes in code-switched language. *Journal of Memory and Language, 89*, 138-161. https://doi.org/10.1016/j.jml.2015.12.002.

Traxler, M. J. (2011). *Introduction to psycholinguistics: Understanding language science.* Wiley-Blackwell.zz

Appendix A: List of Used Abbreviations

1	first person	N	neuter
2	second person	NOM	nominative
3	third person	MOD	modal particle
ACC	accusative	PASS	passive
F	feminine	PL	plural
FUT	future	PRF	perfect
GEN	genitive	PRS	present
INF	infinitive	PST	past
IPF	imperfective	SG	singular
M	masculine		

Language Distance and its Role in the Relationship Between Exposure and Language Dominance in Dutch Bilingual Children

Research report

[1]Lotte Arendsen-Kalıncı, [2]Annemiek Pouw

[1]Bachelor *Psychology*, Radboud University Nijmegen

[2]Master *Pedagogical Sciences,* Radboud University Nijmegen

Abstract

In research, bilingual language development differs from monolingual language development due to high variability present in bilinguals. This variance includes degrees of exposure to either language and language distance, which is the degrees in which language systems differ from each other. A distinction can be made between close language pairs (CLP) and distant language pairs (DLP). Disparities in research findings seem to indicate that there is no clear cut answer to the question of to which extent language distance influences the effects of exposure on language dominance. Using the dataset from the 2in1 project, this paper aims to investigate the research question of to which extent language distance influences the relationship between home exposure and the degree of dominance among Dutch bilingual children. The Sentence Repetition Task (SRT) and Cross-Linguistic Language Task (CLT) were administered among 147 bilingual participants between the ages of 5 and 10 years old. Using ratio scores in both languages, language dominance was calculated for each participant. Relative exposure was measured using the Bilingual Language Experience Calculator (BILEC). Results confirmed the effect of relative exposure on language dominance in both measures of proficiency. This paper also found that the interaction between relative exposure and language dominance seems to be present, but not within all measures of proficiency as it was found for SRT-dominance but not for CLT-dominance. Further research should include additional language proficiency measures to test whether the interaction effect between relative exposure and language dominance could be present in other language domains.

Keywords: cross-linguistic effect, language dominance, language distance, language exposure, bilingual development

1. Introduction

Bilingual development contains higher degrees of variability compared to monolingual development, such as differences in relative exposure and similarities between both languages (Knopp, 2022). Furthermore, some researchers suggest that different language systems, that are present within bilinguals, influence each other (Antonieu et al., 2012). Due to these degrees of variability within bilingual language development, bilingual individuals are rarely ‘balanced’, meaning being equally proficient in both of their languages. Instead, it is more common for individuals to have a dominant language and a weaker language (Knopp, 2022). Researchers often operationalize language dominance as the relative proficiency of a bilingual individual in each language, meaning that the language scores of both languages are compared to each other in order to determine which language is more dominant (Birdsong, 2016; Knopp, 2022). A growing body of evidence suggests that relative language exposure is an acceptable proxy for language dominance (e.g. Unsworth et al., 2018). However, there is no true consensus on the operationalization of language dominance, as it cannot be understood as a static concept in which an individual is either fully dominant or non-dominant within a given language. Rather, it is a dynamic concept, which is relativistic, multidimensional, and gradient in nature (Birdsong, 2016; Knopp, 2022).

Research indicates that relative language exposure to each language in bilinguals correlates to the proficiency in the same language. Various researchers found that high exposure to the majority language has positive effects on both comprehension and production of this language, whilst having a negative effect on the minority language. These patterns were also found for exposure in the minority language (Thordardottir, 2011, 2019; Floccia et al., 2018). Furthermore, participants who were equally exposed to both languages, scored similarly in both languages (Thordardottir, 2011, 2019). This was found to be the case for both expressive and receptive scores in French and English bilingual children. Thordardottir (2011) suggested that similar scores of the equal language exposure group may be due to the similarities of the languages that were tested, which is also referred to as language distance.

According to Radman et al. (2021), language distance is ‘the extent to which two languages have different vocabulary, syntactic structure, phonemes, spelling and pronunciation, orthography and writing systems’ (p.1). Language distance includes close language pairs (CLP), in which language systems are similar, and distant language pairs (DLP), in which language systems are less similar (Radman et al., 2021).

The distance between different language systems could be of equal importance as language exposure when looking at the language dominance of bilingual children, for example, when looking at cross-linguistic influence. This term describes the idea that the different language systems present within bilinguals influence each other (Antonieu, 2012). Various researchers have investigated whether language distance, combined with the degree of language exposure, affects language dominance in the languages spoken by a bilingual child.

Some researchers have found that cross-linguistic influence may cause the language systems to reinforce each other (e.g. Floccia et al., 2018; Blom et al., 2020). For one, there is evidence that bilingual toddlers have a larger production and comprehension vocabulary in their additional language if there is a greater overlap in phonology, typology and morphological complexity between the English language and their additional language (Floccia et al., 2018). This was measured through both production and comprehension tasks in English bilingual children. Other researchers found similar results for receptive language scores in non-English language pairs (Blom et al., 2020).

Other research, such as Knopp (2022), found that the overlap in language systems may not matter as much. Although CLP bilinguals had significantly more exposure to the majority language in comparison to DLP bilinguals, vocabulary scores showed that both groups were dominant in the majority language and had similar language scores (Knopp, 2022).

Alternatively, research suggests that cross-linguistic influence can interfere with language performance of bilingual individuals. This may be the case, for example, when two language systems within an individual are highly developed (Antonieu et al., 2012). Even when two language systems within bilingual individuals are highly developed, cross-linguistic influence remains. This influence allows bilinguals to co-activate different language strategies when faced with a linguistic task. However, if these strategies are drastically different, which is often the case with highly developed DLP bilinguals, this also could result in difficulty to eliminate one of two strategies in favor of the other. This theory is supported by the work of Liu and Ning (2021), who investigated the selective attention of Cantonese-Urdu bilinguals during the processing of segments and tones in Cantonese. They found that Urdu-dominants classified Cantonese stimuli along segments rather than tones, maintaining their L1 strategy, whereas Cantonese-dominants employed similar language strategies to native speakers, being more attentive to tones. However, when the stimuli was manipulated to contain both characteristics of Urdu and Cantonese, Cantonese- dominants had a significantly lower reaction time compared to Cantonese natives and Urdu-dominants, suggesting that Cantonese-dominants had to actively repress their L1-strategy in favor of their L2-strategy.

In sum, there is an indication research that relative language exposure to each language in bilinguals correlates to the proficiency in the same language (Thordardottir, 2011, 2019). As for cross-linguistic influence, current research provides mixed support for its effect on language performance of bilingual children. Though there is evidence that closely related language pairs have a positive influence on the expressive and receptive language skills of the additional language (Floccia et al., 2018), others have found that distance between language pairs has a negative influence on skills in the additional language skills (Liu and Ning, 2021), or does not influence language dominance at all (Knopp, 2022).It seems that there is no clear cut answer to the question of how language distance influences the relationship between exposure and language dominance. This could be because researchers conceptualize language dominance in different ways, which complicates making comparisons between studies. Furthermore, countless combination of language pairs are possible for bilingual children, which makes comparing research on language pairs even more difficult. It is therefore necessary to further examine the relationship between exposure and language dominance, while focusing on language distance within the language pairs of bilinguals.

1.1 Current Research

Using the dataset from the 2in1 project (Unsworth et al., 2022), this paper investigates to which extent language distance influences the relationship between home exposure and the degree of dominance among bilingual children. In order to answer this question, two measures are used. First, the question is answered by looking at the global effect of language distance in comparing CLP- and DLP bilinguals. Next, the different language pairs of the DLP-group will be examined to see if there is a difference in the degree of similarity between language pairs.

Based on the findings by Thordardottir (2011; 2019), it is expected that language pairs that are relatively more distant result in bilinguals being more dominant in either language compared to less distant- and closely related language pairs. This would translate into less discrepancy between the language scores of the CLP bilinguals versus DLP bilinguals.

2. Method

2.1 Participants

A total of 147 participants between the ages of 5 and 10 (*M* = 7,4) and their existing data from other studies of the 2in1 project (Unsworth et al., 2022) were used to answer the research questions. The distribution of studies that were used to form groups of language pairs can be seen in Table 1. The distribution of participants over the language groups can be seen in Table 2.

Table 1

Studies Used per Language Distance Pair (N=147)

Language distance pairs		**Studies**	**Number of participants**
CLP (N = 106)	English - Dutch	cvd1	40
		gjk1	36
		su1	30
DLP (N = 53)	Turkish - Dutch	cvd3	23
	Spanish - Dutch	su1	30

Total	**159**

Note. The number of participants displayed in the table shows the dataset before controlling for possible errors in the data.

Table 2

Distribution of Participants Among Language Groups and Tasks (N = 147)

Language distance pairs		**Scores**	
		CLT	**SRT**
CLP (N = 102)	English - Dutch	60	93
DLP (N = 45)	Turkish - Dutch	16	19
	Spanish - Dutch	29	X
Total		105	112

Note. This table displays the distribution of the combined test scores of the Cross-Linguistic Lexical Task (CLT) and the Sentence Repetition Task (SRT) in Dutch and the additional language among Close Language Pairs (CLP) and Distant Language Pairs (DLP). The SRT was not administered to the Spanish-Dutch children.

2.2 Procedure

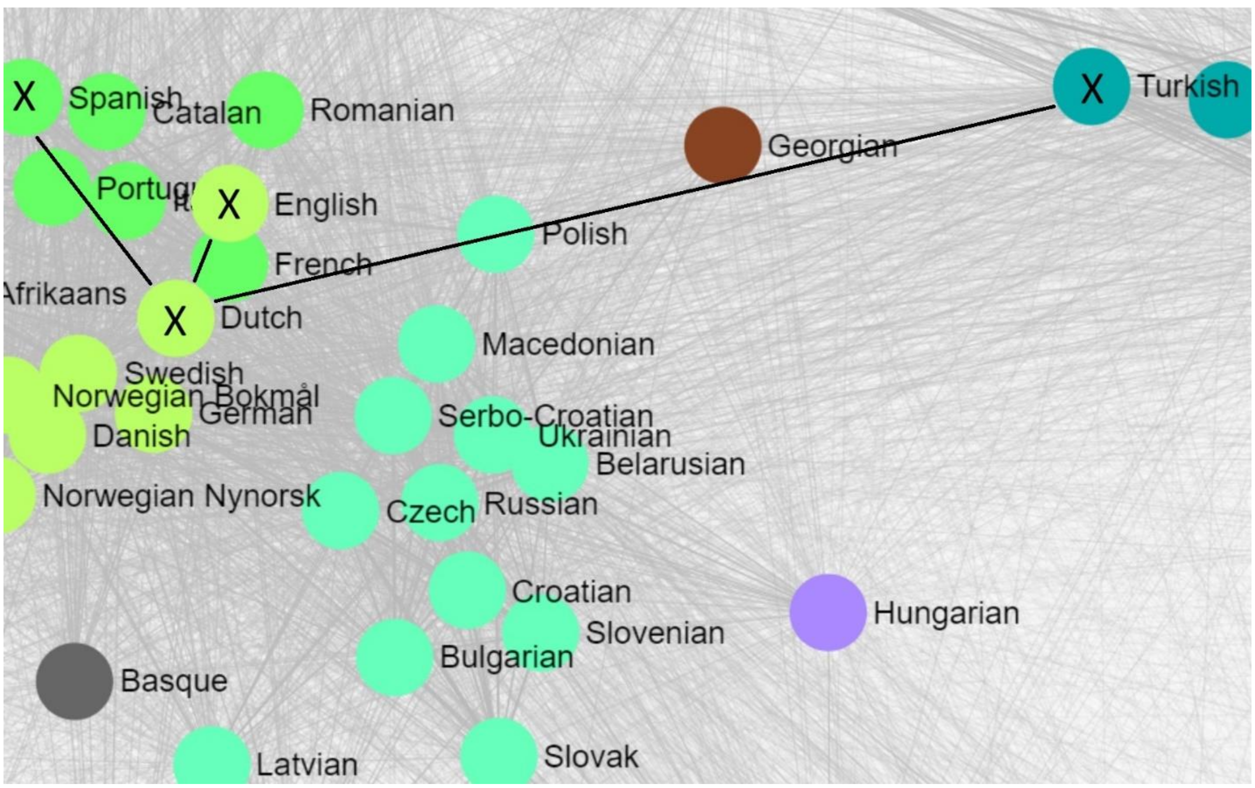
Language dominance was calculated via relative proficiency in both languages. The language pairs were chosen using the available data from the 2in1 project. A language distance classification was created using the Universal Knowledge Core database (Dynamic graphs of lexical similarities, 2021), see Figure 1.

Relative proficiency was measured via the Sentence Repetition Task (SRT) and the Cross-Linguistic Language Task (CLT). The SRT measures proficiency through vocabulary and morphology (Polišenská et al., 2015). Long-term linguistic knowledge has been shown to influence SRT-scores (Armon-Lotem et al., 2015). The CLT is a picture-naming task and measures proficiency via vocabulary capacity (Haman et al., 2015).

A ratio score that represents language dominance for each participant was calculated in the following way. First a percentage score of the right answers for both the SRT and CLT were measured. Secondly, ratio dominance was calculated via this formula: (NLD score / (NLD score + OL1 score)) x 100. Lastly, all scores above 50 were reversed to show the degree of language dominance, with a score of 50 corresponding with being

Figure 1

Dynamic Graph of Lexical Similarities



Note. This graph was retrieved from the Universal Knowledge Core Database on May 21st, 2023 (Dynamic graphs of lexical similarities, 2021). Markings have been added for clarification.

balanced in both languages and a score of 0 corresponding with begin dominant in either language. Before the last step, dominance scores ranged from 0 to 100, with a score of 50 corresponding with being balanced, a score of 0 meaning being dominant in Dutch and a score of 100 meaning being dominant in the other language. Because there was no initial interest in which language children were dominant in, the dominance scores were reformulated. The most convenient way to do this was to leave a score of 50 to mean being balanced and making a score of 0 mean being dominant in either language. For an overview of the formula used in this study, see Table 3.

Cumulative exposure to Dutch and Age of testing were taken from the Bilingual Language Experience Calculator (BILEC) questionnaire of each previous study. The two variables showed a strong, positive correlation (r = .433) and would create multicollinearity if they stayed separate variables in the same analysis (Allen, 1997).

Table 3

Steps of Calculating Language Dominance for the Sentence Repetition task (SRT) and the Cross-Linguistic Language Task (CLT)

	Sentence Repetition task (SRT)	**Cross-Linguistic Language Task (CLT)**
Step 1	$\frac{SRT\ NLD\ correct}{SRT\ NLD\ total}\ x\ 100$	$\frac{CLT\ NLD\ correct}{CLT\ NLD\ total}\ x\ 100$
Step 2	$\frac{SRT\ OL1\ correct}{SRT\ OL1\ total}\ x\ 100$	$\frac{CLT\ OL1\ correct}{CLT\ OL1\ total}\ x\ 100$
Step 3	$\frac{NLD\ CLT\ \%}{(NLD\ CLT + OL\ CLT)}\ x\ 100$	$\frac{NLD\ CLT\ \%}{(NLD\ CLT + OL\ CLT)}\ x\ 100$
Step 4	Scores between 50 – 100 are reversed. Scores between 0 – 50 remain the same.	Scores between 50 – 100 are reversed. Scores between 0 – 50 remain the same.

