Proficiency in Speaking English as a Second Language: Effects of Bilingual Profile

ALI KALIB ALI

SUPERVISORS

Linda Ruth Wheeldon

Allison Louise Wetterlin

University of Agder, 2022

Faculty of Humanities and education

Department of Foreign languages and translation

Acknowledgements

I cannot express enough thanks to Linda Wheeldon and Allison Wetterlin for your support and encouragement throughout this entire process. I will strive to match your enthusiasm, knowledge, and remarkable mentorship in my future as a teacher.

Thank you to all the other research team members at the experimental linguistics lab at the University of Agder. Special thanks to Eunice Gomes Fernandes for help with the participant recruitment processes.

I further want to thank my family and friends for their encouragement. A special thanks to my mother for the sacrifices she has made to support me and my brothers. Finally, I would like to thank you Lisanne, for the laughs, tears and for believing in me in the years we have shared so far.

Abstract

This study examines how aspects of bilingual profile, particularly learning and exposure in L2 English, relate to aspects of L2 spoken proficiency. Norwegian-English bilinguals completed a questionnaire describing their language background, proficiency, learning, and exposure in their L2 English. Next, participants performed a vocabulary depth task determining correct similar or opposite words of a target word, after which they completed a sentence comprehension task at varying levels of syntactic complexity determining the correct answer to questions referring to either the agent or patient in the sentence structure. Additionally, participants performed one of four novel partially controlled picture description tasks in which they verbally described what they saw. Results showed that learning and exposure to L2 English increased fluency measures, such as speech rate. Moreover, learning by gaming also influences fluency in that it induces shorter speaking time but enhances the compactness of the speech. Moreover, results indicate that L2 English learning and exposure through school does not enhance fluency, while exposure through friends does. Furthermore, exposure to L2 English through family hinders increased proficient fluency.

TABLE OF CONTENTS

1. Introduction and General Structure

This study investigates which factors influence proficiency in speaking English as a second language (L2). The focus of this study is speech production because it has been a neglected aspect compared to other language processes such as reading, listening, and writing (Harley, 2014), but is an important aspect of L2 language proficiency. This study tests L2 speech production, in Norwegian-English bilinguals. In Norway people learn English from 6 years of age at school but Norwegians are also exposed to English in the media and on-line. The aim of this study is to investigate how different aspects of L2 English learning and exposure, relate to different aspects of proficiency in speaking.

Speech production is a complex process, even for those speaking in their first language. Researchers in the field propose that speakers must first match words with the intended message they want to communicate, which then needs to be encoded with correct phonemes for our articulation system to create a meaningful speech stream (Levelt, 1989, 1995, 1999a, 1999b; Levelt, Roelofs & Meyer, 1999). However, these processes can be error prone in that there are mix-ups, retrieval failures, blockage, and the like. Speech production is particularly difficult for those who master more than one language, because they have not one, but two languages to select words from and the corresponding phonemic inventory from both languages they must correctly process (e.g., De Bot, 1992; Poulisse & Bongaerts, 1994; Hermans et al., 1998; Green, 1998a, 1998b; Poulisse, 1999: Costa, 2005; La Heij, 2005; Kormos, 2006; Schwartz, Kroll & Diaz, 2007 and others).

How speakers manage two or more languages has generated a debate on possible advantages (e.g., Peal & Lambert, 1962; Bialystok & Craik, 2010; Adi-Japha, Berberich-Artzi & Libnawi, 2010) and disadvantages of being bilingual (e.g., Smith, 1923, Yoshioka, 1929). Mounting evidence has found the focus of the current study, bilingual spoken production is at an disadvantage (e.g., Gollan & Acenas, 2004; Gollan, Montoya & Bonanni, 2005a, Gollan, , 2005b; Gollan et al., 2007; Portocarrero, Burright & Donovick, 2000; Ivanova & Costa, 2008; Kroll et al., 2008; Runnqvist et al., 2013).

Furthermore, bilinguals vary in how proficient they are in their languages. The interest in our study, spoken proficiency has been measured in a myriad of ways, for example, in terms of fluency (Kormos & Denes, 2004; Ginther, Dimova & Yang, 2010; de Jong et al., 2015), complexity (Foster, 2000; Tavakoli & Foster, 2008) and accuracy (Kormos & Denes, 2004). However, bilinguals differ not only in proficiency, but also in other ways including how and when they acquired or used their languages, and in the amount of exposure they have had in each language (Krulatz, Dahl & Flognfeldt, 2018; Luk & Bialystok, 2013). A number of studies (discussed in detail below) have examined some of these factors (Costa, Caramazza & Sebastian-Galles, 2000; Flege, Yeni-Komshian & Liu, 1999; Moyer, 2004; Baker, 2011; Patkowski, 1980; Klein et al., 1994; Yetkin et al., 1996; Abutalebi, Cappa & Perani, 2005). Nevertheless, there is no consensus about how these factors relate to proficiency in speaking.

The aim of the present project is to investigate the relationship between aspects of language learning, exposure and usage and measures of spoken fluency and complexity. In the chapter below I will start by describing Levelts' (1989) classical theory of speech production. Since speech production is more complex for bilinguals (Costa, 2005), the following section will then contain a review of models of bilingual spoken production and the evidence to support them. Since these models do not address how bilinguals differ and what has been found in relation to these differences, the next section will be dedicated to describing and reviewing some of these effects. Additionally, I will briefly review a recent model that does regard, to some degree, usage, and proficiency differences in bilinguals (Green & Abutalebi, 2013) as the model also supports the bilingual advantage's view. Afterward, I will describe some of the advantages, disadvantages, and proposed hypotheses that attempt to explain the disadvantages-in spoken production. The next section will examine the various way in which spoken proficiency has been measured. I will then describe the current study which involves a questionnaire, two tasks ,and an experiment aimed at capturing bilingual differences and proficiency both through speech and objective measures.

2. Monolingual spoken production

Levelt's classical model of speech production (1989, 1995, 1999a, 1999b) originally addressed monolingual language production but has since been adopted in research and modeling of bilingual language production and comprehension (e.g., De Bot, 1992; Poulisse & Bongaerts, 1994; Hermans et al., 1998; Green, 1998; Poulisse, 1999: Costa, 2005). According to this model, language production occurs in three stages: conceptualization, formulation, and execution (see figure 1). The first stage, conceptualization, determines what content is meant to be said. In this stage, the intended message is generated through *macroplanning* and *microplanning*. The first refers to elaborating the purpose of the communication, while microplanning involves pre-linguistic specification such as deciding on a perspective while conveying the message (e.g., perspective-taking, i.e., theory of mind). Moreover, one considers what information is new and old, which helps determine accessibility status (i.e., previously addressed items in discourse). The outcome of macro and microplanning results in a *preverbal message* that is not linguistic (yet accessible) but contains the necessary information to convert the intended message into meaning and, finally, spoken language.

