Exploring the foreign language effect in bi and multilinguals

Sebastian Olsson

Psychology, bachelor's level
2023

Luleå University of Technology
Department of Health, Education and Technology
[This page intentionally left blank]
Abstract

Would your decision-making be different if you used your foreign language? This is a topic that has gained attention in the field of cognitive psychology in recent years. Work today often involves two or more languages and at times requiring processing of information in a language that is not one’s native. This raises a question: does use of a foreign language influence reasoning? The study aims to explore how a foreign language affects reasoning in a bi and multilingual population, looking specifically at age of acquisition, which order participants rank their foreign language and if linguistic distance has associations to performance on a reasoning task. The present study examines fifty Swedish bi and multilingual participants with the languages English and Meänkieli. Unequal years of education makes the results of the reasoning task inconclusive. A possible relationship with the order and age of acquisition, and linguistic distance could be found. Overall, results suggest further research is needed to explore the variables in detail.
Exploring the foreign language effect in bi and multilinguals

A vast variety of professions rely on information presented in a foreign language and make both big and small decisions on a language that is not their native. While most would argue that important decisions are taken rationally it is likely they are affected by biases and heuristics (Stanovich et al., 2008). Understanding when a less biased and more elaborate thought is activated could be beneficial in a setting where a poor decision might be costly. Research has indicated that thinking in a foreign language (FL) can reduce bias and influence decision outcome compared to thinking on your native language (NL). This is known as the foreign language effect (FLe) (Costa et al., 2019; Keysar et al., 2012).

The present study will examine FLe through the perspective of increased deliberation which states that if the task requires more effort to depict, the individual might become more conscious at the task and less susceptible to one’s prejudgement (Frederick, 2005). Language can quickly become a mosaic of confounding variables; age of acquisition, proficiency, order of acquisition and education - just to name a few. In addition to the above-mentioned variables this study will also investigate linguistic distance and in which order of FL the test was taken. The latter has received very little attention in the FLe research to this day.

Bilingualism and multilinguals

In the present study bilingualism refers to speakers of two languages and multilinguals as speakers of three or more languages. Studies of FLe have often occurred in the context of bilingualism, where there is a Language 1 (L1) and a Language 2 (L2). However, there is reason to believe that multilinguals ads to the complexity of FLe by being a very diverse group, for example when it comes to age of acquisition (AoA) and language dominance. This could be the case in countries without a powerful language like English, French or Chinese (Jessner, 2006). Amongst Swedes French, Spanish and German are common optional languages to choose from grade 6 in elementary school. English is a mandatory subject and often taught from the age of ten, it is also a common language in tv, media, music etc. and has a role as a lingua franca. Swedes are therefore often sequential bilinguals in at least English i.e., taught in a school setting but also highly exposed to English which could make it influential in a cultural way as well. Meänkieli is a minority language that derives from Finnish, mainly spoken in the Torne Valley, located along the Finnish border in the northeast part of Sweden. The language
is often taught in a non-academic setting, by family or the surrounding society, giving the language a strong cultural connection, as it is often used within a social context, and showing group identity when spoken (Ridanpää, 2018).

Linguistic distance is complex and there is no exact measurement to describe the actual distance between languages, but languages can be placed in families (e.g., Swedish, Danish, and Norwegian). Languages within families will require less effort to understand compared to a language distant to that family (Jaekel et al., 2023). Another dimension that is examined in the present study is the linguistic distance of bi and multilingual. Following earlier studies in categorization of language families, Swedish and English belong to the Indo-European family, while Meänkieli which is a Finnish language belongs to the Uralic family of languages thus creating a linguistic distance (Körning-Ljungberg et al, 2020).