Therefore, cumulative exposure was combined with the age of testing and turned into a new, single variable: Relative Exposure, through the following formula: (Cumulative exposure to Dutch / Age of testing) x 100. Language distance is based on relative proficiency in both languages (Unsworth, 2018) via two separate percentage scores: CLT and SRT.

2.3 Data Analysis

The dataset of this current study contained the following variables: Participant number, Age of testing, Cumulative exposure, Relative exposure, Language distance (CLP/DLP), type of other language (ENG/TUR/SPA), Dutch SRT-score, other language SRT-score, SRT Dominance, Dutch CLT-score, other language CLT-score, and CLT Dominance. Three Univariate ANCOVA's were done to answer our questions.

The first analysis contained CLT-dominance as dependent variable, language distance as between subject-factor and relative exposure as covariate. The second analysis contained the same independent variables, but SRT-dominance as dependent variable. The third analysis used only the two types of DLP bilinguals. This analysis contained CLT-dominance as dependent variable, language distance (TUR/SPA) as between subject-factor and relative exposure as covariate. All analyses examined main effects of each independent variable and the interaction effect of language distance and relative exposure.

3. Results

3.1 Analysis 1: Cross-Linguistic Lexical Task

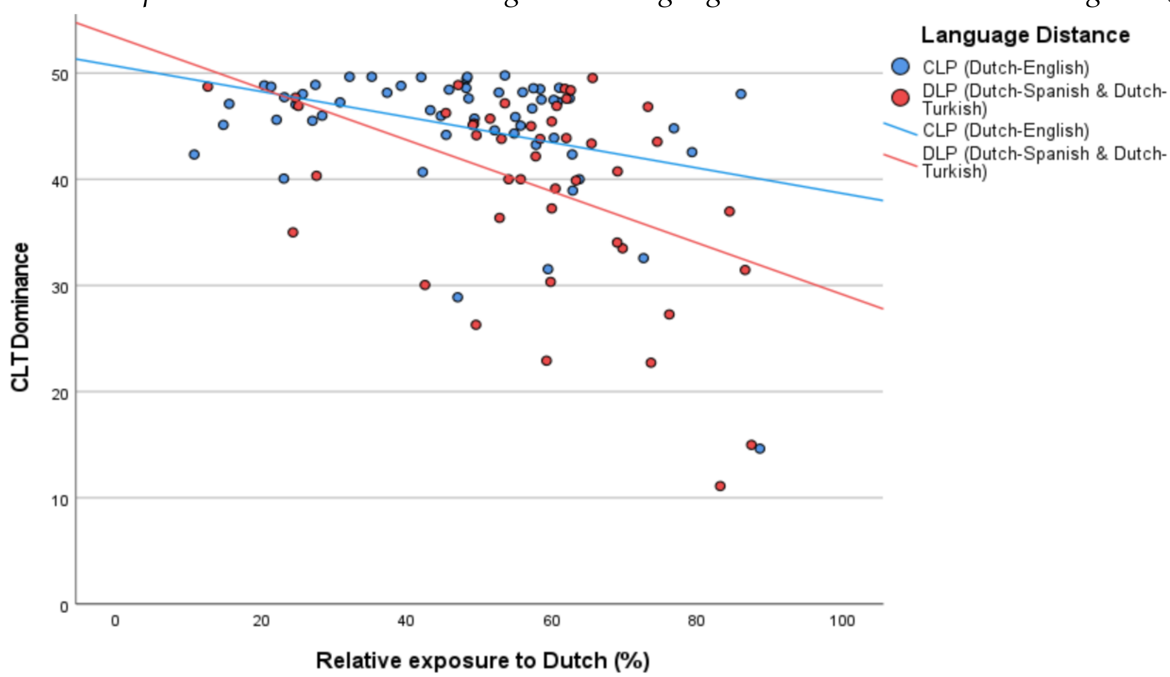
The ANCOVA showed a moderate significant effect for the relationship between relative exposure and language dominance ($F(101,1) = 19.151$; $p<.001$; $R^2 = .165$). There was no significant effect of language distance on language dominance ($F(101,1) = 0.346$; $p =.558$) The CLT-dominance scores did not differ between the CLP-group and DLP-group. This means that language distance did not seem to have an effect on how dominant children scored in either language. Furthermore, there is no interaction between relative exposure and language dominance, which can be seen in Figure 2. This figure shows that higher levels of exposure to Dutch coincide with a greater distance of CLT-dominance scores between the CLP-group and DLP-group. However, this increase in difference is non-significant.

3.2 Analysis 2: Sentence Repetition task

The analysis revealed a significant effect for the relationship between relative exposure and language dominance ($F(102,1) = 8.527$; $p = .004$). This effect was weak ($R^2 = .080$). There was a marginally significant effect from language distance on language dominance ($F(102,1) = 2.935$; $p = .090$). A significant effect of language distance on the relationship between relative exposure and language dominance was found ($F(102,1) = 4.274$; $p = .041$).

Figure 2

Relative Exposure to Dutch in Percentages and Language Dominance Scores in Bilinguals (CLT)



Note. This graph shows the relation between the relative exposure in Dutch and the language dominance scores. Scores closer to 50 mean that the NLD and OTL1 CLT-scores are similar, which reflects balanced bilingualism. Scores closer to 0 mean that there is a large discrepancy between the NLD and OTL1 CLT-scores, which means participants are dominant in one of two languages.

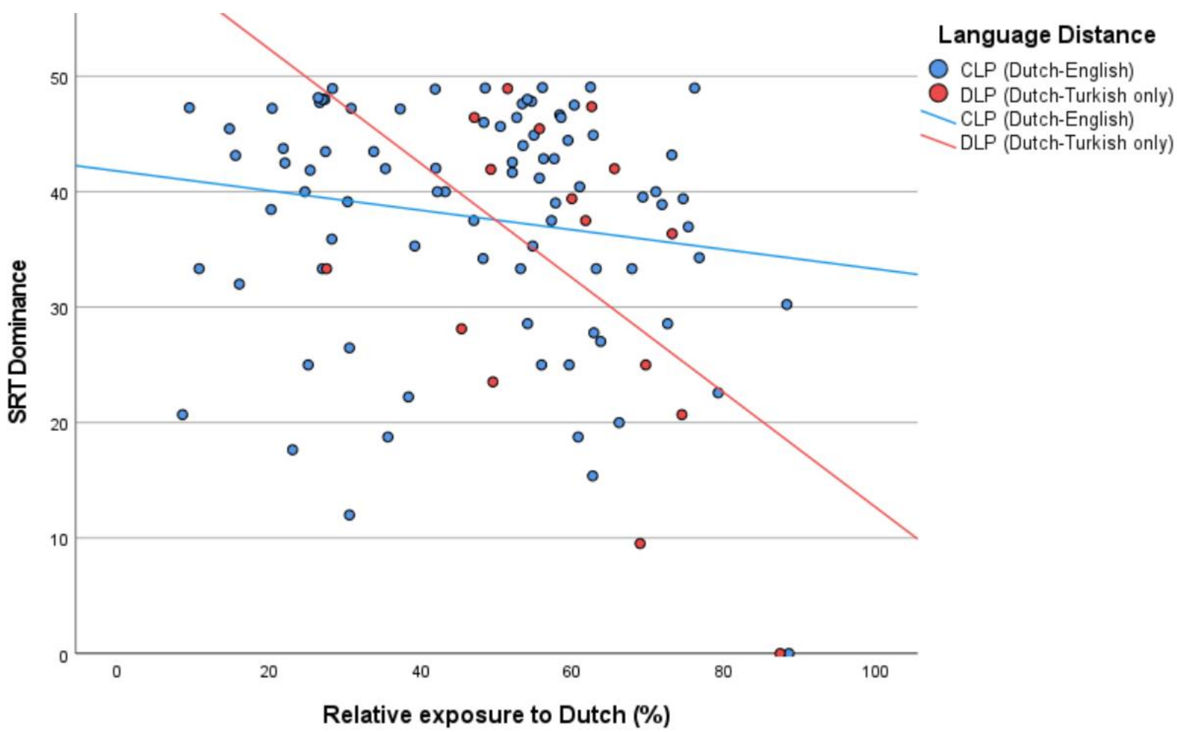
This effect was weak (R^2 = .042). There is a difference between the CLP-group and DLP-group in how SRT-dominance scores change as the exposure as the exposure to Dutch increases. The specific distribution of dominance scores can be seen in Figure 3. This figure shows that as exposure to Dutch increases, DLP-bilinguals become significantly more dominant in either language, whereas CLP-bilinguals stay more or less balanced, regardless of exposure to Dutch.

3.3 Analysis 3: Cross-Linguistic Lexical Task; DLP-group only

The ANCOVA revealed a significant effect of the relationship between relative exposure and language dominance with a moderate effect size (F(45,1) = 7.765; p = .008; R^2 = .159).

Figure 3

Relative Exposure to Dutch in Percentages and Language Dominance Scores in Bilinguals (SRT)

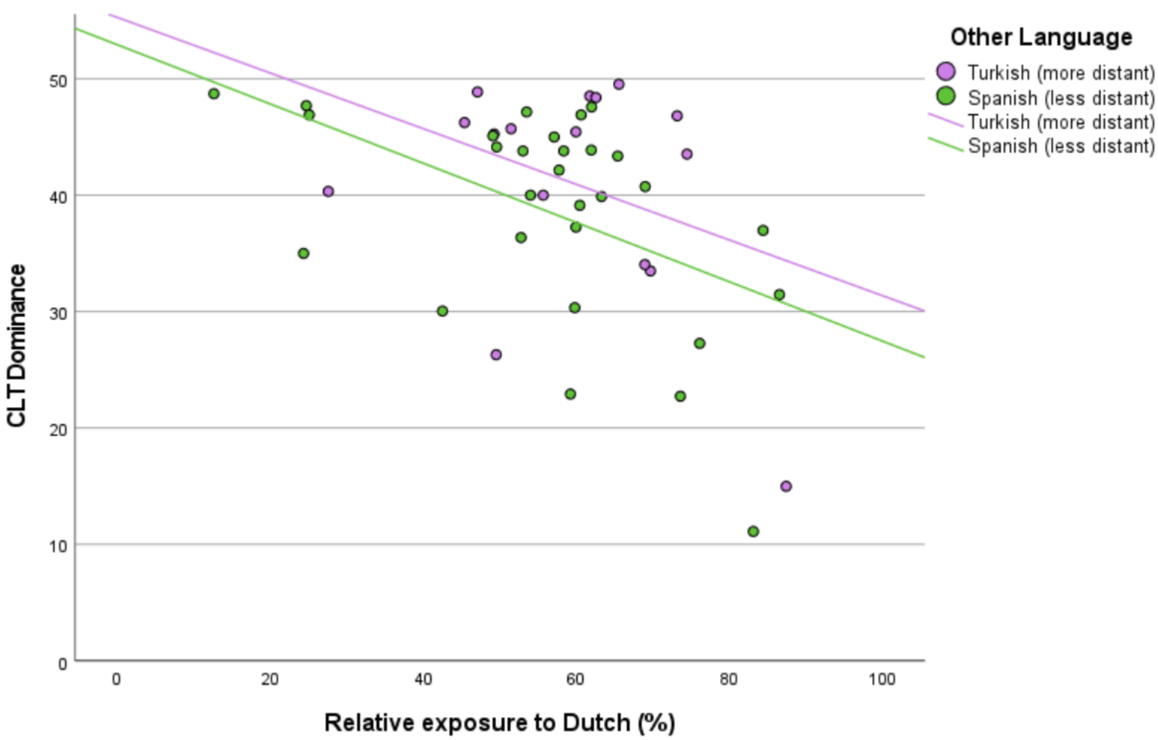


Note. This graph shows the relation between the relative exposure in Dutch and the language dominance scores. Scores closer to 50 mean that the NLD and OTL1 SRT-scores are similar, which reflects balanced bilingualism. Scores closer to 0 mean that there is a large discrepancy between the NLD and OTL1 SRT-scores, which means participants are dominant in one of two languages.

There was no significant effect of language distance on language dominance, when only looking within the DLP-group ($F(45,1) = 0.48$; $p = .827$). This means that between the two types of bilinguals within the DLP-group there seem to be no significant difference in CLT-dominance scores. The interaction between language distance and relative exposure was also non-significant ($F(45,1) = .008$; $p = .930$). The distribution of CLT-dominance scores can be seen in Figure 4. This figure shows that the changes in CLT-dominance as exposure to Dutch increases are quite similar between the Turkish bilinguals and Spanish bilinguals.

Figure 4

Relative Exposure to Dutch in Percentages and Language Dominance Scores in Distant Language Pairs (CLT)



Note. This graph shows the relation between the relative exposure in Dutch and the language dominance scores of the Dutch – Turkish and Dutch – Spanish participants. Scores closer to 50 mean that the NLD and

TUR/SPA CLT-scores are similar, which reflects balanced bilingualism. Scores closer to 0 mean that there is a large discrepancy between the NLD and TUR/SPA CLT-scores, which means participants are dominant in one of two languages.

4. Discussion and Conclusion

It was expected that language distance has a positive effect on the relationship between exposure and language dominance, specifically that as exposure to Dutch increases, DLP bilinguals become dominant in one language and CLP bilinguals stay more or less balanced. It was also expected that within the DLP-group, more distance between languages means a stronger effect compared to less distance.

These expectations were partly borne out. The interaction-effect of relative exposure and language distance was only significant on SRT dominance, but not CLT dominance. Relative exposure was significantly correlated to language dominance for both dominance measures, as expected. The analysis of SRT-dominance revealed that DLP bilinguals showed increased language dominance scores as exposure to Dutch increased. This was not true for the CLP-group as their dominance scores remained more or less similar, even when exposure to Dutch increased. These findings do support the idea that cross-linguistic influence may cause CLP language systems to influence each other and therefor create equal language scores in both languages and that this is not the case with DLP language systems (Floccia et al., 2018).

Closer examination revealed that CLP bilinguals have a more varied dominance pattern than DLP bilinguals, with individuals ranging from being dominant in Dutch to being dominant in English, whereas DLP bilinguals have dominance scores ranging between being balanced to being Dutch dominant. These patterns do not match exposure patterns shown by Knopp (2022), who found opposing results. These patterns do, however, support findings by Thordardottir (2011; 2019) and Floccia et al. (2018), who both found that higher amounts of exposure in one language leads to higher language scores in the same language.

One explanation of these findings is that the sample sizes of the CLP and DLP groups are disproportionate. Within the SRT-analysis, the sample size of the DLP-group only consisted of the Dutch-Turkish bilinguals (N = 19), as the SRT was not available for the Dutch-Spanish bilinguals. Compared to the sample size of the CLP-group (N = 93), this is quite small. A small sample size usually holds the limitation of the type II error

(Hackshaw, 2008), however this seems not to be the case, as the interaction effect was significant.