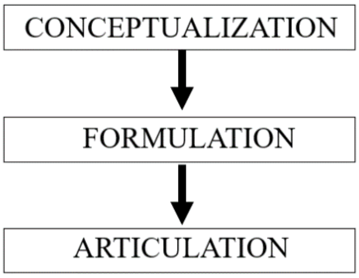

Figure 1: Simplified version of Levelt's (1989) model. Adapted from Harley, 2014: 395

The second part involves formulation. Here the preverbal plan is translated into a linguistic form which entails choosing words, putting those together into sentences, and finally encoding those in articulatory sounds. The process coined *lexicalization* involves retrieving information from a speaker's *mental lexicon,* which, according to Levelt (1989), consists of *lemmas* that contain semantic and syntactic information about the lexical item (e.g., grammatical class, gender) followed by the phonological form of the word, coined *lexemes,* which entail morphological and phonological information about the lexical item (thus it is a two-stage process). Noteworthy, it is assumed that lemmas do not account for modality (i.e., representations are the same across modalities such as speaking, understanding and so forth). Levelt (1989) explains that during *lemma selection,* a speaker retrieves the lemma most suited given the semantic information available in the preverbal plan, the syntax is activated, which in turn activates the syntactic building procedure before morpho-phonological encoding. Given that the lemmas are amodal, it follows that lexical syntax is accessed before the phonological form is accessed.

A range of studies have since found supportive evidence for the two-stage lexicalization model; these include evidence from speech errors (Fay & Cutler, 1977), neuropsychological data (Indefrey & Levelt, 2000, 2004), electrophysiological data (Van Tureout et al., 1998), tip of the tongue (TOT) investigations (Brown & McNeill, 1966; Harley & Brown, 1998), and experimental evidence from psycholinguistic studies (Wheeldon & Monsell, 1992; Starreveld & La Heij, 1995, 1996). For example, Wheeldon & Monsell (1992) examined 106 participants across three object naming experiments after reports of repetition priming of object naming, that is, implicitly strengthening the memory for identification by providing prior stimulus presentation. They found that the repetition priming has long-lasting effects spanning over 100 naming trials. They also found that repeated production of phonological forms (homophones and homographs) was not an effective prime, as such the facilitation could not be phonologically mediated. Instead, it must have been semantic or lemma-based mediation, adding support to the two-stage model.

As mentioned, it is assumed that speech production entails the three differentiated levels of representation, conceptualization, formulation, and articulation (e.g., Levelt, 1989, but also see Levelt et al., 1999; Caramazza, 1997; Dell, 1986). However, the time-course of how speakers progress through these levels is a matter of debate. Two mechanisms have been distinguished in attempts to explain the time-course as they are key across all levels, namely *activation,* and *selection.*

Activation refers to the availability of the representations at the three stages. The degree of activation, either high or low, corresponds to whether the representation is more, or less available for production. In speech production, the first action occurs at the conceptualization stage where conceptual representations are activated. Following the two-stage model it is assumed that not only the semantic representation of the intended concept, but also other semantically related concepts are activated to various degrees. For example (see figure 2), when naming a pictured dog, the target concept (i.e., "DOG") is activated along with other related concepts (i.e., "CAT", and others). The multiple activation of semantic representations spread to the lexical system which in turn activates lexical nodes/words accordingly. If the activated semantic representations flow to the activate the corresponding lexical node, the processes then encounter multiple words for speech production (i.e., "DOG", "CAT", and others). This is where the second mechanism, the selection is employed to decide which lexical node to choose for further processing. The selected lexical node will make its morphosyntax available for further construction of syntactic frames. Finally, the corresponding phonological nodes are activated, resulting in phonological retrieval of /dɒg/

for production.

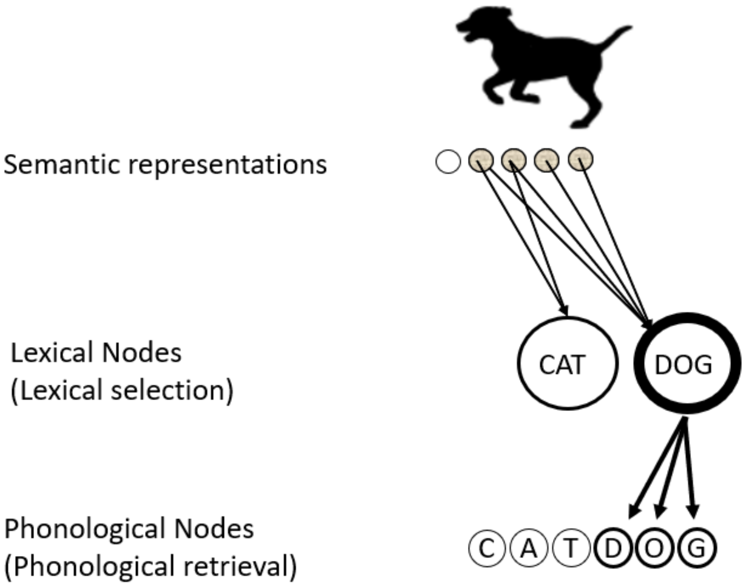

Figure 2: Representation of monolingual speech production. The arrows indicate the flow of activation, and the thickness of the circles indicates the level of activation. Adapted from Costa, 2005: 309

It was assumed thus far that the selected lexical nodes activate proportionally at the phonological stage. However, there is a controversy regarding the activation of the corresponding phonological nodes. The debate concerns whether only the selected lexical nodes or all activated lexical nodes spread activation to the phonological nodes. Two differentiated models have been proposed to explain the activation process: discrete models (Levelt, 1989; Meyer, & Levelt, 1990; Schriefers et al. 1990; Levelt et al. 1999) and cascade models (Dell, 1986; Dell & O'Seaghdha, 1992; Caramazza, 1997; Peterson & Savoy, 1998). The discrete models assume that only the lexical node selected during the first of the two-stages sends activation to the phonological stage. As such, the model predicts that the phonological activation *only* in the further stages of speech production. On the contrary, the cascade models assume continuous activation from the lexical stages, unlimited by the selected node, to the phonological stage. We turn now to speech production processes in bilinguals, where research has provided data relevant to discrete vs cascading activation debate.

3. Bilingual Spoken Production

Like monolinguals, bilinguals must select appropriate words to achieve their communicative goals. However, the fact that bilinguals have not one, but two languages have raised fundamental questions regarding how the speakers mediate the relationship during production, both in terms of activation and selection. For example, if the bilinguals' two languages, are both activated, or is only the intended language activated for further production? If both languages are activated, do they both activate corresponding lexical nodes, and if so, how does the selection processes proceed? There are two models that address these questions, namely language-specific and language non-specific models.