**Dual process theory**

The phenomenon of FLe has been explained through the dual process theory which divides human taught into two cognitive processes, System 1 and 2. Both are fictious systems (i.e., they have no actual physical part in the human brain) but are used to describe our cognitive processes of thought and decision making (Kahneman, 2011). The parting of the processes into two systems allows for a quantitative way of measuring how we make decisions (Alter et al., 2007). System 1 refers to a quick, automated and intuitive response, System 2 is deliberate, analytical and requires effort (Alter et al., 2007). System 1 demands less cognitive resources than its counterpart and is therefore the initial system to process new information and allows for fast decisions. The trade-off for an effortless and quick decision is that bias and heuristics from prior experience and beliefs will be used for guidance (Stanovich et al., 2008). The function of System 2 is to recognize when a task requires deliberate and analytical thought to be solved. If a task is noticeably difficult System 2 will activate quickly, when difficulty is more elusive System 2 could disregard the task and stay inactivated and thereby enable System 1 (Kahneman, 2011). The activation of deliberate and analytical thought has been studied with different tasks for increased deliberation, but the cognitive reflection test (CRT) might be the most common. It has been shown that System 2 could be trained into recognizing a conflict earlier in the process, education has shown to be a predictor for the ability to resist System 1 and an increase in System 2 (Frederick, 2005). Reips and Stieger (2016) found similar results where higher education predicted accuracy on the CRT, also familiarity to similar tasks predicted a higher score, arguing that reasoning tasks such as the CRT could be highly affected by prior experience and education.
The way a conflict is recognized could give an indication of a person’s cognitive characteristics as either deliberative or intuitive (Capraro et al., 2017), although many factors could have an impact on deliberate and rational thought such as physiological arousal, time of the day, and mood, to name a few (Costa et al., 2019). According to Oppenheimer et al. (2007) understanding when System 2 is activated to overwrite or suppress system 1 is an important factor to understand how we make decisions.

**Foreign language effect and deliberate thinking**

FLe predicts that when using your FL certain types of decision-making will be more deliberate and trigger System 2 or emotionally distant reducing System 1 compared to NL (Stankovic et al., 2021). The FLe has mostly been studied in the context of emotional and moral decision making and framing effects. Framing refers to how phrasing in a positive or negative way creates bias even when the meaning holds the same information in both alternatives. Keysar et al. (2012) found that compared to the NL, FL did reduce bias in the context of framing, because of having a greater emotional distance than NL. Del Maschio et al. (2022b) presented similar results even when the FL was considerably less fluent than the NL. Moral decision-making studies have indicated that when using FL participants are more likely to use utilitarian reasoning, for example sacrifice one to save five. Both framing and moral decision-making are seen through the reduced emotion in the FL hypothesis where age of acquisition could play a big part, a language learnt later in life could have less emotional attachment. Ng & Wong, (2018) presented in a study of moral decisions in Chinese and English bilinguals that age of acquisition could influence FLe, if the language (English) was acquired early in life. The exact factors that would cause FLe in a moral context are still uncertain. Criticism towards moral FLe research has been the lack of standardized ways of measuring FL (see Del Maschio et al., 2022a; Stankovic et al., 2022). Del Maschio et al. (2022b) states that cultural and social norms could play a part but is largely overlooked in research. Champoux-Larsson and Dylman (2017) concluded in their study of boundaries of FLe that cultural influences on FL and language similarity between FL and NL could be important factors in a moral setting.

Increased deliberation hypothesis predicts that if a task is presented in an unclear way and increases the cognitive load, activating the more rational System 2, could serve as a more obvious cue for activation (Stankovic et al., 2022). In a study by Alter et al., (2007) the cognitive disfluency was tested on a reasoning task that was written in either an easy-to-read font or a difficult-to-read font (i.e., fluent, and disfluent condition). The results showed that
participant in the disfluent condition answered more accurately than in the fluent condition, concluding that disfluency did activate a more systematic reasoning (i.e., System 2.)

Language context, as a factor, could increase deliberation due to processing information in an FL, especially when learnt later in life, is often less fluent than of NL and creates a larger cognitive load when used (Costa et al., 2019). The increased cognitive load of a FL would trigger an activation of System 2, leading to increased deliberation. However, in the context of FLe previous research has had difficulty to confirm the increased deliberation theory in a FL context. In a study by Vega-Mendoza et al., (2021) on a task without an emotional component with older participants similar in age and education, performed equally on the CRT, no FLe was found, but the study showed that education and age predicted performance on the task. Mækelæ and Pfuhl (2019) did a similar study with young adults and reported no difference in deliberate reasoning in the FL condition. They too found that significant predictors were age, education, gender, and perceived effort. Interestingly, the study also showed that, when no time limit, proficiency (that would affect the fluency/disfluency paradigm) was not a good predictor of deliberate reasoning.

The study aim

The study aims to explore the FLe in a bi and multilingual population, looking specifically at AoA and in which order participants rank their FL. This will be explored with five hypotheses starting with examining if there is a difference on cognitive reflection between the FL and NL (i.e., a FLe)?

Second, the study wants to explore if perceived confidence has any associations to the actual performance, for example high confidence could be a sign of a more intuitive answer and vice versa (see Mækelæ and Pfuhl, 2019). An additional analysis will be performed on how well the group estimates their correct answers, Coutinho et al. (2021) has shown that overestimating score could be a sign of higher intuitive thinking. The present study will use a similar method but look for differences between language conditions.