Another possible explanation for why only the SRT offers a significant result, is that the SRT and CLT measure different domains of language proficiency. The SRT is made to measure proficiency through vocabulary and morphology, whereas the CLT only measures vocabulary through picture-naming. Even though vocabulary size correlates with other aspects that are important for proficiency, such as grammatical ability, it is not as complete of a measure for proficiency as the SRT (Miralpeix & Muñoz, 2018; Van Wonderen & Unsworth, 2020). Furthermore, CLT scores in different languages do not lend themselves well to direct comparison, which is a limitation in this research project (Van Wonderen & Unsworth, 2020). As Birdsong (2016) states, language dominance is relative, multidimensional and gradient, meaning that there are many aspects to language dominance. The addition of morphology in the SRT in combination with the differences in methodology of both tests, may explain why the results differ between the two tests.

Within the DLP-group, no significant effect was found of language distance on the relationship between relative exposure and language dominance. The change in degree of dominance as exposure to Dutch increased was not different among the DLP-group. Dominance patterns of CLT-dominance revealed that the variation of both the Dutch-Turkish and Dutch-Spanish bilinguals were similar to each other, as they both ranged from being balanced to being Dutch-dominant.

These findings raise several questions that could be investigated in future research. First, it would be interesting if future research also includes other language domains that measure proficiency, as the differences in scores of the SRT and CLT could be due to the contents of the tests. In short, the SRT measures proficiency through vocabulary and morphology (Polišenská et al., 2015) and the CLT measures proficiency through vocabulary capacity only (Haman et al., 2015). Secondly, future research should investigate the boundary regarding language distance. This study found hardly any significant differences between the DLP and CLP groups, while other studies did find significant differences between them, such as Blom et al. (2020). Thirdly, it would be interesting to also zoom in on the CLP group by comparing two closely related language pairs. While this study only focused on the degree of variability within the DLP-group, a closer look at the degree of variability among CLP-bilinguals could reveal a broader pattern of language distance related to dominance in bilinguals. Lastly, as mentioned above, it is interesting that the small size of the DLP-group within the SRT-analysis yielded a significant result for the interaction between language distance and relative exposure on language dominance. Future research could go into depth to what happens if both the DLP-group and CLP-group are similar in size.

Concluding, the interaction between relative exposure and language dominance seems to be present, further affirming the results found by Thordardottir (2011, 2019) and Floccia et al. (2018), who found that exposure to L1 increases dominance for L1 while decreasing dominance for L2. However, the effects of the interaction between relative exposure and language distance on language dominance is not present in all measures of proficiency. Moreover, these results are only present in the comparison between CLP bilinguals and DLP bilinguals, and not when looking at two different language pairs within the DLP-group.

Acknowledgements

We would like to thank dr. Sharon Unsworth for her permission to use the dataset from the 2in1 project. Furthermore, we would like to thank her for her constructive criticism which allowed us to improve our paper.

Context of research

This research was conducted as part of the masters course Child Bilingualism (LEX-REMA-LCEX27).

References

Allen, M. P. (1997). The problem of multicollinearity. In *Understanding Regression Analysis* (pp. 176–180). Springer, Boston, MA. https://doi.org/10.1007/978-0-585-25657-3_37

Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode?. *Journal of phonetics*, 40(4), 582–594. https://doi.org/10.1016/j.wocn.2012.05.005

Armon-Lotem, S., De Jong, J., & Meir, N. (2015). Assessing multilingual children: disentangling bilingualism from language impairment. In *Multilingual Matters eBooks* (Issue 13). https://ci.nii.ac.jp/ncid/BB19110548

Blom, E., Boerma, T., Bosma, E., Cornips, L., van den Heuij, K., & Timmermeister, M. (2020). Cross-language distance influences receptive vocabulary outcomes of bilingual children. *First Language, 40*(2), 151–171. https://doi.org/10.1177/0142723719892794

Floccia, C., Sambrook, T. D., Delle Luche, C., Kwok, R., Goslin, J., White, L., Cattani, A., Sullivan, E., Abbot–Smith, K., Krott, A., Mills, D., Rowland, C., Gervain, J., Plunkett, K., Hoff, E., Core, C., & Patricia, J. (2018). Vocabulary of 2-year-olds learning English and an additional language: Norms and effects of linguistic distance. *Monographs of the Society for Research in Child Development*, 83(1), 7–29. https://doi.org/10.1111/mono.12350

Hackshaw, A. (2008). Small studies: strengths and limitations. *The European Respiratory Journal*, *32*(5), 1141–1143. https://doi.org/10.1183/09031936.00136408

Knopp, E. (2022). Effects of dominance in xposure and linguistic distance in the bilingual proficiency profiles of Dutch-German and Greek-German school-age bilinguals. *Journal of Home Language Research,* 5(1): 2, pp.1–20. https://doi.org/10.16993/jhlr.42

Liu, Y., & Ning, J. (2021). The Effect of Language Dominance on the Selective Attention of Segments and Tones in Urdu-Cantonese Speakers. *Frontiers in psychology, 12,* 710713. https://doi.org/10.3389/fpsyg.2021.710713

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching*, 56(1), 1–24. https://doi.org/10.1515/iral-2017-0016

Polišenská, K., Chiat, S., & Roy, P. (2015). Sentence repetition: what does the task measure? International Journal of Language & Communication Disorders, *50*(1), 106–118. https://doi.org/10.1111/1460-6984.12126

Radman, N., Jost, L., Dorood, S., Mancini, C., & Annoni, J. M. (2021). Language distance modulates cognitive control in bilinguals. *Scientific reports*, *11*(1), 24131. https://doi.org/10.1038/s41598-021-02973-x

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism, 15(*4), 426–445. https://doi.org/10.1177/1367006911403202

Thordardottir, E. (2019). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism, 23*(1), 236–255. https://doi.org/10.1177/1367006917722418

Universal Knowledge Core (n.d.). *Lexical Data Visualisations*. Universal Knowledge Core. Retrieved on May 21st, 2023, from http://ukc.datascientia.eu/lexdist

Unsworth, S., Chondrogianni, V., & Skarabela, B. (2018). Experiential Measures Can Be Used as a Proxy for Language Dominance in Bilingual Language Acquisition Research. *Frontiers in psychology*, *9*, 1809. https://doi.org/10.3389/fpsyg.2018.01809

Unsworth, S., Kootstra, G. J., Van Dijk, C., Karaca, F., Koutamanis, E., Van Bosch, J., Van Wonderen, E., & Van Zwet, J. (2022). 2In1: Bilingualism in children [Dataset]. In *2 in 1 research project*.

Van Wonderen, E., & Unsworth, S. (2020). Testing the validity of the Cross-Linguistic Lexical Task as a measure of language proficiency in bilingual children. *Journal of Child Language*, *48*(6), 1101–1125. https://doi.org/10.1017/s030500092000063x

What are the Teachers' and Students' Attitudes Towards English as a Medium of Instruction Within the European Higher Education Area?

Literature review

Athanasios Athanasiou

Research Master *Language and Communication Sciences*, Radboud University Nijmegen

Abstract

Several universities are establishing study-programs taught in English (ETPs) even when the local language is not English, an initiative that has sparked controversy. The sudden increase in these programs has been heavily criticized, as there are concerns about a potential lack of planning and a disregard for teachers' and students' views. The present review looked at the studies examining the teachers' and students' attitudes regarding these English-taught study-programs within the European Higher Education Area aiming to put their opinions at the forefront of relevant research. The current literature review indicated that teachers and students' acknowledged the importance of *ETP*s for boosting their mobility and employability prospects, even though they agreed about the lack of a shared official university policy. Furthermore, both groups admitted having language-related difficulties, also claiming that teacher training was neglected. The importance of these findings is undeniable for policymakers, teachers, and students. Crucial structural weaknesses of the *ETP*s were revealed, with further research considered vital for delving deeper into teachers' and students' struggles.

Keywords: *EMI*, Europe, attitudes, students, teachers

1. Introduction

Within the broader framework of a European multilingualism policy (Eurydice, 2006), many countries are incorporating English as the language to be used in tertiary education. Such a policy was based on participating in the Bologna Process, a series of multi-national educational reforms initiated by the Bologna Declaration in 1999 (Dafouz et al., 2013). Many countries either members of the European Union or geographically located in the European continent formed a barrier-free European Higher Education Area (*EHEA*) by signing the Bologna Declaration, aiming to ensure homogeneity across the European universities. The goal was also to provide people with the 'varied lenses needed to capture, understand and create our global reality' (Skutnabb-Kangas et al., 1995, p. 224–225) and to educate multilingual, multiliterate, and multicultural professionals able to work and communicate in a globalized world (García & Beardsmore, 2013; Skutnabb-Kangas 1995). To this end, European universities encouraged scholars and students from other continents to enroll in European universities (Bolsmann & Miller, 2008; Papatsiba, 2006), using English as the Medium of Instruction (*EMI*) to facilitate mobility. Using the definition provided by Macaro et al. (2018), *EMI* is defined as:

> "the use of the English language to teach academic subjects (other than English itself) in countries or jurisdictions where the first language (*L1*) of the majority of the population is not English" (p. 37).

EMI has been found to improve students' English competence (Dalton-Puffer, 2011; Yang, 2016), foster learners' mobility and employability (Wächter & Maiworm, 2014), and is believed to be a powerful motivator factor for English learning (Doiz et al., 2014; Yang, 2016). Furthermore, this expansion of English in tertiary education can be tangibly reflected in the number of English-taught programs (*ETP*s) in the *EHEA*, which, according to the study by Wächter and Maiworm (2014) skyrocketed to 8,089 in 2014. This signalled a 239% increase compared to their previous study in 2007, although *ETP*s with partial *EMI* were excluded from the recent research design.

It is noteworthy that such a widespread domination of English can also be attributed to socioeconomic and political forces stemming from globalization which have boosted this trend (Kuteeva, 2018). More specifically, according to Wilkinson (2012), the popularity of EMI is due to multiple macro factors such as the economy, society, politics, and education. In a broader context, this acceleration could also be linked to globalization, the incessant flow of resources, higher education's ambition of becoming international, and the growing international, dominant status of English

(Hu, 2019; Macaro 2018).

However, the vagueness surrounding the existence of a common official policy has created a heterogeneous sample within the *EHEA*. In other words, Nordic countries have officially adopted the parallel language use policy (Dimova, 2017) allowing teachers to use the local language when necessary. Even the admission requirements are not the same in all countries of the *EHEA*, with Turkey constituting a prominent exception, as universities require students to be part of a preparatory year *(PYP*) and pass a language test before being admitted to an *ETP* (Ekoç, 2020).

This complex situation in the *EHEA* illustrates that language policies are adopted without careful consideration employing top-down and not bottom-up approaches (Macaro et al., 2018). This means that policies are established by policymakers and (inter)national organizations, with key actors in teaching and learning not always being consulted (Dearden & Macaro, 2016), which contradicts the need for *EMI* programs to be carefully planned (Lasagabaster et al., 2014). However, revealing attitudes towards *EMI* and understanding what teachers think and believe (Brown, 2016; Simbolon, 2018) can be considered a precondition for the long-term success of any language policy Moreover, attitudes towards language, from a medium of instruction perspective, can considerably influence students' academic achievements and career opportunities (Garrett et al., 2006), and should be taken into account for the development of language planning and policy at universities.

Striving to ensure teachers' language competence, some institutions have established certification mechanisms, adapted to the specific university context, which assess proficiency in the language of instruction. Prominent examples of such practices include the Test of Oral English Proficiency for Academic Staff (*TOEPAS*) Certification at Copenhagen University (Dimova & Kling, 2018), *HELA* (Higher Education Lecturing Accreditation) (Álvarez, 2014) at the University of Vigo. Furthermore, in a survey of 79 Higher Education Institutions (*HEI*s) across Europe, more than 60% of them stated that they were already providing training courses lasting from one to 60 hours or, in some cases, even longer (O'Dowd, 2018). However, O'Dowd (2018) stated that "the training of teachers in *EMI* is far from being treated as an important issue in European university education" (p. 557).

Regarding students, the same anxieties about the reliability of standardized language tests, including *IELTS* and *TOEFL*, are voiced. Researchers are questioning the adequacy of these tests as an admission requirement (Gundermann, 2014), believing that they promote native-speaker norms (Saarinen & Nikula, 2013), and recommending implementation of post-entry screening procedures to identify unprepared students who need language support (Wilkinson et al., 2006). The success of *EMI* is also questioned, because of the inadequate levels of students' and faculty's English language proficiency (Macaro, 2018; West & Aşık, 2015). Hence, it becomes crucial to understand students' views and check whether the aforementioned concerns are fathomed.

Overall, the lack of a shared official policy, the concerns about teachers' ability to cope with this new reality, and the doubts raised about the validity of language tests for students make it imperative that teachers' and students' attitudes be investigated. Macaro et al. (2018) also explained that teachers' and students' attitudes is an issue not emphasized enough.

In a similar vein, most studies have focused on *EMI* in universities that have a history of teaching through English and in countries where English is a more "integrated" foreign language, such as Finland (Mauranen, 2006), Sweden (Söderlundh, 2013), and The Netherlands (Wilkinson, 2013). However, the South European countries are also represented in *EMI* studies, portraying a different picture. The Italian context, for instance, is quite different, as the teaching of English has only recently begun throughout primary and secondary education, and it is not commonly used outside school. Therefore, the present systematic review highlights these issues by presenting teachers' and students' attitudes in countries within the full spectrum of the *EHEA*. In short, the following research question was investigated:

***RQ**: What are the teachers' and students' attitudes towards EMI within the EHEA?*

Considering the vague nature of the term attitudes, we decided to focus only on papers investigating teachers and students' views pertaining to the presence of an official *EMI* policy, the benefits of *EMI,* comments about the admission requirements and teacher training programs. The reason being that these constitute recurrent themes in the current *EMI* literature.

2. Methods

2.1 Selection Criteria

The goal of this review is to focus on studies examining students' and teachers' attitudes toward *EMI*, with certain criteria being established to ensure comparability. More specifically, the focus shifted only to the *EHEA*, thus guaranteeing a relatively homogeneous sample. This constituted an objective inclusion criterion, as countries, including Russia and Turkey, are notoriously difficult to group as either European or Asian. Furthermore, African and Latin American countries were excluded, as they were underrepresented (Macaro et al., 2018), while Asian countries were not included due to the complicated role of the English language and them not belonging to the *EHEA*.

On top of these, the next inclusion criteria referred to the year of publication and the accepted sample size. To begin with, Macaro et al.'s paper (2018) suggested that the number of *EMI* programs initially increased from 2006 to 2010 before skyrocketing in 2010. Therefore, the papers' accepted publication dates ranged from the first day of the year 2010 until the end of October 2022, when the reviewing process commenced. Specifying the range for publication dates allowed us to narrow the scope to the most relevant and recent papers, avoiding older publications, which would not have reflected the current academic reality. Regarding the sample size, studies were incorporated in the present review, provided that the sample size was equal to or higher than 20. The aim was to have samples that had sufficient statistical power to uncover significant effects, hence aiding us in reaching robust conclusions.

Overall, the final list of papers abided by the selection criteria and focused on teachers' and students' attitudes toward *EMI*, aiming to verify or disprove the initial hypotheses. Also noteworthy, though, is that the use of English was associated with ICT tools and blended learning being used in *EMI*. However, reporting on this issue would be only indirectly linked with how teachers and students perceive the use of English during the lectures, thus constituting a by-product of *EMI* implementation. In a similar vein, papers using students' grades as a proxy for the efficiency of *EMI* were also eliminated, due to the existence of confounding factors, including individual differences, subjective grading criteria, and different grading scales, influencing performance. Finally, partial *EMI* programs were also rejected, as they were not considered entirely informative.