Regarding whether both languages are activated at the conceptual level, the language-specific account assumes that only the intended-language receives activation (e.g., McNamara, 1967; McNamara & Kushnir, 1972). Simply put, the activation flow from the conceptual stage is restricted to only one language, which in turn creates similar situations as described in the monolingual speech production above. However, in contrast most researchers favor the language-nonspecific account (e.g., De Bot, 1992; Poulisse & Bongaerts, 1994; Hermans et al., 1998; Green, 1998; Poulisse, 1999: Costa, 2005; La Heij, 2005; Kormos, 2006; Schwartz et al., 2007). The language non-specific activation model assumes that activation from the conceptual system flows to the lexical representations of both languages. An often-cited study to support this account is from Hermans et al. (1998) picture-word interference inquiry.

Hermans et al.(1998) investigated the activation of the non-intended language during production. To do so, they examined Dutch-English bilinguals in whether their L1 equivalent translation (e.g., "berg" meaning *mountain*) is activated when asked to produce a word as fast as possible during picture naming in their L2 (e.g., English "mountain"). Distractor words were simultaneously played auditorily (L2 distractor in experiment 1 and L1 in experiment 2). These words were manipulated to follow a set of conditions. For the naming of a mountain picture in Experiment 2, the conditions were as follows; first, a phonologically related word (e.g., *Dutch* "*mouw*" meaning ‘sleeve’), second, a semantically related word (e.g., "*dal*" meaning ‘valley’), third. an unrelated word (e.g., "*kaars*" meaning ‘candle’), and fourth a word which is an orthographic neighbor of the Dutch translation (e.g., "*berm*" meaning ‘verge’ instead of "*berg*"). Hermans et al. hypothesized that in the final condition, hearing "berm" would activate the translation equivalent "berg" and thus delay the activation of "mountain." Additionally, they also manipulated the stimulus onset asynchrony (SOA, i.e., onset of interrupting words) to determine speed influences *and* the time course of lemma and lexeme selection.

Results showed that the fourth condition (e.g., *berm*) resulted in the slowest activation of the correct "mountain" at SOA 0ms (simultaneous interruption) while the third condition (e.g., *kaars*) was the fastest. When the SOA is at -150 ms (150 ms after the picture is displayed), the first condition (e.g., *mouw*) had a faster correct answer activation, while the second condition (e.g., *dal*) was the slowest. What could be understood from these results, is first, that the question of activation is a time-sensitive process. Second, the results show that the non-selected equivalent was activated during what is the lemma selection in Levelt's model but not during lexeme selection. These results were taken as evidence of a language non-specific activation. However, Hermans et al. (1998) were soon criticized for observers' paradox in that they unwittingly activated Dutch by providing Dutch stimuli in the interrupting conditions. Nevertheless, a subsequent study by Hermans et al. (2011) that controlled for this replicated the results. As mentioned, most researchers are in favor of the non-specific activation account, in fact, a consensus has since been reached on the matter (Ishikawa, 2018). Less agreed upon, however, is how words are selected when both the intended and unintended languages are activated.

There are two theoretical views regarding how lexical selection occurs: language-specific selection and language nonspecific selection models. These models differ with respect to whether the selection mechanisms are interfered by the semantically activated lexical items from unintended language. Language-specific selection models assume that the selection mechanisms are blind, or ignorant of the activation of the unintended language. As shown in figure 3 b), the language-specific selection model predicts that only intended-language lexical items (e.g., “CAT”) are potential competitors for the intended word (i.e., “DOG”), meanwhile, the Spanish translation equivalents (e.g., “PERRO” and “GATO”) are being blindsided and thus have no impact on the selection processes. In this sense, bilingual lexical selection through language-specific models may be considered similar to the monolingual lexical selection displayed in figure 2. Additionally, as shown in figure 4, the activation of lexical items from the unintended language is ignored regardless of the activation degree. However, most researchers seem to favor the language-non specific models (e.g., Hermans, 2000; Hermans et al., 1998; Caramazza & Sebastián-Gallés, 2000; Schwarts et al., 2007). The language non-specific models assume that the activated lexical items of *both* languages are in a selection competition. As shown in figure 3 a) selection takes place by considering the activation of all lexical items across both languages (i.e., “DOG” ,“CAT” *and* “PERRO”, “GATO”).

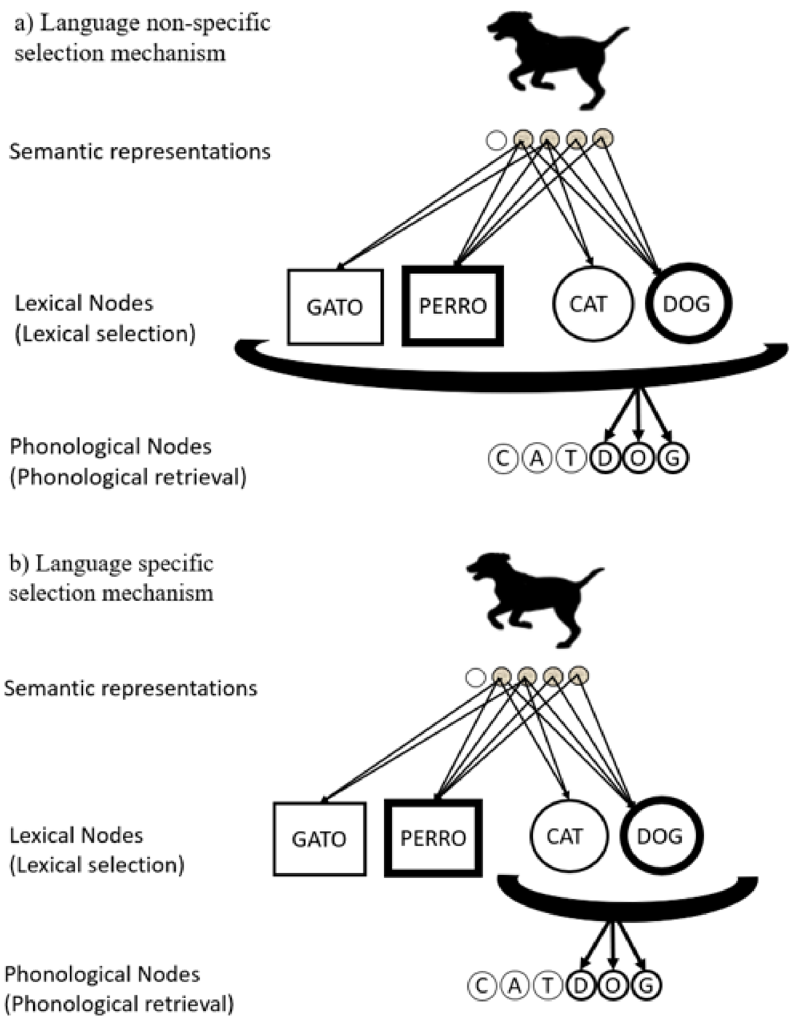

Figure 3: Models of bilingual selection Adapted from Costa, 2005: 314-315. Squares indicate Spanish lexical nodes, while circles indicate English lexical nodes. The thickness of the letters indicates activation levels. Arrows indicate activation flow.