The third aspect of exploration is to investigate how proficiency associates to deliberative reasoning (i.e., accuracy on test). Former studies have found that low proficiency tended to make participant read the questions more carefully and slower, but not contributing to FLe (Mækelæ and Pfuhl, 2019), and in a similar study there was no relations to deliberative reasoning at all (Vega-Mendoza, 2021). This is contradicting evidence to the dual process theory that states, higher effort to read would require increase in deliberation (Alter et al., 2007).
The fourth enquiry investigates if the order of FL affects accuracy on test. The order of the FL in multilinguals has received little attention in previous research, it is possible that when measuring English as FL the participant could have it as their L2, L3 or L4, how the order of FL affects FL is could be an important variable to explore (also see discussion by Champoux-Larsson & Dylman, 2017). The possibility that one’s L2 requires a higher effort than L1, and L3 requires more than L2 and so forth.

In conclusion, bi and multilingual are terms that are used quite broadly, and studies often compare languages that are linguistically similar. In the present study we aim to explore if there is a difference on accuracy between the three languages and determine if the FL is present, does it extend with the languages. Since Swedish and English are from the same family there could be an addition of effort that is needed from the linguistically more distant Meänkieli condition.

**Hypotheses 1**

H0: There is no difference in accuracy on a reasoning task between NL and FL group.
H1: There will be a difference in accuracy between the FL and NL group.

**Hypotheses 2a**

H0: Perceived confidence has no significant association to actual performance on the reasoning task.
H1: Perceived confidence has significant associations to actual performance on the reasoning task.

**Hypotheses 2b**

H0: The estimated confidence-score will not differ between the conditions.
H1: The estimated confidence-score will differ between the conditions.

**Hypotheses 3**

H0: Proficiency in the FL has no significant association to performance on the reasoning task.
H1: Proficiency in the FL significantly associates to performance on the reasoning task.
Hypotheses 4

H0: Deliberation will be equal on the reasoning task between languages (L1, L2, L3 and over).

H1: Deliberation will vary significantly on the reasoning task between the languages (L1, L2, L3 and over).

Hypothesis 5

H0: Linguistic distance has no effect on accuracy on the reasoning task.

H1: Linguistic distance will have effect on accuracy on the reasoning task.

Method

Participants

Fifty participants took part in the study, the inclusion criteria of the study were to have Swedish as NL and English and/or Meänkieli as FL. Eight participants failed the inclusion criteria and were excluded, two participants were excluded for inconclusive or contradicting answers on language proficiency. The final sample was composed of the remaining 40 participants. Demographic information and participant characteristics on the remaining participants are presented in Table 1.

Table 1

Participant characteristics

<table>
<thead>
<tr>
<th>Condition</th>
<th>All ( n = 40 )</th>
<th>Swedish ( n = 15 )</th>
<th>English ( n = 15 )</th>
<th>Meänkieli ( n = 10 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, ( M (SD) )</td>
<td>28 (14,4)</td>
<td>45.5 (15,8)</td>
<td>45 (17)</td>
<td>52,4 (13,8)</td>
</tr>
<tr>
<td>Female %</td>
<td>47,5</td>
<td>37,5</td>
<td>60</td>
<td>50</td>
</tr>
<tr>
<td>Education* ( Mdhn (SD) )</td>
<td>4 (0,9)</td>
<td>3(0,9)</td>
<td>4(0,9)</td>
<td>4(0,8)</td>
</tr>
<tr>
<td>Proficiency* ( M (SD) )</td>
<td>4,8 (0,7)</td>
<td>4,7 (1)</td>
<td>3,7 (0,7)</td>
<td>3 (0,6)</td>
</tr>
<tr>
<td>CRT-accuracy, ( M (SD) )</td>
<td>3(1,9)</td>
<td>2(1,4)</td>
<td>3,4(2,3)</td>
<td>3,6(1,6)</td>
</tr>
<tr>
<td>Confidence scores* ( Mdn (SD) )</td>
<td>0(1,7)</td>
<td>1(1,6)</td>
<td>0(1,8)</td>
<td>-1(0,9)</td>
</tr>
<tr>
<td>Completion of survey* ( M )</td>
<td>14</td>
<td>10,5</td>
<td>17</td>
<td>15,5</td>
</tr>
</tbody>
</table>
Table 1 (continued)

Note. \(^a\) Senior high school 3 yrs = 3. University 2 yrs or less = 4.\(^b\) Proficiency is self-rated on a scale of 0–5. \(^c\) Confidence scores was on a scale of -6 to 6. \(^d\) Average time spent on all parts of the survey presented in minutes.