2.2 Compiling Literature

To find relevant materials for this review, reliable tools were used. More specifically, Web of Science (*WoS*), a highly reputable database, was chosen because of its peer-review system, which filters out material that is not peer-reviewed and could hence be of lower quality. This aspect constitutes an advantage over Google Scholar, as the latter includes more papers but of lower quality.

Regarding the search in the database, different keywords were used to find the most representative terms. Starting from a broad search of the phenomenon of *EMI*, I then proceeded to alternate between the abbreviation and the full name of

EMI, before seeking papers about teachers' and students' attitudes within the *EHEA* (Table 1). However, noticing that the inclusion of both "attitudes" and "Europe" in *WoS* yielded an insufficient number of papers, meant that an artificial method of

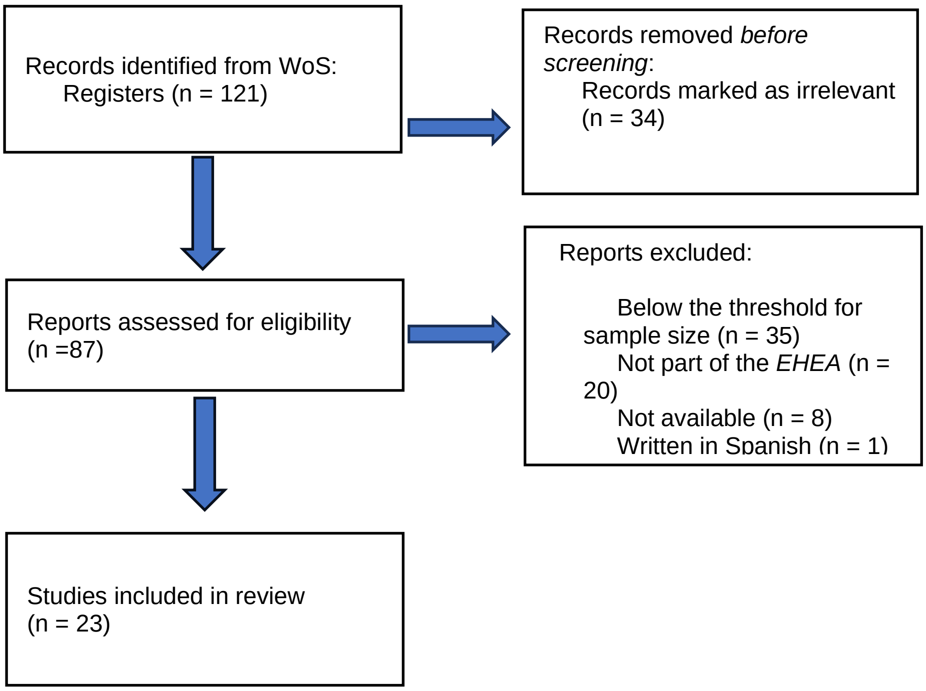
Table 1

The Search Terms Used and the Number of Results They Yielded

Search Terms (WoS)	Number of results
EMI	23,818
English as a Medium of Instruction	4,331
EMI Higher Education	2,310
English as a Medium of Instruction Higher Education	1,412
English as a Medium of Instruction attitudes	429
EMI attitudes	206
English as a Medium of Instruction in European Higher Education	121
English as a Medium of Instruction in European Higher Education attitudes	15
EMI in European Higher Education attitudes	10

Figure 1

The Screening Process Regarding the Inclusion of Relevant Papers Under "EMI attitudes".

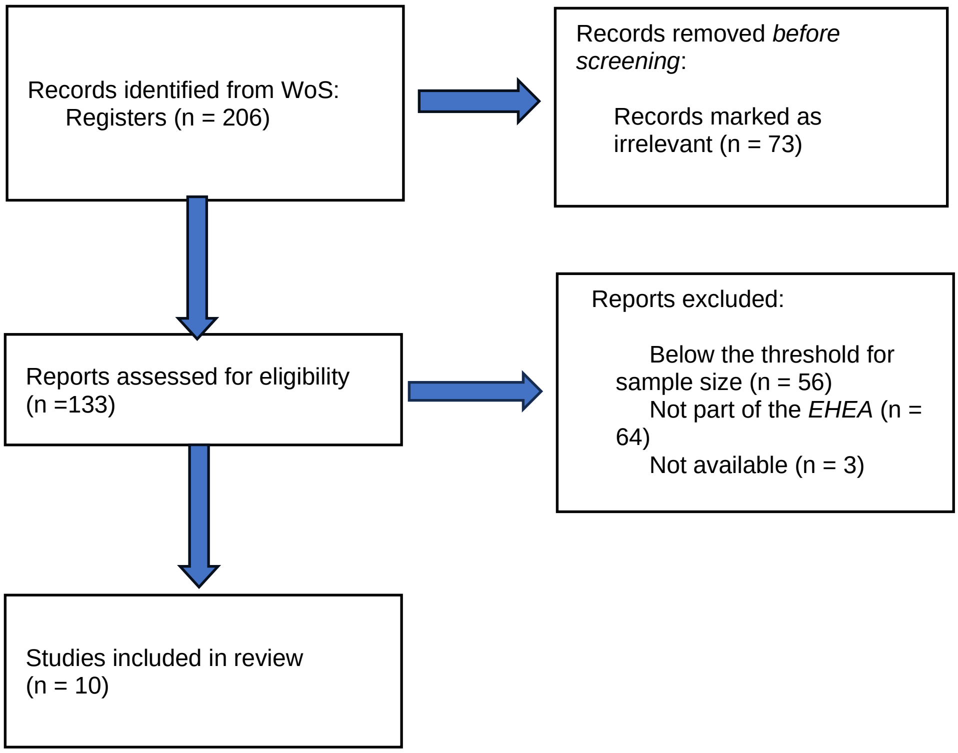


combining them was opted for. Therefore, emphasis was put on the last two keyword entries, as one focused on "EMI attitudes" yielding 206 results, while the other referred to "English as a medium of Instruction in European Higher Education" providing 121 papers for the set timing frame for publication. The latter offered ten papers, whereas the former gave five more, thus adding up to fifteen studies in total.

This number was the result of meticulous abstract screening processes to ensure that the papers chosen abided by the set selection criteria. To elaborate, Figure 1 illustrates that out of the 121 results that the search for "English as a Medium of Instruction in European Higher Education" yielded, 35 were excluded due to either limited sample size and/or content-related issues, as they could be assessing *EMI* from the perspective of judging nonnative accent, measuring performance or even referring to Content Language Integrated Learning *(CLIL)*. Another 20 studies were not describing *EMI* within the *EHEA*, 34 were not relevant, eight were not available and one was written in Spanish. From the remaining studies, ten were selected for the results section and the rest were incorporated in the Introduction or the Discussion, as they provided a theoretical framework for *EMI* research.

Figure 2

The Screening Process Regarding the Inclusion of Relevant Papers Under "English as a Medium of Instruction in European Higher Education".



Furthermore, when searching for "EMI attitudes", 73 papers were discarded as irrelevant, 64 were reporting findings on studies outside the *EHEA*, three were unavailable, and 56 were related but violated the set selection criteria (Figure 2).

3. Results

3.1 EMI Policy

Before presenting teachers' and students' attitudes towards an *EMI* policy, we should distinguish between the two aspects of policy. To elaborate, it could refer to either the methods used by universities to attract teachers and students (internationalization policy) or the method of implementing *EMI* (pedagogical policy).

Starting with the first definition of policy, data from Spain (Aguilar, 2017) and Italy, Austria, and Poland (Dearden & Macaro, 2016) indicated that teachers were unaware of any official *EMI* policy enforced by their universities. They claimed that *EMI*'s growth was not the product of meticulous planning", a problem also evident with the lack of support and set criteria for selecting *EMI* teachers.

Even in the presence of university policies, the official documents of 10

European Universities (Orduna-Nocito & Sánchez-García, 2022) did not reflect teachers' views, regarding a threshold for English language proficiency. The established policy, as reflected by those documents, also neglected the educational support that teachers thought they needed as well as the communicative purpose of *EMI*.

Pertaining to a pedagogical *EMI* policy, it was absent from official documents, with teachers often resorting to teaching strategies that are not part of *EMI*. These included the parallel language use, that is the alternation of the local language and English during lectures, in countries such as Spain and Sweden (Orduna-Nocito & Sánchez-García, 2022), Denmark (Dimova, 2020), and Turkey (Ekoç, 2020). This was attributed to teachers' inadequate English proficiency, and the extent of its implementation varied depending on the discipline. More specifically, Science, Technology, Engineering and Mathematics (STEM) teachers at a Spanish university (33.3%) admitted to allowing the use of L1 more than their counterparts in the Humanities (21.7%), with the latter responding that the use of the local language depended on other factors at a higher percentage (30.4 % vs 16.7%) (Roothooft, 2019). Finally, the lack of an official policy also influenced teachers' teaching goals. A study in the Republic of Macedonia showed that teachers in the Language Center of the National University emphasized the communicative aspect of English, whereas those teaching in the English department aimed to raise awareness towards native norms (Agai-Lochi, 2015).

3.2 Goal of EMI

Both students and teachers highlighted the importance of *EMI* for mobility and employability. To elaborate, Engineering lecturers in Spain explained that English is the language of science and the tool that increases employability (Aguilar, 2017), with teachers and students from another Spanish university (Mira et al., 2021) and teachers at a Turkish university also supporting these ideas (Ozer, 2020).

However, despite teachers supporting *EMI*, they also argued that universities and policymakers have a completely different goal in mind. In other words, they claimed that universities implemented *EMI* to compete internationally (Dearden & Macaro, 2016) and/or to increase their revenue (Orduna-Nocito & Sánchez-García, 2022). Such a concern was even voiced by students in Catalonia (Sabaté-Dalmau, 2016) who despite acknowledging the value of *EMI*, were, nevertheless, suspicious about the political and financial forces supporting it, calling for measures to safeguard minority languages and linguistic diversity. All in all, the acceptance of

EMI was not complete, as students and teachers stressed its positive impact, but also raised concerns.

3.3 English Language Proficiency

The issue of English language proficiency was approached from two perspectives, namely participants' language skills and the existence of a threshold allowing participation in *EMI* programs.

Starting from students in Turkey, it was found that only 52% were able to meet the language requirements on their first try, while 6.2% needed four attempts to surpass the threshold (Ekoç, 2020). In view of this, students suggested the establishment of English for Academic Purposes (*EAP*) tests to ensure that they were qualified to follow lectures in an *EMI* context, as only 21.8% of the participants were satisfied with the current *EMI* reality. Teachers from another university in Turkey also underlined the low level of students' proficiency in English as a major problem at 81.4% (Ozer, 2020). The same concern was voiced by Danish teachers, claiming that local students had high general proficiency in English, but they were lacking academic proficiency (Dimova, 2020), which constituted a barrier during lectures and led to the parallel language use policy. Difficulties in coping with English were also in the form of students suffering from communicative anxiety as a corollary of a lack of confidence when judging their linguistic competence.The results of a study in the Basque country (Santos et al., 2018) showed that there is a correlation between these levels of stress with gender and the academic discipline of students, with female students studying Business exhibiting a higher amount of anxiety than those majoring in Education. Another interesting finding regarding students was that those in Catalonia supported teaching and tests that promoted native-like norms and prioritized accuracy, even though they struggled with English (Sabaté-Dalmau, 2016).

Moving on to teachers, it was discovered that they were also facing certain difficulties. Lecturers in Italy claimed that teaching in English threatened their skill to improvise during lectures, worrying that their potentially inadequate language proficiency would be assessed negatively by students, particularly during communication in informal contexts (Helm & Guarda, 2015). In another study in Denmark, teachers admitted that not everyone was adequately prepared for *EMI* (Jensen & Thøgersen, 2011), while teachers from 10 European Higher Education Institutions, suggested that it was not easy to assess students' language proficiency (Orduna-Nocito & Sánchez-García, 2022).

Regarding the existence of a benchmark for proficiency, the selected studies illustrated the absence of an established threshold. To elucidate, Engineering lecturers defined approximately C1 for teachers and B2 for students, as the required language level proving adequate competence (Aguilar, 2017). Students in Turkey also replied that they would have to get 60% on a language test measuring general proficiency to become members of an *EMI* class (Ekoç, 2020). Nonetheless, the fact that they were given four chances to meet the requirements raises questions pertaining to the appropriateness of such a test.

Such a vague threshold was found in other countries as well, which was often the result of an incomprehensive *EMI* policy. In other words, lecturers from Italy, Poland, and Austria could not agree on a specific required language level (Dearden & Macaro, 2016), while their colleagues from 10 other European universities also expressed similar difficulties (Orduna-Nocito & Sánchez-García, 2022). Quite interestingly, despite 61% of Danish teachers finding current tests adequate, they suggested interviewing students and assessing written statements of purpose and students' research experience to ensure that students can follow a lecture in English (Dimova, 2020).

3.4 Teacher Training

Teacher training constituted another neglected aspect of *EMI*. Such a reality was either attributed to the scant offering or even the total absence of educational and language support to teachers.

To begin with, lecturers in Spain (Aguilar, 2017) were favorable towards any language support that the university could offer to them, recognizing the complicated nature of *EMI*. Teachers in Turkey (Ozer, 2020) and Teaching Assistants and Assistant Professors in a Serbian university (Đorđević & Blagojević, 2019) pinpointed the specific domains they needed training including speaking and pronunciation skills (28.4%) and pedagogical training (12.8%). New teaching strategies and terminology for teaching their academic subject also gathered support.

Furthermore, teachers of an Italian University (Helm & Guarda, 2015) as well as those of 10 European *HEI*s (Orduna-Nocito & Sánchez-García, 2022) expressed their desire to receive training pertaining to teaching methods, without explicitly stating their need for support or the specific skill that needed improvement. The importance of training was even highlighted by students in Turkey, who claimed that teachers should be properly trained before being allowed to teach in an *EMI* classroom (Ekoç, 2020).

Despite the training provided by certain universities, though, such as the University of Copenhagen, this was not always appreciated by trainees. To be clearer, *TOEPAS* allowed teachers to engage in a simulated teaching experience, as their performances were recorded and written feedback was given. However, only 54% of the teachers kept the video of their lecture and only 5% watched it claiming that they were too self-conscious (Dimova, 2017), while the written report provided was also overlooked. Nevertheless, they recognized that the training increased their awareness about their teaching methods, strengths, and weaknesses, but it did not change their teaching practices. In contrast, teachers simply inserted the *TOEPAS* certification in their CVs to increase their job prospects.

4. Discussion

This systematic review examined the relevant literature on teachers' and students' attitudes toward *EMI*. The goal was to raise awareness regarding current *EMI* implementation and encourage more research to be conducted, thus constituting a valuable tool for stakeholders, policymakers, teachers, and students engaging with *EMI*.

Firstly, the present review yielded some interesting findings verified our initial hypotheses. To begin with, the absence or unawareness of an official policy confirmed that teachers and students are not always consulted when policymakers establish university policies (Dearden & Macaro, 2016). The positive impact of *EMI* on Higher Education was depicted, although this review revealed teachers' and students' concerns about their language skills. The latter was linked to,the inadequacy of the current tests measuring language proficiency (Macaro et al., 2018) and the need to measure academic English proficiency (Ekoç, 2020). Finally, the overall inadequacy of planning led to its negligence.