Each model, language-specific or non-specific addresses the process of lexical selection in differentiated manners, while the latter assumes competition across both languages, the former assumes that the non-intended language does not interfere as the activation is language-specific. However, when viewing these models, an unanswered variable is how the selection mechanisms restrict itself to the intended language. In the language-specific model, it is a necessity to specify how the selection mechanism omits searching for lexical nodes beyond the intended language. The language non-specific models must address how the selection mechanism prevents the non-intended lexical nodes from selection as opposed to the intended nodes. Simply put, for either model to work, there is a concern about whether the bilingual lexical selection is sufficient on the basis of conceptual activation alone or in the case of dual concept activation, what mechanism prevents erroneous selection.

One solution presented for the language-specific models is found in Levelt, Roelofs and Meyer (1999), namely the binding-by-checking mechanism. This mechanism is postulated to ensure that selected words match the intended communicative goal of the speaker. Critically, the mechanism is sensitive to both languages and the language of the selected lexical node. Thus, when the language of a selected lexical node does not match the intended language, the checking mechanism notices the error and notes a mismatch resulting in the unintended node being discarded. Unintended intrusions are explained as a result of two errors in the selection mechanism. First, the lexical node selected belongs to the unintended language, and second, a failure must occur in the checking mechanism that binds the target conceptual representation to the intended target language with a proportionate lexical node.

Two solutions are proposed for the non-specific models on how the selection of *only* the intended language and not the unintended languages occurs. The first of these assumes that the semantic system itself activates the lexical representation of the intended language with a higher activation due to language cues (Poulisse & Bongaerts, 1994; Poulisse, 1999, Hermans, 2000). The second solution refers to the existence of inhibitory control processes acting on the unintended language. The activation of lexical nodes are, in other words, suppressed, allowing the intended lexical nodes to be *more* activated than the unintended (Green, 1998). The first of these solutions derive from a model of bilingual speech production by Poulisse and Bongaerts (1994).

Poulisse and Bongaerts' (1994) model which was since extended by Hermans, 2000 (figure 4) employs Levelt's (1989) conception of language processing that involves first, the conceptualizer, which results in a preverbal message. However, additionally, the language cue feature is assumed at the conceptual stage (+ English in figure 4). These language cues determine the input prior to lemma selection. In short, Following the preverbal message, a set of lemmas are selected, however, these are simultaneously tagged for a language. Given that a set of lemmas are activated, the model assumes a language non-specific activation model and a spreading activation mechanism. The non-intended but now active lemmas are at this stage competing with intended lemmas for selection (i.e., non-specific selection). The most activated lemma is chosen because it has additional language cues strengthening its activation (i.e., the concept of a bike, +English and lemma "bike" results in bike).

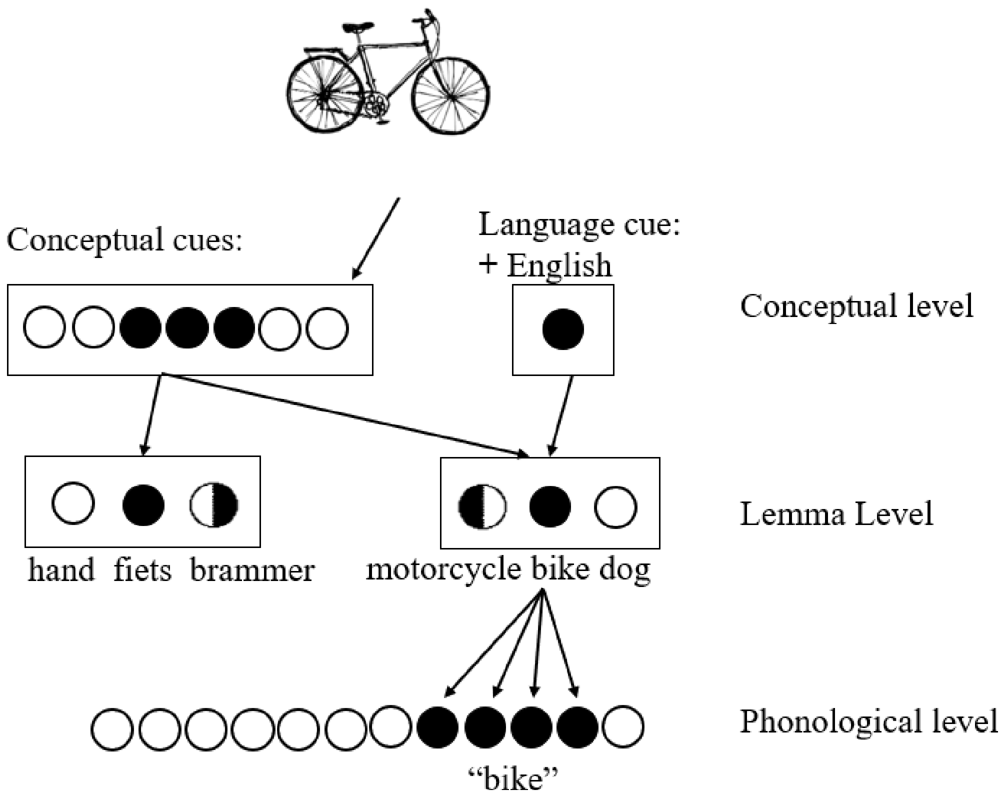

Figure 4: Model of bilingual spoken word production modelled on Hermans, 2000 and Poulisse & Bongaerts, 1994 Adapted from Kroll & Tokowicz, 2005: 539

The second solution, namely through inhibitory processes, is linked with Green's (1993, 1997, 1998a, b) Inhibitory Control Model (IC, see Figure 5). The model describes how bilinguals control production by selection and inhibition processes. As Figure 5 shows, there are multiple components; of these the *conceptualizer, SAS, language task schemas* are related to processing, while the *bilingual lexico-semantic system* is a linguistic component.

Like the other production models described, Green adopts Levelt's model, where the production begins at the conceptual stage. The output of the conceptualizer spreads to other areas, such as the language task schemas. The latter, in Greens model, refers to L2 production schemas. These schemas are particular action sequences appropriate to reach the goal of the action they are selected for, e.g., L1 to L2 translation. The SAS, short for supervisory attentional system, controls and establishes the activation of the task schemas to reach the Goal. For example, if a Norwegian bilingual is presented with the Norwegian word *jakke* it may elicit task schema activation "L2 translation" (or a range of other schemas such as selecting synonyms, antonyms, and more), resulting in