Materials

Survey

The survey was created in Microsoft Forms via OneDrive connected to Luleå University of Technology. Organizations that in some way work with Meänkieli as a language were contacted via e-mail to assist in spreading the survey. Seven organizations were contacted and four agreed to assist in spreading the survey to speakers of Meänkieli. The responding organizations were Nationellt Språkcentrum för Meänkieli (ISOF), Svenska Tornedalingars Riksförbund (str-t), Met nuoret, and Tornedalsteatern (Pajala) which agreed to help using their social media and/or posting the survey in their physical surroundings. The author also used personal contacts to spread the questionnaire. The questionnaire was made for speakers of Meänkieli in three versions (Swedish, Meänkieli and English) with 29 questions (Group 1). A second questionnaire was made for Swedish and English speakers (Group 2) with 27 questions (questions concerning Meänkieli was excluded). Besides the exclusion of Meänkieli questions for Group 2 the survey was the same as outlined above. The language in the questionnaire for demographic information and language background was in Swedish only CRT questions were translated.

Participation was possible by using a QR-code or a link to the questionnaire and could be completed where the participant regarded as appropriate, via smart phone or computer. For randomizing between conditions participants were asked to choose the QR-code or link that had their month of birth below. The months for Group 1 was divided between conditions by adding every third month per condition until all conditions had four in total. For Group 2 same procedure was made but using every second moth until six in each condition.

The cognitive reflection test (CRT)

The CRT is a commonly used instrument to detect heuristics and biases (Oppenheimer & Thomson, 2016). The function of the CRT is to trigger system 1 that will

7
produce a quick and intuitive but wrong answer. Activation of System 2 is required to solve the questions, when or if its activated can give a measure of cognitive processing (Frederick, 2005). The CRT has developed into different versions since the original three questions by Frederick (2005). In the present study the CRT-total had six items. The first three presented in the questionnaire was the original three questions by Frederick (2005) and three additional questions (4, 5 and 9) from Primi et al. (2015) resulting in a total of 6 items.

Questions were chosen depending on how well they would translate linguistically and culturally to Swedish and Meänkieli. Translations from the original language English was made to Meänkieli and Swedish. The questions for the Meänkieli condition were sent to four different teachers who taught Meänkieli in elementary and/or high school. Two of the teachers were unable to assist in the translation. The other two translated all questions and both agreed on one final translation. The Swedish CRT 1 (Frederick, 2005), and CRT 2 (Primi et al., 2015) was translated by the researcher, see Appendix A for translations.

Confidence

Confidence was a subjective measure of the participants perceived certainty over correctness in their answers. Participants chose on a 5-point Likert-scale from, “not confident” to “very confident” after each question. A mean was calculated for each participant and used as an overall indication of the participants perceived confidence on the CRT questions.

Estimated confidence-score.

After completing the full CRT task participants gave an estimate on how many correct answers they thought they had from 0-6. The raw score was subtracted from their number of correct answers in the CRT-total (estimated score-actual score). A positive number indicated over confidence and a negative number indicated under confidence, this method has been used previously in a study by Coutinho et al. (2021) when analyzing overconfidence on the CRT.

Foreign language order of acquisition

Data was collected by letting the participants write in their language in order of acquisition and the age of acquisition on their languages (L2, L3 and L4 etc.) L1 was their NL Swedish for all participants, this information was compared to order of preference and would state which language that was their L2, L3 and so forth. The order of preference was applied as a control to age of acquisition since Finnish and Meänkieli could be hard to separate since
Meänkieli can be seen as a dialect of Finnish to some and as an own separate language to others.

**Language proficiency**

Proficiency was a subjective measure on the languages; Swedish, English and Meänkieli. The participants chose a number on a 6-point Likert-scale from “I don’t know the language” (0), “low proficiency” (1) to “high proficiency” (5).

**Procedure**

The questionnaire started with information about the study and consent to participate and was thereafter divided into three main parts: 1) demographic questions about age, gender, and education. 2) Language background of order and age of acquisition, language preference, proficiency, and language modality. 3) CRT test in either Swedish, Meänkieli or English with confidence and estimated confidence. Participants were asked not to go back, and redo earlier answered questions in the CRT. All questions were mandatory and had to be answered before submitting the survey. After submitting the participant received a “thank you” and confirmation of their participation in the study. The questionnaire was estimated to take fifteen minutes to complete. The average time for completion between the groups are presented in Table 1.