A strength of this review refers to the inclusion of countries that are not traditionally *EMI* supporters. To be clearer, it has been hypothesized that South European countries do not have many available *ETP*s compared to the Nordic countries and the Netherlands, with students also having lower English proficiency due to various socio-linguistic factors (Dafouz et al., 2013). Therefore, exploring students' and teachers' opinions contributes to forming a comprehensive account of *EMI* and to assessing the success of the Bologna Declaration (1999) by examining the extent of comparability within the *EHEA*.

However, the limited number of studies included constitutes a shortcoming of the present review. To elucidate, the sample of 15 papers may not lead to robust conclusions, especially if combined with the relatively limited sample size in some of the selected studies. As a corollary, there were no studies on most of the 47 countries within the *EHEA*, while it also seemed difficult to recruit participants. Additionally, the overrepresentation of a specific set of countries is not likely to compensate for the limited quantity. Despite such problems, though, this review simply aims to draw attention to some of the problematic aspects of *EMI* and does not claim to fully explain the topic. In essence, its importance for the future of *EMI* cannot be doubted, as it paves the way for subsequent research to be conducted.

4.1 Implications

The present article focuses on teachers' and students' attitudes in countries within the *EHEA* to gain insights into how *EMI* is understood and implemented. Publishing the results could then have major implications on most aspects of the current *EMI* reality, aiming to improve the teaching and learning process.

To begin with, many universities lacked an official *EMI* policy and thus comparability within the *EHEA* was not ensured. This aggravates potential difficulties and poses an obstacle for changes to be made at an international level. Hence, before modifying parts of the *EMI* experience, a certain degree of comparability should be established by forming international committees, that should create a general international framework for *EMI* implementation. Adhering to these rules would ensure compatibility amongst the *EHEA*, while also allowing for some flexibility due to each country's socio-political context.

As for the students' struggles with English, it is likely to trigger the conversation regarding establishing *EAP* tests that would substitute for the tests measuring general proficiency. In other words, the validity of the traditional *IELTS* and *TOEFL* scores is likely to be questioned, as even when these criteria were met, language-related difficulties were expressed. Hence, these tests may be complemented by other tests or even be substituted by *EAP* cut-off scores. To this end, teachers' and students' arguments should be at the forefront of *EMI* research to apply pressure for changes to be made by policymakers.

Finally, the absence or underestimation of teacher training programs is another problematic part of *EMI*. This stresses the need for training to be established or upgraded, which should be tailored to teachers' needs. Incentives should be offered for completing them, while the difficulties caused due to insufficient training should be highlighted. Emphasis should also be put on lifelong learning and on creating official international accreditation procedures as a requirement for teaching in an *EMI* classroom.

5. Conclusion

To conclude, it is evident that many *HEI*s have increased the number of *ETP*s offered. The present review illustrated that an official university policy was often absent, while the value of English for future employability was appreciated. The absence of a threshold for teachers' and students' English proficiency and the negligence of teacher training were also highlighted. However, the lack of consensus surrounding these crucial aspects of *EMI* is a major problem requiring cooperation amongst all parties involved. Hence, more studies should be conducted, examining students' and teachers' attitudes to gain a greater understanding and apply any necessary changes.

References

Agai-Lochi, E. (2015). English as medium of instruction in university education. *Procedia - Social and Behavioral Sciences*, *199*, 340–347. https://doi.org/10.1016/j.sbspro.2015.07.557

Aguilar, M. (2015). Engineering lecturers' views on Clil and Emi. *International Journal of Bilingual Education and Bilingualism*, *20*(6), 722–735. https://doi.org/10.1080/13670050.2015.1073664

Álvarez, G. D. (2014). HELA (Higher Education Lecturing Accreditation):: la acreditación del profesorado universitario para la docència en inglés en la Universidad de Vigo. In *Nous reptes per als centres de llengües: estàndards de qualitat* (p. 44). essay.

Bolsmann, C., & Miller, H. (2008). International student recruitment to universities in England: Discourse, rationales and globalisation. *Globalisation, Societies and Education*, *6*(1), 75–88. https://doi.org/10.1080/14767720701855634

Brown, H. (2016). Current trends in English-medium instruction at universities in Japan. *CUE*, *10*(1), 3–20.

Dafouz, E., Camacho, M., & Urquia, E. (2013). 'surely they can't do as well': A comparison of business students' academic performance in English-medium and Spanish-as-first-language-medium programmes. *Language and Education*, *28*(3), 223–236. https://doi.org/10.1080/09500782.2013.808661

Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics*, *31*, 182–204. https://doi.org/10.1017/s0267190511000092

Dearden, J., & Macaro, E. (2016). Higher education teachers' attitudes towards english medium instruction: A three-country comparison. *Studies in Second Language Learning and Teaching*, *6*(3), 455–486. https://doi.org/10.14746/sllt.2016.6.3.5

Dimova, S. (2017). Life after oral English certification: The consequences of the test of oral English proficiency for academic staff for EMI lecturers. *English for Specific Purposes*, *46*, 45–58. https://doi.org/10.1016/j.esp.2016.12.004

Dimova, S. (2020). English language requirements for enrolment in EMI programs in Higher Education: A European case. *Journal of English for Academic Purposes*, *47*, 100896. https://doi.org/10.1016/j.jeap.2020.100896

Dimova, S., & Kling, J. (2018). Assessing english-medium instruction lecturer language proficiency across disciplines. *TESOL Quarterly*, *52*(3), 634–656. https://doi.org/10.1002/tesq.454

Doiz, A., Lasagabaster, D., & Sierra, J. M. (2014). Clil and motivation: The effect of individual and contextual variables. *The Language Learning Journal*, *42*(2), 209–224. https://doi.org/10.1080/09571736.2014.889508

Đorđević, J. P., & Blagojević, S. N. (2019). University teachers' attitudes towards the implementation of English as a medium of Instruction in Serbian Higher Education, *Nasle e, 153.*

Ekoç, A. (2018). English medium instruction (EMI) from the perspectives of students at a Technical University in Turkey. *Journal of Further and Higher Education*, *44*(2), 231–243. https://doi.org/10.1080/0309877x.2018.1527025

Eurydice (2006). *Content and Language Integrated Learning (CLIL) at School in Europe*. Brussels: Eurydice European Unit.

García, O., & Beardsmore, H. B. (2013). *Bilingual education in the 21st Century: A global perspective*. Wiley-Blackwell.

Garrett, P., Coupland, N., & Williams, A. (2006). *Investigating language attitudes: Social meanings of dialect, ethnicity and performance*. University of Wales Press.

Gundermann, S. (2014). *English-medium instruction: Modelling the role of the native speaker in a lingua franca context*. Albert-Ludwigs-Universität.

Helm, F., & Guarda, M. (2015). “improvisation is not allowed in a second language”: A survey of Italian lecturers’ concerns about teaching their subjects through English. *Language Learning in Higher Education*, *5*(2). https://doi.org/10.1515/cercles-2015-0017

Hu, G. (2019). English-medium instruction in higher education: Lessons from China. *The Journal of AsiaTEFL*, *16*(1), 1–11. https://doi.org/10.18823/asiatefl.2019.16.1.1.1

Jensen, C., & Thøgersen, J. (2011). Danish University lecturers’ attitudes towards English as the medium of instruction. *Ibérica, Revista de La Asociación Europea de Lenguas Para Fines Específicos*, *22*, 13–33.

Kuteeva, M. (2018). Researching English-medium instruction at Swedish universities. *English-Medium Instruction from an English as a Lingua Franca Perspective*, 46–63. https://doi.org/10.4324/9781351184335-5

Lasagabaster, D., Doiz, A., & Sierra, J. M. (2014). In *Motivation and foreign language learning: From theory to practice*. essay, John Benjamins Publishing Company.

Macaro, E. (2018). *English medium instruction: Content and language in policy and Practice*. Oxford University Press.

Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching*, *51*(1), 36–76. https://doi.org/10.1017/S0261444817000350

Mauranen, A. (2006). Signaling and preventing misunderstanding in English as Lingua Franca Communication. *International Journal of the Sociology of Language*, *2006*(177). https://doi.org/10.1515/ijsl.2006.008

Mira, S., San José Martínez, F., Hontoria, C., Sanz-Cobena, A., Rodríguez Sinobas, L., & Cuadrado, G. (2021). Attitudes of academics and students towards English-medium instruction in engineering studies. *European Journal of Engineering Education*, *46*(6), 1043–1057. https://doi.org/10.1080/03043797.2021.1987392

Orduna-Nocito, E., & Sánchez-García, D. (2022). Aligning higher education language policies with lecturers’ views on EMI practices: A Comparative Study of ten european universities. *System*, *104*, 102692. https://doi.org/10.1016/j.system.2021.102692

O’Dowd, R. (2018). The training and accreditation of teachers for english medium instruction: An overview of practice in European universities. *International Journal of Bilingual Education and Bilingualism*, *21*(5), 553–563. https://doi.org/10.1080/13670050.2018.1491945

Ozer, O. (2020). Lecturers' experiences with English-medium instruction in a state university in Turkey: Practices and challenges. *Issues in Educational Research*, *30*(2), 612–634.

Papatsiba, V. (2006). Making Higher Education more European through student mobility? revisiting EU initiatives in the context of the bologna process. *Comparative Education*, *42*(1), 93–111. https://doi.org/10.1080/03050060500515785

Roothooft, H. (2019). Spanish lecturers' beliefs about English medium instruction: Stem versus humanities. *International Journal of Bilingual Education and Bilingualism*, *25*(2), 627–640. https://doi.org/10.1080/13670050.2019.1707768

Saarinen, T., & Nikula, T. (2012). 7. implicit policy, invisible language: Policies and practices of international degree programmes in Finnish Higher Education. *English-Medium Instruction at Universities*, 131–150. https://doi.org/10.21832/9781847698162-011

Sabaté-Dalmau, M. (2016). The englishisation of higher education in Catalonia: A critical sociolinguistic ethnographic approach to the students' perspectives. *Language, Culture and Curriculum*, *29*(3), 263–285. https://doi.org/10.1080/07908318.2016.1153108

Santos, A., Cenoz, J., & Gorter, D. (2017). Attitudes and anxieties of business and education students towards English: Some data from the Basque Country. *Language, Culture and Curriculum*, *31*(1), 94–110. https://doi.org/10.1080/07908318.2017.1350189

Simbolon, N. (2018). EMI in Indonesian Higher Education: Stakeholders' perspectives. *TEFLIN Journal - A Publication on the Teaching and Learning of English*, *29*(1), 108. https://doi.org/10.15639/teflinjournal.v29i1/108-128

Skutnabb-Kangas, T., Phillipson, R., & Rannut, M. (1995). *Linguistic human rights: Overcoming linguistic discrimination*. Mouton de Gruyter.

Söderlundh, H. (2013). Applying transnational strategies locally: English as a medium of instruction in swedish higher education. *Nordic Journal of English Studies*, *12*(1), 113. https://doi.org/10.35360/njes.278

Wächter, B., & Maiworm, F. (2008). *English-taught programmes in European Higher Education: The picture in 2007*. Lemmens.

Wächter, B., & Maiworm, F. (2014). *English-taught programmes in European Higher Education the State of play in 2014*. Lemmens.

West, R., & Aşik, G. (2015). *The state of English in higher education in Turkey: A baseline study*. British Council.

Wilkinson, R. (2012). 1. English-medium instruction at a Dutch university: Challenges and pitfalls. *English-Medium Instruction at Universities*, 3–24. https://doi.org/10.21832/9781847698162-005

Wilkinson, R., Zegers, V., & Leeuwen, C. van. (2006). *Bridging the assessment gap in English-Medium Higher Education*. AKS-Verlag.

Yang, P. (2016). Intercultural Nonverbal Communication Competence as intercultural responsiveness in the Second language learning classroom. *Intercultural Responsiveness in the Second Language Learning Classroom*, 127–147. https://doi.org/10.4018/978-1-5225-2069-6.ch008

The Impact of Parental Gender and Home Input on the Minority Language Proficiency of Dutch Bilingual Children

Research report

Nivedha Narayanaswamy

Research Master *Linguistics and Communication Sciences*, Radboud University Nijmegen

Abstract

Home input is a major factor that contributes to the acquisition of the minority or heritage language (HL) in bilingual children that belong to One Parent, One Language (OPOL) families. However, the role of parental gender in OPOL families is unexplored. The present study examined this with 7-10-year-old Dutch-German (*n*=31) and Dutch-English (*n*=9) children. First, we investigated the relationship between HL home input and the gender of the HL-speaking parent. Then, we examined if this relationship predicts their children's performance on a sentence repetition task (SRT) and a cross-linguistic lexical task (CLT). Analyses revealed that mothers contribute more to HL input compared to fathers. Further, HL home input was indeed a significant predictor of CLT scores but not SRT scores. An interaction effect was also seen between parental gender and HL home input: as home input increased, children with HL-speaking fathers performed better than children with HL-speaking mothers. However, follow-up models suggest that this could be due to the thresholding of maternal input, suggesting that parental gender differences do play a crucial role in children's HL development.

Keywords: bilingualism, parental input, OPOL, gender, home input

1. Background

Input majorly contributes to language development in both monolingual and bilingual children (Thordardottir, 2017; Huttenlocher et al., 1991; Pearson, 2007). Bilingual children may be exposed to the majority language in various scenarios outside the home, but exposure to the minority or heritage language (henceforth, *HL*) may be more restrictive (Unsworth, 2016).

Home input primarily arises from parents, caregivers, and siblings (Unsworth, 2016). Various aspects of home input across languages have been linked to children's growing language skills, including parental proficiency (Paradis, 2011; Chondrogianni & Marinis, 2011), input quantity (Buac et al., 2014; Thordardottir, 2011; Altinkamis & Simon, 2020; Daskalaki et al., 2018; Flores et al., 2016), input quality (Rowe, 2012; Daskalaki et al., 2020), and family language patterns (Verhagen et al., 2022). Family language patterns refer to which languages are used with the child at home by different caregivers in the family. There is great heterogeneity in the family patterns adopted by caregivers to ensure HL home input (Unsworth et al., 2019). Children are most likely to be bilingual if both parents speak the HL and at least one parent speaks the majority language (De Houwer, 2007). However, this strategy may not be possible for all families because not all parents may speak the HL or majority language natively. A popular alternative is the One Parent, One Language (henceforth, *OPOL*) approach. Here, each parent speaks exclusively one language to the child, either the majority language or the HL. The present study concerns the effect of HL home input quantity (henceforth, *HL home input*) on children's HL proficiency in OPOL families.

The success of OPOL as a language strategy greatly depends on the above-mentioned aspects of home input. However, it has not been examined if the gender of the HL-speaking parent affects the success of this bilingual parenting strategy (De Houwer, 2007). Few studies have examined the contribution of input from both parents separately in improving children's HL proficiency. Sun et al. (2022) found that maternal HL proficiency mediated the relationship between socio-economic status and children's HL proficiency after controlling for home input in English-Mandarin bilinguals. They also found a positive correlation between mothers' HL proficiency and use, but not fathers'. Additionally, Hammer et al. (2012) found evidence of the mothers' education predicting Spanish-English children's proficiency in both L1 and L2 tasks, since high maternal (but not paternal) education correlated to high home input. Place and Hoff (2011), also studying Spanish-English bilinguals, found that children had lower native HL input with HL-speaking fathers compared to HL-speaking mothers. Romanowski (2022) found that the *motivation* and *willingness* to indulge in joint childcare practices influence the extent to which fathers pass on the HL in Polish-English bilinguals, not parental gender per se.