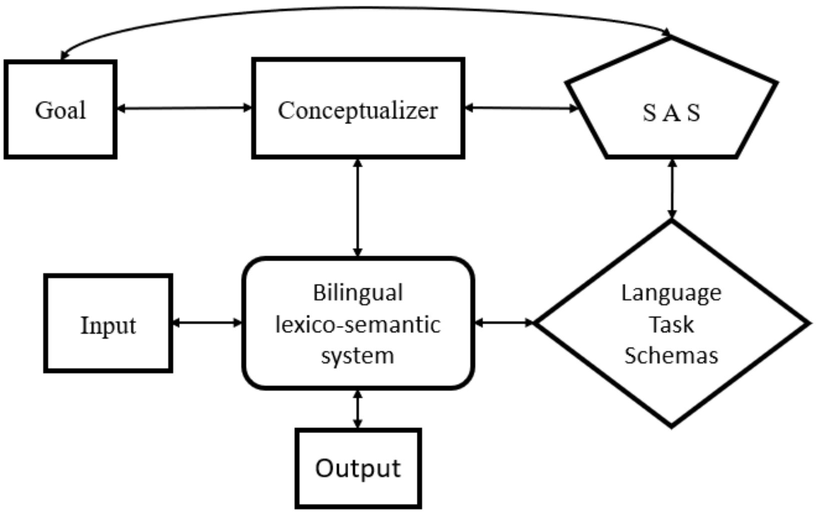

Figure 5: The inhibitory control model Adapted from Green (1998). SAS refers to the supervisory attentional system.

the English equivalent/translation *jacket*. Effectively, this makes the processes competitive. Therefore, to reach the goal, (*jacket* in this case) L1 production schemas are assumed to be inhibited/suppressed. In short, the L2 translation schema boosts the L2 production schema while simultaneously actively inhibiting the L1 production schema. Green (1998a, b) argues that the processes described occurring regardless of the linguistic or lexico-semantic processes. Thus, Green names the task schemas as *functional control circuits* that are an output of the conceptualizer and executed through the SAS

Regarding lexical selection, Green's IC model assumes that each lemma has a language tag for either language, which is specified during a language-independent conceptualization. Moreover, he assumes that lexical selection occurs at the lemma stage. Thus, each lexical node is associated with a language tag which helps inhibit the non-target lexical nodes tagged by the unintended language (see figure 6). In short, three critical features of the IC model are as follows. First, the semantic system activates the lexical nodes in both languages. Secondly, the lexical nodes of the non-intended language are inhibited reactively, and thirdly, the inhibition is proportional to the activation levels of the non-intended lexical node (i.e., more robust semantic activation of the wrong language equals stronger inhibition). The IC model assumes that the L1 is more strongly activated than the L2, therefore, inhibition will be more potent when the L1 is the non-intended language. The opposite will be valid for the L2 (weaker inhibition).

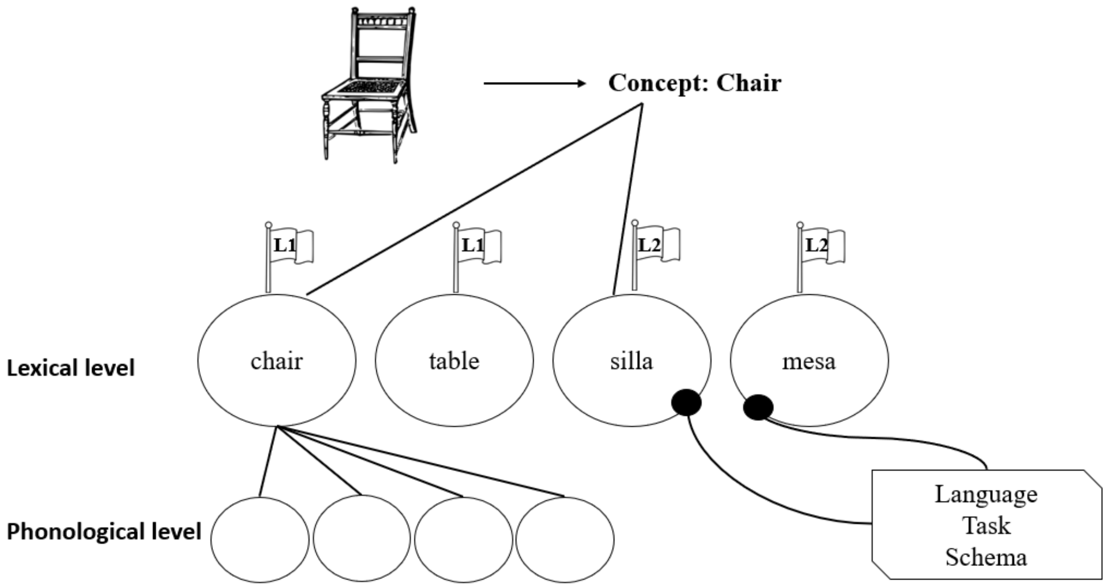

Figure 6: the IC model (Green, 1998) in action adapted from Finkbeiner, Gollan & Caramazza (2006). L1 English and L2 Spanish. The task schema connections inhibit the L2 lexical nodes in the non-intended L2.

A range of studies have since reported empirical data in support for the IC model (Meuter & Allport, 1999; Costa & Santesteban, 2004; Schwieter & Sunderman, 2008; Schwieter, 2010, 2013; Linck, Schwieter & Sunderman, 2012 and others). All report that when participants are asked to switch between languages, the switching cost into the dominant L1 takes longer than their L2 (or L3). The earliest of these, Meuter and Allport (1999) also investigated switching costs in an experimental study where bilinguals named numerals in either their L1 or L2. Participants were asked to name the numerals either in a trial with the same language (no switching) or to name the other language than the prior trial. Meuter and Allport reported that response latencies on switching trials (response language changed from the initial trial) were slower than in the non-switching trials. The results support the IC model in that the switching cost associated with switching to the dominant L1 was consistently slower than switching from the weaker L2. The authors argue that the inhibition of the L1 is a result of strong inhibitions of the dominant L1, as the prior trial required it for naming in the weaker L2. Simply, Meuter and Allport suggested that the naming in L2 requires active inhibition of the more active competitive L1 nodes. These results were later replicated in Linck et al. (2012), where they examined the performance of multilingual (trilingual in this case) in both the Simon task and a switching task.

Another question relates to how the phonological repertoire of a bilingual's two languages are represented. For example, in the case of a language-specific mechanism, two separate phonological inventories would be necessary to presuppose a retrieval mechanism sensitive to the activation of only one phonological inventory In the contrary, if there is any certain overlap between the phonological systems of a bilinguals two languages there must be a measurable impact of the phonological information from the unintended language on the retrieval of phonological information for the intended language. While there is no consensus on the matter, the study reviewed earlier by Hermans et al., (1998) did indicate that activation flow was not language-specific. The results have also been interpreted to suggest that lexical selection mechanisms consider the activation of the unintended language during lexeme selection.