**Ethics**

Participants were informed that all information that is collected in the project is processed in such a way that no unauthorized person can access it. Participants were also informed about their anonymity in the handling and processing of the collected data, and that received answers would be summed up at group level. Contact instructions for further information was presented and where to find the finished research paper when completed. For the reason that the survey was made through a web-based survey tool that was outside Luleå University of Technology matters of reliability were presented. Participation was voluntary and the survey could be stopped and aborted at any time, no information about why was needed. Informed consent was obtained from participants before beginning the survey.
Data analyses

Tests of normality (Kolmogorov-Smirnov) and the skewness/kurtosis values are listed in Appendix B. Age, education was significantly non-normal, therefore differences in the conditions and group FL/NL were analyzed with non-parametric tests. A hierarchical multiple regression was used to further outline the results. Mann-Whitney U test was used for group comparisons when data was not normally distributed. All performed analyses are declared in the result section. All analyses were performed on IBM SPSS v. 29.

Results

Table 1 shows demographic information, language proficiency, confidence score and CRT-score in NL and each of the two FL conditions. In a first set of analyses both FL groups were clustered to into one to be compared to the NL group. There were no significant differences between the groups NL and FL with regards to age ($U = 208$, $z = 0.567$, $p = 0.58$), and gender ($\chi^2 (1) = 1.9$, $p = 0.16$). The groups were not comparable on education NL (Mdn = 3) from FL (Mdn = 4), $U = 258.5$, $z = 2.15$, $p = 0.046$, $r = 0.34$.

Regarding the predictions, hypotheses 1 (H1) stated that a difference in accuracy between NL and FL group in a reasoning task. Results showed that the FL group (Mdn = 4) was significantly more accurate than NL (Mdn = 2) on a reasoning task ($U = 271.50$, $z = 2.38$, $p = 0.018$, $r = 0.37$). When controlling for education on the relationship between the groups and accuracy on test, the following partial correlation was found stating that accuracy no longer had a significant relationship with FL group performance, $r (39) = 0.285$, $p = 0.079$.

Since the group differed in education, and literature has shown education and age to impact on performance on CRT (Brañas-Garza et al, 2019; Frederick, 2005; Vega-Mendoza et al, 2021) a hierarchical regression was performed to further the pattern of results. Although education was non-normal in both groups a hierarchical multiple regression was performed on the variable education as predictors in the first step, the language groups (NL and FL coded as NL=1 and FL=0) in the second step and CRT score as the outcome. Table 2 shows summary of the models and that education and age explains a significant part of the CRT score.
Table 2

Hierarchical Multiple regression with variables education and age on CRT score

<table>
<thead>
<tr>
<th>Model</th>
<th>(b)</th>
<th>SE (b)</th>
<th>(\beta)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1 Constant</td>
<td>-0.745</td>
<td>1.334</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.023</td>
<td>0.019</td>
<td>0.191</td>
<td>0.23</td>
</tr>
<tr>
<td>Education</td>
<td>0.667</td>
<td>0.324</td>
<td>0.323</td>
<td>0.046</td>
</tr>
<tr>
<td>Step 2 Constant</td>
<td>0.371</td>
<td>1.441</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.024</td>
<td>0.018</td>
<td>0.198</td>
<td>0.2</td>
</tr>
<tr>
<td>Education</td>
<td>0.474</td>
<td>0.333</td>
<td>0.229</td>
<td>0.16</td>
</tr>
<tr>
<td>FL/NL group</td>
<td>-1.055</td>
<td>0.594</td>
<td>-0.273</td>
<td>0.084</td>
</tr>
</tbody>
</table>

Note: Dependent variable CRT score. \(R^2 = 0.179\) for Step 1; \(R^2 = 0.245\) for Step 2.

Hypotheses 2a (H\(_1\)) predicted that the perceived confidence had associations to performance on the reasoning task. For the NL group there was no association between confidence on test and performance \(r(39) = -0.069, p = 0.8\). The FL group showed significant association between confidence and performance on test, high confidence predicted a greater score on the test \((r(39) = 0.552, p = 0.004)\) in the FL condition.