In sum, while HL home input has been shown to affect children's HL acquisition, the role of parental gender and its interaction with HL home input remains unclear.

2. Research Questions (RQs)

The present study used this premise to examine data from the 2in1 project (Unsworth et al., 2022), which investigated cross-linguistic influences in Dutch bilingual children. This project comprises data from a parental questionnaire, the BiLEC (Bilingual Language Experience Calculator; Unsworth, 2012), scores of the Cross-linguistic Lexical Task (CLT; Haman et al., 2017) and the Sentence Repetition Task (SRT; Armon-Lotem & Meir, 2015).

> ***RQ 1****: To what extent does the use of the HL differ between mothers and fathers in OPOL families?*

Prediction 1 (Main Effect 1): HL input from the mother is significantly higher than that of fathers, as observed by Place and Hoff (2011).

> ***RQ 2****: To what extent does the relation between parental input and children's proficiency in the HL differ depending on which parent uses the HL?*

Prediction 2 (Main effect 2): HL home input predicts the child's HL proficiency as seen by Buac et al. (2014), Thordardottir (2011), and Altinkamis and Simon (2020). We predicted higher HL proficiency (CLT and SRT) scores for children with higher HL home input.

Prediction 3 (Interaction effect): There is a significant interaction between parental gender and HL home input. Since not many studies have examined this interaction, we made no predictions but could speculate:

- As home input increases, children with HL-speaking mothers perform better than children with HL-speaking fathers, since mothers are seen as "caretakers of first-language maintenance" (Akoğlu & Yagmur, 2016, p.9).

- As home input increases, children with HL-speaking fathers perform better than children with HL-speaking mothers, since HL-speaking fathers may play an even more significant role than mothers: fathers represent children's "outside world" (Hammer et al., 2012, p. 1261).
- A lack of significant interaction would indicate no impact of parental gender.

3. Methods

3.1 Participants

Since the RQs examine HL proficiency scores and information about family constellations, only datasets that had no missing entries for CLT scores, SRT scores, and the BiLEC were examined. The datasets that met these criteria comprised 7-to-10-year-old Dutch-German, Dutch-Turkish, and Dutch-English children. We defined OPOL families as those in which only one parent speaks the HL to the child and does so at least 70% of the time. This cutoff was chosen because we noted that using 95% cutoff used by Verhagen et al. would be too restrictive for the families in our sample. The majority of families with only one HL-speaking parent in our sample did not speak the HL 95% of the time, reflecting the fact that the OPOL strategy may not be systematically enforced in practice (Venables et al., 2013). Although the cutoff value is arbitrary as a reviewer pointed out, 70% is more practically relevant for our sample.

Only three Dutch-Turkish children met this criterion. Since this number was too low to carry out inferential statistical analysis, the Dutch-Turkish data was excluded. Our final participant pool was unbalanced with respect to the gender of the HL-speaking parent because of the nature of our sample. More HL-speaking mothers were surveyed compared to HL-speaking fathers. Table 1 details the final participant pool.

Table 1
Overview of Participants (N=40)

Language pair	*n* (after exclusion)
Dutch-German	31
Dutch-English	9

3.2 Methodology and Analyses

For each subset, corresponding BiLEC data was used to create a new variable 'HL_Parent.' This variable coded the gender of the HL-speaking parent, with two values, M (mother) or F (father).

Since there was no direct of HL home input in the questionnaire, another new variable HL_Home_Input, which indicated the percentage of HL home input per week, was created. This was calculated as the complement of the percentage of Dutch home input per week ('nl_pc_home' in the BiLEC document). This variable took values between 0 (no HL home input) to 1 (HL home input 100% of the time). Additionally, a variable 'HL_Name' was created to code the HL of our participants, with two values ENG (English) and DEU (German).

SRT scores were obtained from the variable VerbatimPropCorr (the proportion of verbatim correctly repeated sentences), with values between 0 (the child did not repeat any sentences verbatim) to 1 (the child repeated all sentences verbatim). Finally, CLT scores were obtained from the variable 'Percentage,' which denoted the percentage of correct scores. This variable took values from 0 (no correct answers) to 100 (all correct answers).

One of the factors that contributes to differences in bilingual acquisition is the typological distance between the majority language and HL (Blom et al., 2019). Dutch and German are both Germanic languages. Turkish, in contrast, is from the Turkic family, which is not as closely related to Dutch as German. To check if the variation in HL (English/German) in our dataset was a confound, an independent samples t-test was run to see if there was a systematic difference in HL home input between the two groups. The *t*-test showed no significant differences, $t(20.25)=-.33$, $p=.742$. Dutch-English ($M=.48$, $SD=.11$) and Dutch-German children ($M=.46$, $SD=.17$) did not differ in the amount of HL home input. Therefore, language distance was not included as a control variable in further analyses.

Regression analyses were performed using the *lm* function of the *lme4* package (Bates et al., 2015) in R (v4.2.3; R Core Team 2023) and RStudio (Rstudio Team, 2023), and summaries were created with the *car* package (Fox & Weisberg, 2019) . One base model each was constructed with CLT scores or SRT scores as the dependent variable and HL home input as a predictor. Subsequently, predictors of interest (HL Parent and an interaction between HL home input and HL Parent) were added. Sum and polynomial contrast coding were applied to the categorical fixed effects variable (HL Parent). For each model, a stepwise variable selection procedure was used to remove non-significant predictors to obtain the most parsimonious model. Likelihood ratio tests were performed to compare the goodness of fit using the *anova* function of the *base* package (R Core Team 2023).

4. Results

The number of HL-speaking mothers (n=27) was higher than HL-speaking fathers (n=13). To investigate the relationship between parental gender and HL home input, an independent samples t-test was run. This revealed significant differences in HL home input as a function of parental gender, $t(32.93)=-4.17$, $p<.001$. Maternal HL home input (M=.53, SD=.15) was significantly higher than paternal (M=.36, SD=.11), confirming our first prediction.

For the second hypothesis, regression analysis revealed that HL home input was a significant predictor for CLT scores but not SRT scores. An analysis of variance revealed that adding HL Parent as a predictor improved the model fit for the CLT score, $F(2, 38) = 5.16$, p=.014. These results are summarized in Table 2. Since HL home input did not significantly predict SRT scores, only CLT scores were used for further analyses.

Table 2

Summary of Regression Models for SRT and CLT Scores

	SRT Score					CLT Score				
Model	Adjusted R^2	F	df	*p*	AIC	Adjusted R^2	F	df	*p*	AIC
Base	.06	3.5	38	.07	7.69	0.3	15.44	38	<.001	316.34
Final	.14	3.19	36	.05	5.78	0.41	9.719	36	<.001	310.26

For CLT scores, there was indeed a main effect of HL home input, in line with Prediction 2. In addition, a significant interaction was observed between HL home input and HL Parent (Table 3). As HL home input increased, children with HL-speaking fathers performed better than those with HL-speaking mothers (Figure 1).

Table 3

Final Multiple Linear Regression Models for Children's CLT Scores With Significant Predictors, Estimate (b-coefficient), Standard error (SE), t-value, and p-value

Model	Significant predictors	Estimate	*SE*	*t*	*p*
Base	HL home input	47.96	12.21	3.93	<.001
Final	HL home input	58.37	16.72	3.49	.001
	HL Parent	-19.82	6.81	-2.91	<.01
	HL home input × HL Parent	39.08	16.72	2.34	.02

To further analyse this interaction, one follow-up model was created for each level of the variable HL Parent. HL home input was added as a predictor for CLT scores. HL home input was a significant predictor only at the level of the father but not the mother (Table 4). Therefore, while children with high HL home input from

HL-speaking fathers had high proficiency, this was not observed with HL-speaking mothers.

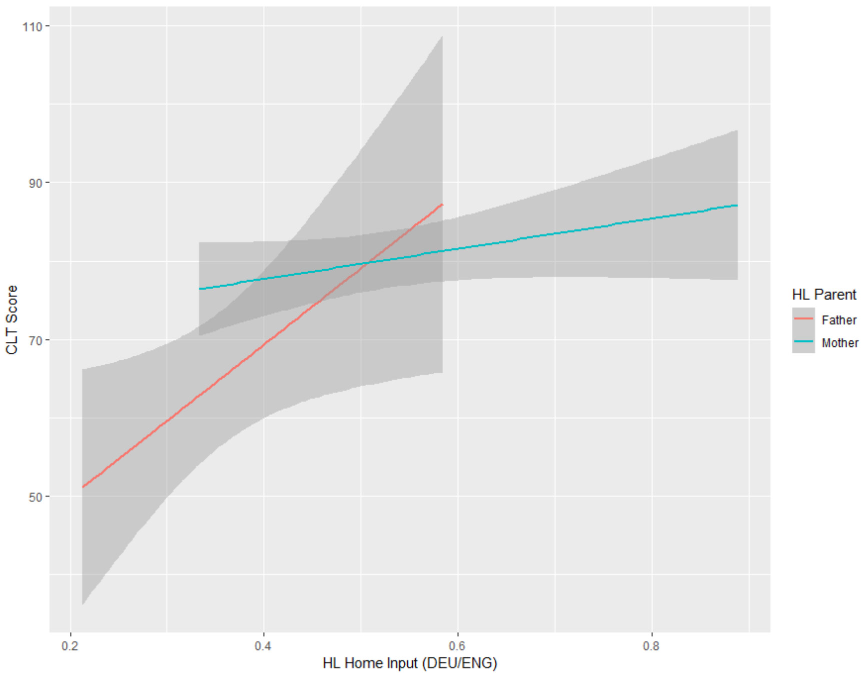
Table 4

Follow-up Models to Predict CLT Scores With Levels 'Father' and 'Mother' of HL_Parent With Model Summaries, Estimate (b-coefficient), Sandard Error (SE), t-value, and p-value.

Level	Adjusted R^2	F	df	Estimate	*SE*	*t*	*p*
Mother	.06	2.62	25	19.29	11.92	1.62	.12
Father	.31	6.19	11	97.45	39.18	2.49	.03

Figure 1

Interaction Effect of Parental Gender and HL home input on CLT Scores



5. Discussion

This study examined the relationship between parental gender and HL home input in influencing the HL proficiency of Dutch bilingual OPOL children.

First, it was found that HL-speaking mothers provide much higher HL home input than HL-speaking fathers, in line with our prediction. Such a trend motivates researchers to rely on maternal metrics to study bilingual development (Duncan & Paradis, 2018; Akoğlu & Yağmur, 2016).

Secondly, it was seen that HL home input predicted only CLT and not SRT scores, partially confirming our second prediction. The CLT and SRT scores are significantly positively correlated but it is not unusual that only the CLT score predicts HL home input. The CLT assesses comprehension and production of nouns and verbs, while the SRT is concerned with syntactic representation. Differential outcomes for the two variables show that different amounts of HL home input are necessary for different aspects of proficiency. While more input is needed to expand children's vocabulary, less input may prove sufficient to learn grammatical structures and build a receptive framework. In fact, Thordardottir (2011), studying Spanish-English bilinguals, observed that while bilinguals need at least 40% exposure to a language to match monolingual scores in receptive tests, 60% is needed for expressive scores.

Analyses also revealed that as HL home input increases, children with HL-speaking fathers perform better than children with HL-speaking mothers. While this suggests that parental gender does indeed influence HL acquisition, these results must be interpreted with caution. Additional input beyond the 50-60% threshold does not result in significantly better scores (Maas, 2014; Thordardottir, 2011). In our study, the mean home input from HL-speaking mothers is 53%. The lack of significant interaction for HL-speaking mothers could be a consequence of reaching the threshold, since additional input may not be necessary beyond this point (Gathercole, 2007). For HL-speaking fathers, the mean HL home input is much below the threshold (36%). Therefore, the more input, the better the expressive test scores.

6. Limitations and Future Research

While the present study highlights the role of HL-speaking fathers in HL transmission, future research could address a few limitations.

First, the current results should be interpreted cautiously due to the low sample size and lack of data points at all levels of home input. To further tease apart the significant interaction, subsequent regression analyses were attempted after thresholding HL home input into three levels – low (0 – 33%), medium (34 – 66%), and high (67 – 100%). However, the low and high levels could not be analysed as there were no data points for both genders at these levels. Additionally, the number of HL-speaking mothers and fathers was not equal due to the nature of our sample. More HL-speaking mothers were surveyed than HL-speaking fathers. Data points are needed across the entire range of HL home input from both parents to draw stronger conclusions.

Secondly, current study examined only two groups of Dutch bilinguals, as the Dutch-Turkish data had to be discarded. Future research could examine more diverse language pairs to understand the role of language distance. Our study focussed only on OPOL families. Future studies could examine other family constellations, following Unsworth et al. (2019). This would help draw stronger inferences about the role of parental gender in HL transmission.

Finally, our study examined children from age 7-10 and was restricted to *current* HL home input. Expanding this age range and studying the cumulative amount of exposure could provide insights into HL transmission at various developmental milestones.

7. Conclusion

These findings highlight the differences in distribution of HL home input from the father and the mother. Parental gender is indeed crucial in how HL home input affects HL proficiency. HL-speaking fathers must increase their HL home input to ensure HL transmission. A trend is noticed in the field of child bilingualism research about the role of each parent in the child's bilingual development – the mother's role is so established that maternal metrics are used as predictors of the child's language growth. Mothers indulge in language activities to pass on the first language as agents of HL transmission (Akoğlu & Yagmur, 2016). However, the role of fathers is unclear. The current study highlights this discrepancy significantly, as reflected in the disproportionately high number of HL-speaking mothers compared to HL-speaking fathers. These circumstances are particularly concerning when considering that fathers may represent the "outside world" to children (Hammer et al, 2012, p. 1261). If fathers do not engage sufficiently in language transmission practices, children may experience an increased difficulty in HL acquisition. In sum, this study highlights the role of HL-speaking fathers in facilitating HL transmission.

Acknowledgements

I would like to thank dr. Sharon Unsworth for her feedback and permission to use the dataset for this report from the 2in1 research project, an NWO-funded project that researches language acquisition of bilingual children.

References

Akoğlu, G., & Yagmur, K. (2016). First-language skills of bilingual Turkish immigrant children growing up in a Dutch submersion context. *International Journal of Bilingual Education and Bilingualism*, *19*(6), 706–721. https://doi.org/10.1080/13670050.2016.1181605

Altinkamis, F., & Simon, E. (2020). Language abilities in bilingual children: The effect of family background and language exposure on the development of Turkish and Dutch. *International Journal of Bilingualism*, *24*(5–6), 931–951. https://doi.org/10.1177/1367006920909889

Armon-Lotem, S., & Meir, N. (2015). Sentence repetition. In J. De Jong (Ed.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (Vol. 12, pp. 95–122). Multilingual Matters.

Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects models usinglme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Blom, E., Boerma, T., Bosma, E., Cornips, L., Van Den Heuij, K., & Timmermeister, M. (2019). Cross-language distance influences receptive vocabulary outcomes of bilingual children. *First Language*, *40*(2), 151–171. https://doi.org/10.1177/0142723719892794

Buac, M., Gross, M., & Kaushanskaya, M. (2014). The Role of Primary Caregiver Vocabulary Knowledge in the Development of Bilingual Children's Vocabulary Skills. *Journal of Speech Language and Hearing Research*, *57*(5), 1804–1816. https://doi.org/10.1044/2014_jslhr-l-13-0055

Chondrogianni, V., & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism*, *1*(3), 318–345. https://doi.org/10.1075/lab.1.3.05cho

Daskalaki, E., Blom, E., Chondrogianni, V., & Paradis, J. (2020). Effects of parental input quality in child heritage language acquisition. *Journal of Child Language*, *47*(4), 709–736. https://doi.org/10.1017/s0305000919000850

Daskalaki, E., Chondrogianni, V., Blom, E., Argyri, F., & Paradis, J. (2018). Input effects across domains: The case of Greek subjects in child heritage language. *Second Language Research*, *35*(3), 421–445. https://doi.org/10.1177/0267658318787231

De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, *28*(3), 411–424. https://doi.org/10.1017/s0142716407070221

Duncan, T. S., & Paradis, J. (2018). How does maternal education influence the linguistic environment supporting bilingual language development in child second language learners of English? *International Journal of Bilingualism*, *24*(1), 46–61. https://doi.org/10.1177/1367006918768366

Flores, C., Santos, A. J., Jesus, A., & Marques, R. C. (2016). Age and input effects in the acquisition of mood in Heritage Portuguese. *Journal of Child Language*, *44*(4), 795–828. https://doi.org/10.1017/s0305000916000222

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* [Software]. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Gathercole, V. C. M. (2007). Miami and North Wales, So Far and Yet So Near: A Constructivist Account of Morphosyntactic Development in Bilingual Children. *International Journal of Bilingual Education and Bilingualism*, *10*(3), 224–247. https://doi.org/10.2167/beb442.0

Haman, E., Łuniewska, M., Hansen, P. B. L., Simonsen, H. G., Chiat, S., Bjekić, J., Blažienė, A., Chyl, K., Dabašinskienė, I., De Abreu, P. E., Gagarina, N., Gavarró, A., Håkansson, G., Harel, E., Holm, E., Kapalková, S., Kunnari, S., Levorato, C., Lindgren, J., . . . Armon-Lotem, S. (2017). Noun and verb knowledge in monolingual preschool children across 17 languages: Data from Cross-linguistic Lexical Tasks (LITMUS-CLT). *Clinical Linguistics & Phonetics*, *31*(11–12), 818–843. https://doi.org/10.1080/02699206.2017.1308553

Hammer, C. S., Komaroff, E., Rodriguez, B. L., López, L. M., Scarpino, S. E., & Goldstein, B. A. (2012). Predicting Spanish–English Bilingual Children's Language Abilities. *Journal of Speech Language and Hearing Research*, *55*(5), 1251–1264. https://doi.org/10.1044/1092-4388(2012/11-0016

Huttenlocher, J., Haight, W., Bryk, A. S., Seltzer, M. L., & Lyons, T. S. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2), 236–248. https://doi.org/10.1037/0012-1649.27.2.236

Maas, E. L. H. (2014). *The role of bilingual language input in children's receptive and productive vocabulary development* [MA Thesis]. Utrecht University.

Paradis, J. (2011). Individual differences in child English second language acquisition. *Linguistic Approaches to Bilingualism*, *1*(3), 213–237. https://doi.org/10.1075/lab.1.3.01par

Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics*, *28*(3), 399–410. https://doi.org/10.1017/s014271640707021x

Place, S., & Hoff, E. (2011). Properties of Dual Language Exposure That Influence 2-Year-Olds' Bilingual Proficiency. *Child Development*, *82*(6), 1834–1849. https://doi.org/10.1111/j.1467-8624.2011.01660.x

Romanowski, P. (2022). Paternal agency in heritage language maintenance in Australia: Polish fathers in action. *International Journal of Bilingual Education and Bilingualism*, *25*(9), 3320–3332. https://doi.org/10.1080/13670050.2022.2050994

Rowe, M. L. (2012). A Longitudinal Investigation of the Role of Quantity and Quality of Child-Directed Speech in Vocabulary Development. *Child Development*, *83*(5), 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x

Sun, H., Low, J., & Chua, I. (2022). Maternal heritage language proficiency and child bilingual heritage language learning. *International Journal of Bilingual Education and Bilingualism*, *26*(7), 861–875. https://doi.org/10.1080/13670050.2022.2130153

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15*(4), 426–445. https://doi.org/10.1177/1367006911403202

Thordardottir, E. (2017). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism*, *23*(1), 236–255. https://doi.org/10.1177/1367006917722418

Unsworth, S. (2012). Assessing the role of current andcumulativeexposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, *16*(1), 86–110. https://doi.org/10.1017/s1366728912000284

Unsworth, S. (2016). Quantity and quality of language input in bilingual language development. In *American Psychological Association eBooks* (pp. 103–121). https://doi.org/10.1037/14939-007

Unsworth, S. (2016). Quantity and quality of language input in bilingual language development. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 103–121). American Psychological Association. https://doi.org/10.1037/14939-007

Unsworth, S., Brouwer, S., De Bree, E., & Verhagen, J. (2019). Predicting bilingual preschoolers' patterns of language development: Degree of non-native input matters. *Applied Psycholinguistics*, *40*(05), 1189–1219. https://doi.org/10.1017/s0142716419000225

Unsworth, S., Kootstra, G. J., Van Dijk, C., Karaca, F., Koutamanis, E., Van Bosch, J., Van Wonderen, E., & Van Zwet, J. (2022). 2In1: Bilingualism in children [Dataset]. In *2 in 1 research project*.

Venables, E. H., Eisenchlas, S. A., & Schalley, A. C. (2013). One-parent-one-language (OPOL) families: is the majority language-speaking parent instrumental in the minority language development? *International Journal of Bilingual Education and Bilingualism*, *17*(4), 429–448. https://doi.org/10.1080/13670050.2013.816263

Verhagen, J., Kuiken, F., & Andringa, S. (2022). Family language patterns in bilingual families and relationships with children's language outcomes. *Applied Psycholinguistics*, *43*(5), 1109–1139. https://doi.org/10.1017/s0142716422000297

Older Sibling Influences on Bilingual Children's Dutch Language Development and the Effect of Age Difference: the Closer the Better

Research report

Elsa Opheij[1], Anna Gorter[2]

[1]Research Master *Linguistics and Communication Sciences*, Radboud University Nijmegen
[2]Master *Language Variation and Multilingualism*, Radboud University Nijmegen

Abstract

This study investigates the effect of older siblings on their younger sibling's Dutch language development and proficiency (RQ1). Moreover, this study focuses on the age difference between the older and the younger sibling to investigate whether a smaller or larger gap leads to a higher language proficiency (RQ2). Dutch language proficiency was measured by using sentence repetition task-scores (SRTs). For the first research question the participants were children growing up with Dutch and another language (i.e., German, English, Greek, Turkish, Spanish). Only children with one older sibling or no siblings were included. For the second research question a subset of these data was used (only the children with siblings) and the influence of age difference was investigated. The results showed that having a sibling leads to lower SRT-Dutch scores (RQ1). However, these results might have been impacted by various factors, like the different languages spoken by a child. In addition, a smaller age difference between the two siblings led to higher SRT-Dutch scores (RQ2). Lines for future research include a more extensive analysis of factors that could affect older sibling influences on language development, resulting in a more comprehensive overview of how the presence of an older sibling in a bilingual home setting can affect the younger sibling's societal language development.

Keywords: child bilingualism, language development, language input, older siblings

1. Introduction

Throughout the day, children receive a lot of language input, which they can use to acquire a language. This input is usually not produced by one single person but by different people, like parents, siblings, and peers. Previous research has shown that language input from different input providers affects the bilingual child's language proficiency differently (Tsinivits & Unsworth, 2020). Regarding the input provided by siblings particularly, it has been found that the input bilingual children get from their older siblings is positively related to L2 proficiency scores (Duncan & Paradis, 2020). More specifically, Duncan and Paradis (2020) found a positive relation between the relative quantity of L2 (English) input offered by an older sibling and the younger sibling's L2 abilities on a range of linguistic subdomains. Another study found that children with older siblings are more advanced in English, but children without siblings are more advanced in the minority language spoken at home (Bridges & Hoff, 2012). The authors suggest that this is because older siblings bring the school language, in this case English, into the home. However, older studies looking at monolingual families have found an effect of the presence of an older sibling on the younger sibling's language input and output in mother-child interaction (Wellen, 1985). More specifically, mothers read stories to the younger sibling and asked questions about each story either in presence or in absence of the older sibling. The presence of an older sibling led to fewer an less informative answers on questions in the tasks.

In this study, we want to explore the role of older siblings on their younger sibling's proficiency in the majority language. Notably, the younger children are from varying language backgrounds, acquiring Dutch and another language, which allows for inspection of an older sibling's role in a rich sample regarding language background.

> ***RQ 1****: To what extent do children with older siblings have a higher Dutch language proficiency than children without older siblings?*

Although previous studies have found varying results, we hypothesise that having an older sibling has a positive effect on Dutch language proficiency scores in bilingual children, in line with previous research (e.g., Duncan & Paradis, 2020; Tsinivits & Unsworth, 2020). These studies have been conducted quite recently, and focussed on bilingual rather than monolingual children. The aim of this study is to investigate whether the positive effect of an older sibling on English proficiency scores found in previous studies (e.g., Duncan & Paradis, 2020; Bridges & Hoff, 2012) is maintained for Dutch as well as for children from various language backgrounds. Moreover, this study investigates children from a different age group compared to previous studies (e.g., Tsinivits & Unsworth, 2020), namely children in elementary school instead of toddlers.

In a study on monolingual siblings, the age difference between siblings was found not to be associated with language development, contrary to predictions (Havron et al., 2019). However, the role of age difference on language development has not been subjected to much scientific research, and the influence of age difference between bilingual siblings in particular has not been studied before. Therefore, this topic will also be addressed.

> ***RQ 2****: To what extent is Dutch language proficiency affected by age difference between siblings?*

Formulating a clear hypothesis for this research question is more challenging. Although Havron et al. (2019) did not find any relation between age difference and language development for monolinguals, several ideas can be posed to suspect a different relation for bilingual children. On the one hand, one could argue that a larger age difference means that the first-born sibling is relatively older, hence has acquired the Dutch language to a larger extent and can provide the younger sibling with better language input and feedback. Furthermore, a larger age gap results in less competition for parents' resources (Havron et al., 2019; Her et al., 2021).

On the other hand, the input and feedback received from an older sibling might be more attuned to the younger sibling's acquisition process when the siblings are closer together in age. Besides, siblings are more likely to be role models with a smaller age gap (Her et al., 2021), possibly also improving their language learning. Apart from the expectation that there is an effect of age difference, no specific hypothesis is formulated regarding the directionality of this effect. This is due to the low availability of research on this topic, so this research question is of a more exploratory nature.

2. Methods

2.1 Participants

Participants were selected based on information provided through a parental questionnaire on children's language experience (BiLEC; Unsworth et al., 2022). The BiLEC contains data from 346 children with various language backgrounds, but all children are native speakers of Dutch and another language. For the current study, only bilingual children (raised with two languages; German-Dutch, English-Dutch, Greek-Dutch, Spanish-Dutch, or Turkish-Dutch) who had either no siblings or only one older sibling were selected. Including children with more than one sibling or with younger siblings would pose a problem. It would obscure studying the effect of age difference and explaining any potential difference between the two experimental groups, as it would be unclear what specific factors might have led to these results. Hence, this categorisation allowed for a sound comparison between the sibling groups. Only children who received at least 10% of the language input from their sibling in Dutch were included (as indicated by parental estimation of the percentage of language input provided in every language by every input provider for the BiLEC questionnaire) . In total, 250 children were excluded due to different reasons (missing SRT data for French-Dutch children, children with no siblings, children with more than one sibling, children with a younger rather than an older sibling, and children receiving less than 10% of Dutch input) leading to a total of 96 children, aged between 4;9 and 10;10 years. Table 1 provides an overview of the number of children per sibling group and language pair.

Table 1

Number of Children per Group for Each Language Pair

Language pair	**Sibling group** $M_{age} = 7.49$; $SD_{age} = 1.67$	**No-sibling group** $M_{age} = 7.67$; $SD_{age} = 1.60$
German-Dutch	18	22
English-Dutch	18	16
Greek-Dutch	6	4
Spanish-Dutch	5	4
Turkish-Dutch	2	1
Total	49	47

Children completed a sentence repetition task (SRT) and a productive vocabulary task (cross-linguistic lexical task, CLT) as measures of language proficiency. As the data from the Dutch SRT were collected for nearly all participants, which was less true for the CLT, SRT-Dutch scores were employed as a measure of Dutch language proficiency. In the sentence repetition task (LITMUS-SRep, see Marinis & Armon-Lotem, 2015, for a detailed description), children hear and repeat sentences of varying levels of complexity which are too long to retain in short-term memory. Hence, this task requires sufficient proficiency to process the sentence heard and repeat it accurately. For this task, an utterance was coded as a correct (i.e., verbatim) repetition if the child's utterance was identical to the target sentence. However, errors related to pronunciation were ignored, as long as the target word could still be recognised. Additionally, alternative forms in the spoken language variety a child speaks were coded as correct (e.g., the abbreviated form *gister* instead of *gisteren* 'yesterday' or *hun* 'their' as a subject pronoun instead of *zij/ze* 'they').

2.2 Design

The SRT-Dutch scores were employed as a measure of Dutch language proficiency, which is the outcome variable for both research questions. These scores were computed as the percentage of correct verbatim items (out of 30). The following variables from the BiLEC were included as predictor variables:

- **Sibling**: the two participant groups entail one group of children with no siblings and one group of children with only one older sibling;
- **Age at testing**: participants' age at testing in years and months;
- **Cumulative LoE Dutch**: cumulative length of exposure to Dutch, i.e., the average percentage of exposure to Dutch at daycare/school and home over time, considering that one year of exposure for a bilingual child is not the same as for a monolingual child;
- **Language pair**: the child's language background (either German, English, Greek, Spanish, or Turkish and Dutch);
- **Age difference**: computed by calculating the difference in age (in months) between the sibling and the child tested;

- **Sibling input % Dutch**: the percentage of input provided by the sibling in Dutch;
- **Item**: included as a random effect to account for random variance caused by particular items in the SRT.

2.3 Data Analysis

The data analysis was conducted by fitting Linear Mixed Effects Models using R (version 3.4.0; R Core Team, 2023). R-package *lme4* (Bates et al., 2015) was used to perform multiple linear mixed-effects regression analyses for SRT-Dutch scores. Modelling commenced with a base model, including only SRT-Dutch scores and sibling and age difference as a predictor for RQ1 and RQ2, respectively. Afterwards, the remaining predictor variables were added to the base model one by one based on expected relevance to the specific research question. Each two subsequent models were compared using ANOVA comparisons of the package *lmerTest* (Kuznetsova et al., 2017) to test if the addition of another covariate improved the model fit. If AIC and BIC scores were significantly lower for the novel model, it could be assumed that including this covariate significantly improved the model fit. Finally, R-package *ggplot2* (Wickham, 2016) and *effects* (Fox, 2003) were used for visualising the data.

3. Results

The descriptive results and mixed effects model results for RQ1 are shown in Table 2 and Table 3, respectively.