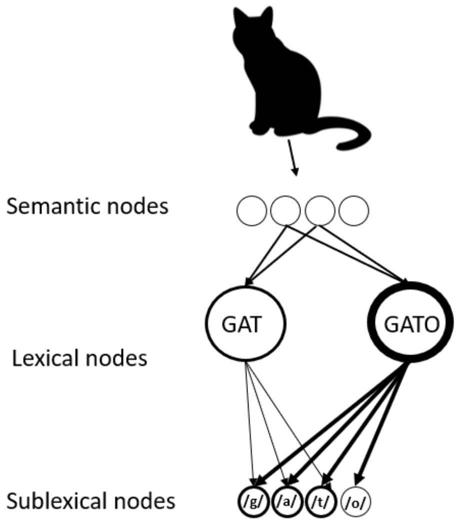

Figure 7: Representation of lexical and sub lexical access for cognate words Catalan "*gat*" with Spanish "*gato*". Arrows indicate activation. Thickness of the circles indicates levels of activation of sub lexical nodesAdapted from Costa et al., 2000: 1285

Costa et al. (2000) addressed how the phonology is represented from a different perspective. They explored whether the cognate status would impact the speed at which they are produced. Cognates are, at least in the psycholinguistic research domain, translations with similar orthography and phonology (e.g., the Catalan and Spanish words for 'cat', *gat* and *gato* respectively). Non-cognates are translations with dissimilar orthography and phonology (e.g., Spanish *pandereta* and '*tambourine*'). Costa et al. (2000) hypothesized that the selection of phonological representations during picture naming of the target translation (e.g., *gato*) is achieved quickly due to cross-language activations at the phonological stage. Simply put, retrieval of phonological properties of the target word is easier for cognates than for noncognate words due to phonological features that overlap across cognates. For example, the

Catalan *gat* (i.e., /g/, /a/, /t/) and Spanish (/g/, /a/, /t/, /o/) share segments as illustrated by thickness at the sublexical nodes as shown in figure 7.

Adhering to the cascading activation model, Costa et al. (2000) made two predictions. First, that the two languages of a bilingual share a common semantic system, and second, that there is parallel activation between the languages. To their predictions, Costa et al., conducted two investigations on Catalan-Spanish bilinguals and Spanish monolinguals through a series of picture naming tasks. The picture names were either cognates or noncognates. To render their predictions, correct, bilinguals should name cognates more quickly than non-cognates. In the first of two, experiments participants were separated into two groups, including highly proficient Catalan-Spanish bilinguals and monolingual Spanish speakers. All participants, regardless of the group were asked to name one of two sets (one of these sets contained cognate names) of pictures in Spanish. Results revealed that the bilingual speakers had faster naming latencies when naming cognates in contrast to the monolinguals. The results supported their theory in that firstly, activation flow from the semantic system to the lexical system is non-specific and possibly shares a semantic system, and secondly, that lexical nodes form the non-intended language spread activation to the corresponding phonological nodes indicating parallel activation.

However, given the bilinguals were Catalan-Spanish speakers responding in their L2, the intended language, Spanish, Costa et al. (2000) further investigated, in a second experiment, whether the cognate facilitation corresponded proportionally to the level of activation in the non-selected lexical node. The second experiment compared the performance of the Catalan-Spanish bilinguals who reported a dominance for the Catalan language and highly proficient Spanish-Catalan bilinguals with Spanish as the dominant language. The number of pictures to be named was doubled. Results confirmed the findings of the first experiment. Moreover, Costa et al.(2000) found that cognate facilitation related to the language of response, being larger when the bilinguals responded in the non-dominant language. The second experiment provided further support for parallel activation at the phonological stage and language-nonspecific activation. These results have since been replicated in picture naming (e.g., Kroll et al., 2000), but also in TOT investigations (e.g., Gollan & Acenas, 2000).

Kroll et al., (2000) for example, asked bilinguals to name pictures depending on cues which indicated which of their languages to use for naming each picture (mixed condition) or to name pictures in one language alone (blocked condition). Critically, the authors manipulated the cognate presence in the conditions. Kroll et al. reported cognate facilitation in the mixed condition regardless of language used, however the cognate facilitation was only found in the blocked condition when naming in the L2. In this case the cognate effect is visible in the blocked condition because the unintended phonology was activated given the L2/L1 cross language activation. When bilinguals were alternating between their languages in the mixed condition, the alternation induced cross language phonology activation regardless of the language used.

To explain the stronger facilitation when bilinguals responded in their non-dominant language, Costa et al. proposed that it is a result of translation from a dominant language to the weaker language. They explain that the dominant language has stronger connection to the semantic representation and the corresponding lexical node than the weaker language, an argument addressed earlier by Kroll and Stewart (1994) in their model of language processing, The Revised Hierarchical Model (RHM).

Kroll and Stewarts' RHM (1994, see figure 8), like previously reviewed production models, approaches language processing in terms of conceptual stage. Critically, however, this model assumes that L1 words are more strongly connected to the conceptual stage than L2 words. However, the L2 words are strongly connected to the translation equivalent in the L1, as shown in figure 8. These differences are, according to Kroll and Stewart, a result of bilinguals' proficiency. For example, the late age of acquisition (AoA) of the L2 in bilinguals corresponds to having a fully developed lexicon for their L1. The RHM assumes that during the acquisition of the L2, bilinguals exploit the notion of L1 connections with the conceptual stage to acquire new lexical entries. As such, the bilingual speaker develops a strong network from the L2 to L1, which continuously matures under L2 acquisition. However, as the bilingual becomes more proficient in the L2, the need to mediate through the L1 dissipates, albeit *never* completely gone. Only speakers who attain (or start with) balanced bilingualism are believed to have equal L1/L2 to concept-level connections.

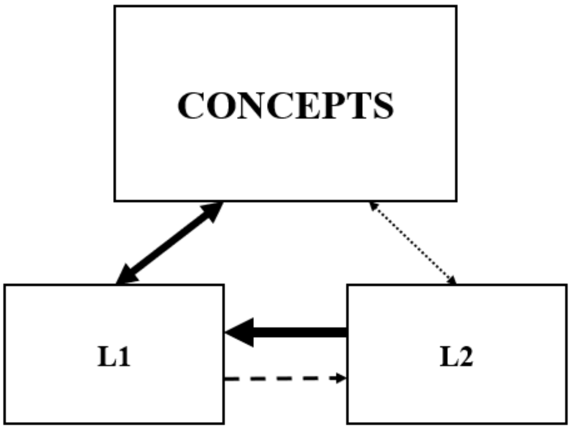

Figure 8: Modelled on Kroll & Stewart (1994)

Given that the connections are as described, translations from the L1 to the L2 in forwarding translation (i.e., from the source language to the intended target language) must be mediated at the conceptual stage. The opposite would be true for backward translation L2 to L1, as that would require a direct translation. Kroll and Stewart tests this hypothesis by investigating whether forward translation takes longer than backward translation. Additionally, since proficiency affects the mediation, one might expect a decrease in direct translation from the L2 to the L1 instead of allowing more access directly to the conceptual stage.