Hypotheses 2b (H\(_1\)) predicted that the estimated confidence-score would differ between the conditions. A Kruskal–Wallis test was performed showing a statistically significant difference across the three conditions, \(H(2) = 10.2, p = 0.006\). Pairwise comparisons with adjusted \(p\)-values showed that there was no significant difference between estimated confidence-score between Meänkieli and English \((p = 0.17, r = 0.3)\) or English and Swedish \((p = 0.45, r = 0.23)\), there was a significant difference between the Swedish and Meänkieli condition \((p = 0.004, r = 0.5)\). Participants in the Meänkieli condition were generally under-confident \((Mdn = -1, SD = 1)\) compared to English \((Mdn = 0.0, SD = 1)\) and the Swedish condition \((Mdn = 1, SD = 1.8)\). Additional analyses were made to see if there was a difference between NL and FL group, a Mann–Whitney test showed that the FL group \((Mdn = 0)\) had a lower estimated confidence-score than the NL group \((Mdn = 1)\) meaning that the NL group were more overconfident than FL group \((U = 98, z = -2.55, p = 0.012, r = -0.4)\).
Hypothesis 3 (H$_3$) stated that Proficiency in the FL significantly associates to performance on the reasoning task. The analysis showed there was no association between performance on test and proficiency and fail to reject the $H_0$.

Hypothesis 4 (H$_4$) predicted that deliberation will vary significantly in the reasoning task between the languages (L1, L2, L3 and over). Accuracy was significantly affected if the test was completed with a language that was acquired later in age, $H(3)=8.66, p = 0.34$. A Jonckheere-Terpstra test revealed that the later the language was acquired, the median score on the test increased, $J=384. z = 2.7, p = 0.007, r = 0.43$.

Hypotheses 5 (H$_5$) predicted that linguistic distance will have effect on accuracy on the reasoning task. Analysis with Kruskal-Wallis test showed that there were no significant difference on the reasoning task when taken in English compared to Meänkieli, $H(2)= 5.83, p = 0.054$. To further investigate if there is a trend in the statistics a Jonckheere-Terpstra test revealed that as the linguistic distance increased (Swedish, English and Meänkieli) so did accuracy on test, $J=347. z = 2.14, p = 0.03, r = 0.34$.

**Discussion**

The present study aimed at exploring if confidence in correctly answering the CRT, linguistic distance, AoA and the order of acquisition contributed to FLe. Group differences regarding education was unevenly distributed and makes it hard to draw clear conclusions when accuracy on test was measured. The major finding was that order and age of acquisition could play a part in adding to the FLe. When the tested language had an order of acquisition that was ranked as their L3 or higher the deliberation tended to increase.

Hypotheses 1 predicted differences in accuracy between the FL and NL group. Analyses indicated a significant difference between the groups where the use of FL on test resulted in higher accuracy. After follow-up analyses accounting for education and age results were non-significant. Education seemed to predict accuracy better than the language on test. Due to the non-normally distributed data the parametric follow-up test was not ideal, but it gave further explanations to use in the interpretation of the results. In the collected data there is no support for the hypotheses. These results fall in line with studies by Vega-Mendoza et al. (2021) and Reips and Stieger (2016). The later gave an interesting argument that the CRT might not be an appropriate test for participant with a high or low education since it seemed to predict a high and low score. Further investigation should be made to establish which interval of education is ideal and if there is a difference in the type of academia.
Hypotheses 2a investigated if perceived confidence has significant associations to performance on the reasoning task. Interestingly positive associations were only seen in the FL group, meaning that the FL group in general were more accurate in predicting their correct answers as correct. No significant associations could be seen in the NL group, this could be an indication of added deliberation when using a FL which would be in line with the theory of Alter et al. (2007) and Stanovich et al. (2008). This is only speculative at this point since a negative correlation in the NL was not seen. Also, the measurement of confidence could be an inaccurate instrument to use since the mean of such score could vary greatly from the score of a single answer. According to Coutinho et al. (2021) the result could be more related to the type of errors that participants made rather than FLe, a predicament that this study has not looked in to. An alternative explanation for the result could be that higher education compensated and activated a more deliberate thinking as covered by Frederick (2005).

Hypotheses 2b accounted for the estimated confidence-score, will groups (NL/FL) or conditions (Swedish, English or Meänkieli) differ regarding over and under confidence? For measurement the confidence-score was subtracted by the actual score giving an estimated confidence-score. In the conditions the Meänkieli group was significantly more underconfident compared to the Swedish condition, similar results were seen when comparing NL and FL. The Swedish (NL) condition seemed to be more predisposed to biases when estimating their result compared to the FL group as a whole, and more specifically in comparison to the Meänkieli condition. Linguistic distance could have an effect over how assured a person is when making a decision or evaluating one’s personal performance. Further research is needed to also examine if intuitive and non-intuitive errors are affected due to linguistic distance.