Children without siblings (M_{score} = 73.83; SD_{score} = 24.09) scored significantly higher on the SRT-Dutch than children with siblings (M_{score} = 69.80; SD_{score} = 23.88), β = -3.40, SE = 0.72, $p < .001$. This effect of sibling group on SRT-Dutch scores, with higher SRT-Dutch scores for children without siblings, is visualised in Figure 1. Additionally, a significant effect of children's age at testing on SRT scores was found (β = 5.18, SE = 0.28, $p < .001$), indicating that older children scored higher on the SRT than younger children. Cumulative length of exposure was also found to have a significantly positive effect on SRT scores (β = 4.12, SE = 0.27, $p < .001$), meaning that the higher the (cumulative) exposure to Dutch, the higher the SRT score. Additionally, Dutch-German (functioning as the reference level in the mixed effects model, to which the other Language pair levels are compared) children scored significantly higher than

Table 2

Descriptive Results for all Fixed Factors

Sibling	**M/SD**	**SRT-Dutch score (in %)**	Age at testing	**Cumulative LoE Dutch**	**Age difference**	**Sibling input % Dutch**
Y	M	69.80	7.49	4.41	2.37	0.72
	SD	23.88	1.67	1.58	1.18	0.28
N	M	73.83	7.67	4.04	-	-
	SD	24.09	1.62	1.55	-	-

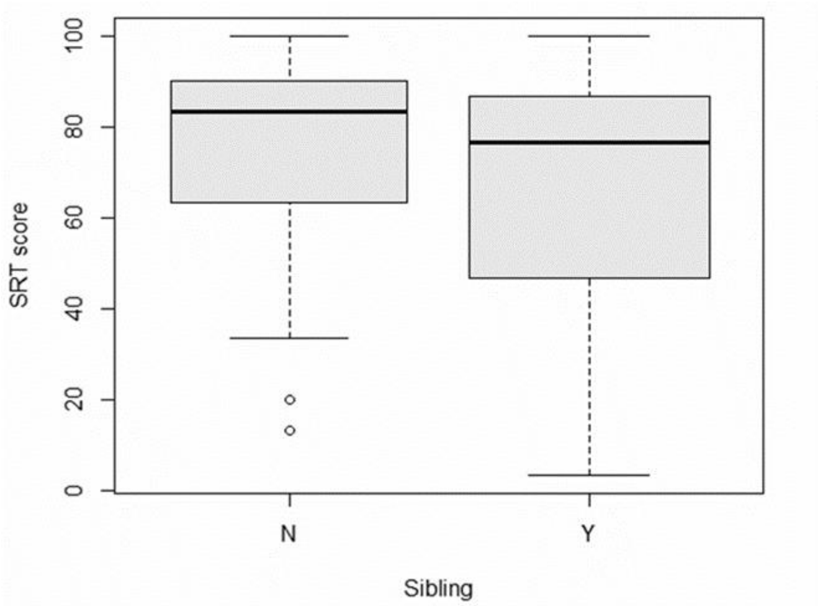
Table 3

Mixed Effects Model for RQ1

Fixed effects	***β***	***SE***	***t*-value**	***p*-value**
(Intercept)	22.26	2.10	10.59	<.001
Sibling: Yes	-3.40	0.72	-4.75	<.001
Age at testing	5.18	0.28	18.31	<.001
Cumulative LoE Dutch	4.12	0.27	15.04	<.001
Language pair: Greek-Dutch	-10.90	1.32	-8.25	<.001
Language pair: English-Dutch	-9.46	0.82	-11.60	<.001
Language pair: Spanish-Dutch	-3.86	1.36	-2.83	.0047
Language pair: Turkish-Dutch	-13.68	2.08	-6.57	<.001

Figure 1

Boxplot of SRT Scores Across the Two Sibling Groups



children in all other language pairs (Dutch-Greek: β = -10.90, *SE* = 1.32, *p* < .001; Dutch-English: β = -9.46, *SE* = 0.82, p < .001; Dutch-Spanish: β = -3.86, *SE* = 1.36, *p* = .0047; Dutch-Turkish: β = -13.68, *SE* = 2.08, *p* < .001).

The mixed effects model results for RQ2 are presented in Table 4. The results showed a significant effect of age difference on SRT scores (β = -8.30, *SE* = 0.74, *p* < .001), indicating that a smaller age gap between siblings was associated with higher proficiency scores. Several covariates showed significant results: age at testing (β = 3.42, *SE* = 0.34, *p* < .001) and cumulative exposure to Dutch (β = 7.43, *SE* = 0.42, *p* < .001) were both positively related to SRT scores, meaning that a higher age at testing and a higher exposure to Dutch led to higher scores. Including *percentage* of input provided in Dutch by the sibling as a fixed effect did not significantly improve the model fit. However, a significant interaction between age difference and *percentage* of input provided in Dutch by the sibling on SRT scores was found (β = 2.29, *SE* = 0.90, *p* = .0115), which indicates that the effect of age difference on score differed based on the *percentage* of Dutch input provided by the sibling. This interaction is visualised in Figure 2; the higher the age difference between siblings, the larger the positive effect of relatively more Dutch input given by the sibling on SRT-Dutch scores. This suggests that it mattered less how much input children receive from an older sibling when the age difference is small.

The fixed and random effects that were included in the final model per research question are presented in Table 5.

Table 4

Mixed Effects Model for RQ2

Fixed effects	*β*	*SE*	*t*-value	*p*-value
(Intercept)	27.14	2.26	12.00	<.001
Age difference	-8.30	0.74	-11.17	<.001
Age at testing	3.42	0.34	9.92	<.001
Cumulative LoE Dutch	7.43	0.42	17.64	<.001
Age difference:Sibling input % Dutch	2.29	0.90	2.53	0.0115

Figure 2

Plot of Interaction Between Percentage of Dutch Input Provided by the Sibling and Age Difference on SRT-Dutch Scores

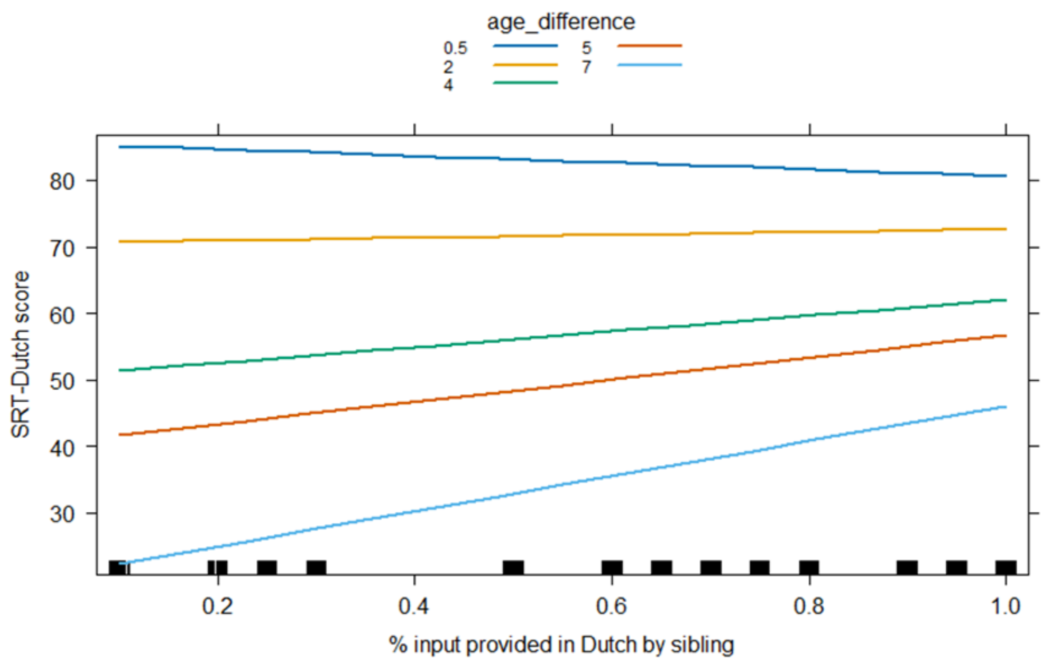


Table 5

Outcome and Predictor Variables for the Best Model per Research Question

Variables	RQ1	RQ2
Outcome	SRT-Dutch score	SRT-Dutch score
Predictor	Fixed effects • Sibling • Age at testing • Cumulative LoE Dutch • Language pair	Fixed effects • Age difference • Age at testing • Cumulative LoE Dutch • Age difference*Sibling input % Dutch
	Random effects • By-item	Random effects • By-item

4. Discussion and Conclusion

The aims of this study were twofold. Firstly, we investigated the effect of having an older sibling versus having no siblings on bilingual children's Dutch language proficiency, using performance on a sentence repetition task as a measure of language proficiency. Participants with different language backgrounds, acquiring Dutch and another language, were selected from the BiLEC dataset (Unsworth et al., 2022).

Mixed effects models showed that having a sibling led to significantly lower proficiency scores than having no siblings, which was not in line with our hypothesis based on previous studies on bilinguals (Bridges & Hoff, 2012; Duncan & Paradis, 2020). However, it is in line with Wellen's (1985) study on monolinguals. To control for a potential influence of the distribution across language pairs per group (sibling versus no-sibling), language pair was included in the analysis. Although SRT-Dutch scores were higher for Dutch-German children than for all other language pairs, including language pair still revealed that SRT-Dutch scores for children without a sibling were significantly higher than for children with an older sibling. A possible explanation for these results might be that children with an older sibling have to compete for their caregivers' attention, which might result in less language input to the younger sibling, hence leading to lower proficiency scores (e.g., Havron et al., 2019).

Secondly, the role of age difference between siblings was studied, which was found to affect SRT-Dutch scores, unlike Havron et al.'s (2019) findings. Our results showed that a smaller age gap had a positive effect on the target children's language development, as measured by performance on a sentence repetition task.

Havron et al. (2019) expected to find an influence of age difference between monolingual siblings on the younger sibling's language development. More specifically, they hypothesised that a larger age gap would have a less detrimental effect for the younger sibling's linguistic skills, since there is less competition for parental resources (less care and supervision required at a higher age), and the older sibling's social and linguistic skills are likely to be more developed, suggesting that better linguistic input could be provided. However, they did not find such an effect in their study. A possible explanation for this finding could be that an older sibling with a smaller age gap is more likely to be a role model for the younger sibling, and input and feedback might be more attuned to the younger sibling's language learning process (Her et al., 2021).

Moreover, Havron et al. (2019) focussed on monolingual children, whereas the current study looked at bilingual children, meaning that more factors might be at play here. The different language(s) spoken in a home setting might alter the role of the older sibling's language input and feedback, as less input in the societal language from the parents could increase the importance of input from siblings. Additionally, sex of the sibling and overall exposure to the societal language (quantity and quality) might have impacted the results.

The effect of age difference was also found to be moderated by the percentage of Dutch input provided by the sibling. This suggests that the positive influence of a higher amount of input provided in Dutch by the sibling on the target child's Dutch language proficiency differed by age difference in the sense that a target child with a larger age gap with their sibling benefited more from a higher percentage of Dutch input compared to children with a smaller age difference with their sibling.

In conclusion, these findings show that having a sibling has a negative effect on the target children's Dutch language proficiency, but the effect of having an older sibling was less detrimental for siblings with a smaller age difference, and this positive effect of age difference was larger for target children who received more Dutch input from their sibling. The older sibling might be functioning more as a role model when the age difference is smaller and input and feedback might be better aligned with the target child's language development process, both positively impacting the target child's societal language proficiency.

It should be mentioned that the exact age of the siblings was not known, as the age was only indicated in whole years and not months. This means that a sibling of, for instance, 7 years old could have had an age between 7;0 and 7;11, leading to an 11-month gap of uncertainty in the older sibling's age. It is unclear to what extent this might have had an effect on our findings. Therefore, this should be kept in mind when interpreting the results.

What was not considered in this study is the quality and the absolute amount of the sibling's input, the sibling's exact age, the sibling's sex, the language spoken at home by the parents (and the quantity and quality of any parental input provided in Dutch), and other measures of Dutch language proficiency besides sentence repetition abilities, as this was beyond the scope of the current study. Future research should address these shortcomings, allowing for a more comprehensive overview of how the presence of an older sibling in a bilingual home setting can affect the younger sibling's Dutch language development and what factors are at play here. Additionally, further studies could investigate the effect of language distance and its interaction with other variables explored in this research (e.g., sibling) to determine whether language distance moderates the effect of sibling on language proficiency in the sense that children who acquire two more distant languages have a higher need for input from a sibling than is the case with a less distant language pair. Finally, it would be interesting to see whether similar results are found when other societal languages than Dutch are studied.

In conclusion, we found that having a sibling resulted in lower Dutch proficiency scores, but that a smaller age difference between siblings did improve scores. However, future research should aim at increasing the degree to which other factors that might have influenced these results are considered, as well as showing whether results can be reproduced for different societal languages.

Acknowledgements

We would like to thank dr. Sharon Unsworth for giving us the opportunity to use the data she and her colleagues collected and for helping us throughout the process of writing this paper during the course Child Bilingualism at the Faculty of Arts at Radboud University. Moreover, we want to thank Gerrit-Jan Kootstra, Chantal van Dijk, Figen Karaca, Elly Koutamanis, Jasmijn Bosch, Elise van Wonderen, and Joyce van Zwet, who were also involved in the data collection process. We would also like to thank the editors and two anonymous reviewers for their insightful comments on previous versions of this report.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Bridges, K. J., & Hoff, E. (2012). Older sibling influences on the language environment and language development of toddlers in bilingual homes. *Applied Psycholinguistics, 35*(2), 225–241. https://doi.org/10.1017/s0142716412000379

Duncan, T. S., & Paradis, J. (2020). Home language environment and children's second language acquisition: The special status of input from older siblings. *Journal of Child Language*, *47*(5), 982-1005. https://doi.org/10.1017/S0305000919000977

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of statistical software*, *8*, 1-27. https://doi.org/10.18637/jss.v008.i15

Havron, N., Ramus, F., Heude, B., Forhan, A., Cristia, A., & Peyre, H. (2019). The effect of older siblings on language development as a function of age difference and sex. *Psychological Science*, *30*(9), 1333–1343. https://doi.org/10.1177/0956797619861436

Her, Y. C., Batur, Z. Z., Vergauwen, J., & Mortelmans, D. (2021). Siblings' similarities and dissimilarities: A review of theoretical perspectives and empirical insights. *Journal of Family Theory & Review, 13*(4), 447-462. https://doi.org/10.1111/jftr.12436

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*, 1-26. https://doi.org/10.18637/jss.v082.i13

Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. *Assessing multilingual children: Disentangling bilingualism from language impairment*, 95-124.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org

Tsinivits, D., & Unsworth, S. (2020). The impact of older siblings on the language environment and language development of bilingual toddlers. *Applied Psycholinguistics, 42*(2), 325-344. https://doi.org/10.1017/S0142716420000570

Unsworth, S., Kootstra, G. J., Van Dijk, C., Karaca, F., Koutamanis, E., Van Bosch, J., Van Wonderen, E., & Van Zwet, J. (2022). 2In1: Bilingualism in children [Dataset]. In *2In1 research project*.

Wellen, C. J. (1985). Effects of older siblings on the language young children hear and produce. *Journal of Speech and Hearing Disorders*, *50*(1), 84–99. https://doi.org/10.1044/jshd.5001.84

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. https://ggplot2.tidyverse.org