Kroll and Stewart asked proficient Dutch-English bilinguals to do both forward and backward translation (L1 to L2 and L2 to L1, respectively). Moreover, they manipulated the translation lists according to semantic relations, thus, one list was semantically mixed and the other categorized (e.g., all animals). They found that the L1 to L2 translations were slower in the categorized lists, while the L2 to L1 showed no difference in the same list. As such, it seems that the L1 to L2 translation was conceptual stage, semantically mediated, confirming their predictions. The same results were replicated in another study by Talamas, Kroll and Dufour (1999), who examined bilinguals with varying degrees of proficiency. In their low-proficiency bilingual pool, the L1 to conceptual level mediation was stronger than for the more proficient group. Further studies found more evidence in support of the RHM (e.g., Kroll, et al., 2002); some varied in their findings (e.g., Francis et al., 2003; Sunderman, 2002), and others found opposing results (De Groot & Poot, 1997).

What we can take from Kroll and Stewarts'(1994) investigation is as follows. First, bilinguals L1 has a stronger connection to the conceptual stage. This aspect confirms Costa et al. (2000) proposal that bilinguals who use their non-dominant language in cognate naming have stronger facilitation because translation occurs through the L1 with a stronger conceptual connection. However, the latter would dissipate with increased proficiency in Bilinguals. The proficiency was also connected to another aspect of bilingualism, namely AoA, in that it affected the proficiency levels. A series of investigations examined AoA and other aspects that may vary in Bilinguals, which will be the topic in the following chapter.

4. Bilingual profile and fluency in L2.

An early definition of bilinguals by Roelofs (2003, p. 175) is "Bilingual speakers are persons who regularly use two or more languages for their verbal communication." Following Roelofs' definition, any person can be considered bilingual simply because of their knowledge of two or more languages for communication in regularity. However, the required degree of knowledge in the respective languages is not mentioned. In contrast, Luk and Bialystoks's (2013) investigation of quantifiable bilingualism claims that bilingualism is not a categorical variable. Luk and Bialystok (2013, p. 605) argue that "the criteria that determine an individual's designation as monolingual or bilingual are fuzzy at best" and that these criteria minimally involve language proficiency and usage. Luk and Bialystok argue that any inquiry into bilingualism should not be categorical but reflect a varied range of possible bilingualism.

Krulatz et al. (2018 but also see Grosjean, 1998) distinguish a range of bilingualism types. These include *simultaneous* bilingualism (i.e., both languages are acquired simultaneously), *successive* bilingualism (languages are acquired in succession), *balanced* bilingualism (i.e., competence is equal across both languages), and *dominant* bilingualism (opposite of balanced). One may also distinguish between *early* or *late* bilinguals, which describes whether acquiring the second language was earlier or later in life (for an expansive overview, see Krulatz et al., 2018). In the case of Norwegian-English bilinguals, which are the focus in this study, also vary regarding these types, however formal instruction in English starts at 6 years of age. This means that if they are successive bilinguals’ chances are chances are they share a common AoA in English.

Unsurprisingly, there is evidence that differences in aspects of bilingual profile affect language ability (Costa et al., 2000; Flege et al., 1999; Moyer, 2004; Baker, 2011; Patkowski, 1980; Klein et al., 1994; Yetkin et al., 1996, Abutalebi, 2005 and others). For example, Costa et al. (2000) suggested that later AoA resulted in lowered proficiency and thus later naming of cognates. AoA can be traced to the *critical period hypothesis* (CPH). CPH is often linked to Noam Chomsky's innatist perspective in which he hypothesize that all animals, humans included, have innate abilities at birth that aids the acquisition of knowledge (see Rowland, 2014). Albeit in congruence with the innatism perspective, CPH adds that this knowledge is acquired in specific periods and that acquisition beyond these periods is difficult. Studies examining CPH found that older learners (later AoA) typically display non-native accents (e.g., Flege et al., 1999; Moyer, 2004; Baker, 2011), indicating the acquisition beyond a specific period hinders native-like phonological proficiency. Hansen-Edwards (2017) describes studies that have found evidence suggesting that the non-native phonological proficiency, despite CPH, may be ameliorated, instead suggesting that other factors may influence linguistic features, regardless of AoA. Given that Norwegian-English bilinguals have little variation in their English AoA, investigating linguistic features in this group may provide insight into which other factors have an effect on linguistic proficiency.

An early study by Patkowski (1980) aimed to explore all relationships between AoA and linguistic knowledge other than pronunciation, specifically syntax acquisition. Patkowski hypothesized that those acquiring a language before 15 years of age would more likely achieve a native-like skill level in the respective language. Albeit his main goal was to discover whether there is a difference between those who learn English before puberty versus those who acquired English later, he also examined other factors that may influence performance and mastery, namely amount of formal/informal instruction, residence time, and exposure to the language in the US. He examined 67 participants who all shared a higher level of education. All participants were immigrants to the USA and had lived in the country for more than five years. However, all participants had started acquiring English at various ages. To compare, he studied 15 participants born in the US with education levels that reflected the other participants. Critically, the dialect of those 15 participants reflected the type of English spoken by the 67 participants. All participants took part in the one task employed.

Participants took part in interviews that lasted between 15 and 35 minutes. From these, 5-minute samples were collected and transcribed (thus avoiding pronunciation measures). Two ESL teachers were given the transcripts to rate on a scale from 0 to 5 (0 = no knowledge of English, 5= level of English expected from native speakers).

The results showed that the transcripts of the all the native speakers, and L2 leaners (except for one) with an AoA before the age of 15 (named pre-puberty group) scored between 4 and 5. For the group (34 participants) who had their AoA after the age of 15 (post-puberty group), the scores contrasted the prior group with ratings ranging from 3 to 4, including one rating of 2. Moreover, length of residence and amounts of formal instruction correlated with ratings in which more extended residence and more instruction meant higher ratings. However, the latter results were explained as a result of age. For example, while the length of residence was found to be a good predictor of ratings, it was often the case that those with lengthier residence had arrived earlier, resulting in an earlier AoA. Likewise, more instruction was found to predict better ratings, but often in speakers who started formal instruction at an earlier AoA. As such, when the non-AoA factors were controlled for AoA, they no longer predicted the ratings as well as AoA. Patkowski concluded that AoA is a crucial item for developing native-like language mastery beyond pronunciation.

While Patkowskis' findings have since been replicated in other psycholinguistic studies (see Johnson & Newport, 1989; Birdsong, 1999; DeKeyser, 2000; DeKeyser et al., 2010; Granena & Long, 2013) little inquiry into whether the non-AoA factors predicted better proficiency rating when the AoA was equal in across participants did not find place, leaving a gap in the literature.