Hypotheses 3 explored the FL group whether proficiency significantly associates to performance on the reasoning task as a variable that would be closely related to score on test. If the participant is high in proficiency less deliberation is needed, then FLe would not occur, and vice versa if proficiency is low, more deliberation is needed. The results were similar to Vega-Mendoza et al. (2021), where proficiency did not associate to performance. This could be explained by the fact that there was no time limit on the test, giving participants time to carefully read through the questions, this was seen by Mækelæ and Pfuhl (2019) where proficiency was correlated to time spent on test rather than performance. When measuring the effect of proficiency on deliberation a key strategy would be to use a time limit.

Hypotheses 4 examined if deliberation would vary significantly on the reasoning task between the order of acquisition (e.g., if the test was taken with their L1, L2, L3 and over).
Results showed that when a participant took the test with a language that was considered to be L3 or higher accuracy on test increased. A pattern was seen stating that language that was further away from their NL also would affect score positively if it was considered L3 or higher. Similar results have been seen in comparisons with simultaneous bilinguals of proficiency in syntax (Hartstone & Li, 2022) but few studies have covered the topic in the context of FLe and deliberate reasoning. The results should motivate for further research investigating specifically multilinguals and the relationship between L1, L2 and L3 or higher.

Hypotheses 5 predicted that linguistic distance would have effect on accuracy in a reasoning task. The study could not confirm a difference in effect on accuracy between the conditions, this indicates that the linguistic distance could not be detected when measuring accuracy and did not add to the FLe. However, there was a pattern showing that participants scored more accurately when in the Meänkieli condition compared to the Swedish condition that was the linguistically most distant language. Despite not statistically significant English compared to Meänkieli contributed to the trend in the results showing an interesting pattern between linguistic distance and accuracy (i.e., further linguistic distance gave greater accuracy). Although the findings are interesting and should encourage further research in the subject, there is a question mark over how the influence of education effected the analysis of the data.

The findings indicate that AoA in combination with linguistic distance are meaningful variables in the research of FLe and deliberative reasoning. This is also an area of the FLe research that has been largely overlooked. Results in the present study should be interpreted carefully, education as a predictor of performance is well established in previous research and the results might be biased by this.

The present study is an important contribution to the FLe research due to its unique disposition of investigating the dynamics between L1, L2 and L3. There is good reason to broaden the field and discover more on what role linguistic distance has on FLe and deliberate reasoning.

Limitations of the study is that the population was quite small in both NL and FL groups, when conditions was measured the population became very small. This is an important aspect since FLe generally has a small effect (Del Maschio et al, 2022a).

Further research is needed to specifically focus the aspect of how the dynamics between FL languages and if linguistic distance plays a part.
References


Appendix A

English version of CRT from Frederick (2005) and Primi et al. (2015)

**CRT 1** (Frederick, 2005)

1. A bat and a ball cost €1.10 in total. The bat costs €1.00 more than the ball. How much does the ball cost?
   [correct answer = 5 cents; heuristic answer = 10 cents]

2. If it takes 5 minutes for five machines to make five widgets, how long would it take for 100 machines to make 100 widgets?
   [correct answer = 5 minutes; heuristic answer = 100 minutes]

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?
   [correct answer = 47 days; heuristic answer = 24 days]

**CRT 2** (Primi et al., 2015)

4. If three elves can wrap three toys in an hour, how many elves are needed to wrap six toys in 2 hours?
   [correct answer = 3 elves; heuristic answer = 6 elves]

5. Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are there in the class?
   [correct answer = 29 students; heuristic answer = 30 students]

6. In an athletics team, tall members are three times more likely to win a medal than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes?
   [correct answer = 15 medals; heuristic answer = 20 medals]

Swedish translation of CRT Frederick (2005) and Primi et al. (2015)

**CRT 1** (Frederick, 2005)
1. En boll och ett slagträ kostar tillsammans 1,10 €. Slagträet kostar 1,00 € mer än bollen. Hur mycket kostar bollen?

2. Om det tar 5 maskiner 5 minuter för att tillverka 5 prylar, hur lång tid skulle det ta 100 maskiner att tillverka 100 prylar?

3. I en sjö finns det en ansamling av näckrosor. Varje dag fördubblas ansamlingen i storlek. Om det tar 48 dagar för ansamlingen att täcka hela sjön, hur lång tid skulle det ta för ansamlingen att täcka halva sjön?

**CRT 2** (Primi et al., 2015)

4. Om tre tomtenissar kan slå in tre leksaker i timmen, hur många tomtenissar krävs för att slå in sex leksaker på två timmar?

5. Jerry fick både det 15:e högsta och det 15:e lägsta betyget i klassen. Hur många elever är det i hans klass?

6. I ett friidrottslag har långa medlemmar tre gånger större chans att vinna en medalj än korta medlemmar. I år har laget vunnit 60 medaljer hittills. Hur många av dessa har vunnits av korta idrottare?

**Meänkieli translation of CRT Frederick (2005) and Primi et al. (2015)**

**CRT 1** (Frederick, 2005)

1. Yks pallo ja yks pelkka maksava yhteensä 1,10 €. Pelkka maksaa yhen euron enämpi ko pallo. Kuinka paljo pallo maksaa?

2. Jos viielä konheela mennee viis minuuttia tehä viis kampetta niin kuinka pitkä aika mennee saala konheela tehä sata kampetta?
kahenkertasesti. Jos mennee 48 päivää siihen ette koko järvä täytyy lumpeenkukkista
niin kuinka pitkä aika mennee siihen ko puolet järvä kasvaa lumpheenkukkia?

CRT 2 (Primi et al., 2015)

4. Jos kolme tonttua käärii pakethiin kolme leikkikampetta tunnissa niin kuinka monta
tonttua tarvitaan käärimhään kuus leikkikampetta kahessa tunnissa?

5. Jerry sai luokan viieksistoista huonoiman toistuksen ja luokan viieksitoista parhaiman
toistuksen. Kuinka monta oppilasta hänen luokassa oon?

6. Fri-itrotlaakissa oon pitkilä urheilijoila kolme kertaa suurempi maholisuus voittaa mitali
ko lyhyilä urheilijoila. Tänä vuonna laaki oon voittanut 60 mitalia. Kuinka monta lyhyttä
urheilijaa oon voittanut mitalin?
**Appendix B**

**Table 3**

*Skewness, Kurtosis and normality tests by group NL and FL*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Skewness (SE)</th>
<th>Kurtosis (SE)</th>
<th>K-S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NL</td>
<td>FL</td>
<td>NL</td>
</tr>
<tr>
<td>Age</td>
<td>0,728(0,6)</td>
<td>0,5(0,46)</td>
<td>-0,63(1,12)</td>
</tr>
<tr>
<td>Education</td>
<td>1,18(0,6)</td>
<td>-0,33(0,46)</td>
<td>-0,73(1,12)</td>
</tr>
</tbody>
</table>

*Note.* K-S = Kolmogorov–Smirnov test with Lilliefors correction. SE = Standard Error. NL = Native language, FL = Foreign Language.

**Table 4**

*Skewness, Kurtosis and normality tests by FL group Meänkieli and English*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Skewness (SE)</th>
<th>Kurtosis (SE)</th>
<th>K-S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Meänkieli</td>
<td>English</td>
<td>Meänkieli</td>
</tr>
<tr>
<td>Age</td>
<td>-0,23(0,7)</td>
<td>1(0,6)</td>
<td>-0,8(1,3)</td>
</tr>
<tr>
<td>Education</td>
<td>-0,4(0,7)</td>
<td>-0,29(0,6)</td>
<td>-1,1(1,3)</td>
</tr>
<tr>
<td>Proficiency</td>
<td>0,09(0,7)</td>
<td>0,43(0,6)</td>
<td>1,5(1,3)</td>
</tr>
</tbody>
</table>

*Note.* FL group is divided by condition.

**Table 5**

*Skewness, Kurtosis and normality tests by test group 1 and 2*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Skewness (SE)</th>
<th>Kurtosis (SE)</th>
<th>K-S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 1</td>
<td>Group 2</td>
<td>Group 1</td>
</tr>
<tr>
<td>Age</td>
<td>0,46(0,5)</td>
<td>0,6(0,5)</td>
<td>-0,9(1)</td>
</tr>
<tr>
<td>Education</td>
<td>-0,2(0,5)</td>
<td>0,5(0,5)</td>
<td>-1,8(1)</td>
</tr>
</tbody>
</table>

*Note.* Group is divided into test group 1 with the Meänkieli condition and test group 2 without Meänkieli condition.