Researchers investigating bilingual L2 acquisition have since explored the topic in neurophysiological studies as it has the advantage of directly imaging the bilingual brain (see Perani & Abutalebi, 2005 for a review). Klein et al., (1994) used PET (Positron Emission Tomography) to investigate whether L2 production involved the same neural substrates as L1 production. The researchers used a word repetition task on twelve Canadians who were native English speakers who learned French after 5 years of age (mean age 7.3 years). All participants had a high level of language proficiency established through pre-tests. Klein et al. (1994) found that the rCBF (regional Cerebral Blood Flow) distribution was the same across both languages, suggesting that the repetition task involves overlapping neural substrates. Critically, one difference was found on the left putamen region of the brain which has been found to be related to articulation and phonetic coding (Harley, 2014), was activated to a larger extent when participants repeated words in their L2. As such, Klein et al. concluded that articulating L2 speech units, when the L2 is learned later in life, increases the demands and thus activates the putamen to mediate the increased processing load. However, since lexical and semantic access is limited in repetition tasks, Klein et al. conducted a second study.

Klein et al. (1995) undertook the subsequent language production study with the same participant pool as the previous study. Like the earlier study, Klein et al. employed PET to measure rCBF. The participants were asked to selectively produce words for the L1 or L2 following three strategies. First, rhyme generation based on phonological prompts, second, synonym generation, which activates semantic-lexical search, and finally, translation which requires activating access in the other languages. Irrespective of the results found in the 1994 study, the 1995 study tasks showed significant activation of the left dorsolateral frontal cortex and other frontal areas of the brain. Like the previous study, there was overlap irrespective of language, and they also, again, found increased rCBF of the left putamen area when subjects generated words in the L2. rCBF increases manifested particularly in the forward-translation word selection. In the activated areas, the left inferotemporal and superior partial cortex were equally active irrespective of the languages and strategy except in the rhyme condition for L2. The PET showed that both forward and backward translation correlated with increased rCBF, particularly in the left putamen with the forward L2-L1 translation.

Klein et al. (1995) concluded that the activated brain areas were engaged in the highly proficient bilinguals despite late L2 AoA (5 years of age). Thus, languages learned later would not differ in brain activity from ones acquired earlier in life, when the bilingual is highly proficient. This would mean that at least for highly proficient bilinguals, AoA is not a crucial determiner of brain activity, an interesting aspect given that Norwegian-English bilinguals often share an AoA of 6 years of age.

Another fMRI study (Yetkin et al., 1996) also tested L2 production through word generation; however, the participants in this study were multilingual. Reportedly the participants were heterogeneous in that they were *fluent* in at least two languages and *non-fluent* in a third. *Fluent* was defined as languages spoken currently and at least for the past 5 years, while non-fluent was defined as languages studied between 2-4 years and without regular everyday use. One interesting result in this study was the extended brain activation for the L3 (*non-fluent*). This result can suggest effects of exposure on the production of a language. Notably, neither proficiency nor AoA were employed as controlled variables.

Abutalebi, Cappa and Perani (2005) report two further fMRI (functional Magnetic Resonance Imaging) studies that employ language production experiments which found contrasting evidence on the importance of AoA. While Kim, Relkin and Hirsch (1997 as cited in Abutalebi et al., 2005) conducted a sentence-generation task where participants were asked to describe events internally. Kim et al. concluded that AoA indeed is a major factor in the second language spoken production. Chee, Tan and Thiel (1999 as cited in Abutalebi et al., 2005) conducted a different study in which participants were asked to produce words which were cued by a word stem visually presented. Chee et al. found no difference in brain activation in early (6 years) and late (12 years) bilinguals. Abutalebi (2005) explains that this discrepancy might be related to participants' differentiated levels of proficiency across languages, as the participants in Kim's study may have had low proficiency while the participants in Chee et al. are expected to be highly proficient given the socio-linguistic background of an integrated bilingual society. Like Kim and Chee's contrasting evidence, other inquiries in language production studies (Perani et al., 2003; Briellmann et al., 2004; Köpke et al., 2021) and, conversely, language comprehension studies (see Abutalebi et al., 2005, pp. 506-510 for a review) also find contrasting evidence in the brain activity from L1 and L2 production/comprehension given the AoA. As such, these discrepancies seem to be related to other factors that modulate activation, namely *proficiency* but also exposure, and possibly others. Despite AoA being an unquestionably important determinant of proficiency in L2, Abutalebi et al. (2005) emphasize that brain activity in bilinguals are a result of the interplay deriving from other factors such as exposure and proficiency (e.g., Hansen-Edwards, 2017; Patkowski, 1980; Klein et al., 1994; Yetkin et al., 1996, Abutalebi et al., 2005).

Indeed, bilingual exposure is a key factor in Green and Abutalebi (2013) model that takes exposure and language use into account, namely the Adaptive Control Hypothesis (see figure 9). As shown in the displayed architecture of the ACH, three additional modules to control processes are displayed. First, the speech pipeline refers to a conceptual-affective-linguistic-sensory-monitor representations deriving from comprehension activity. Meta-control processes refer to processes in working memory to control the speech-pipeline to achieve the communicative goals.

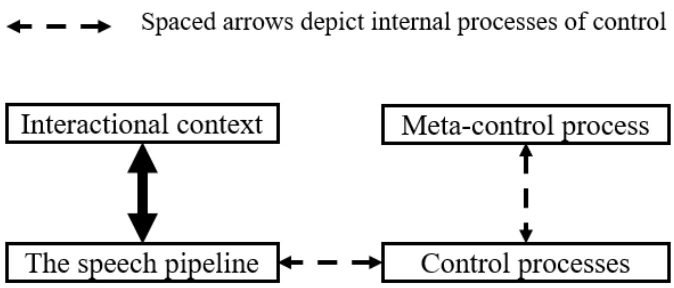

Figure 9: Architecture of the ACH, modelled on Green & Abutalebi, 2013: 517

The critical element, however, is the tripart split patterns of everyday language use defined by the interactional context module, these are as follows. First, there is a single-language context (SLC) which refers to speakers only applying *one* language in each context (i.e., one language at home and another at work). Second, a dual-language context (DLC) refers to contexts in which the speakers use both languages and often with a specific interlocutor. Finally, a dense-code switching (DCS) refers to the same scenario as the SLC; however, in these contexts, both languages are freely mixed within single utterances.

Table 1: Modelled on Green & Abutalebi (2013: 519). + indicates the context increases the demand on that control processes (more so if bolded);= indicates that the context is neutral in its effects.

Demands on language control processes in bilingual speakers as a function of interactional context relative to demands of the processes in monolingual speakers in a monolingual context			
		Interactional contexts	
Control processes	*Single Language* (SLC)	*Dual Language* (DLC)	*Dense Code Switching* (DCS)
Goal maintenance	+	**+**	=
Interference control: conflict monitoring and interference suppression	+	**+**	=
Salient cue detection	=	+	=
Selective response inhibition	=	+	=
Task disengagement	=	+	=
Task engagement	=	+	=
Opportunistic planning	=	=	**+**

Green and Abutalebi (2013, also see Abutalebi & Green; 2016, Green & Abutalebi, 2018) describe how the SLC, DLC, and DCS require different activation of cognitive control processes to achieve effective communication. Table 1 presents the differences between the three language contexts and their interaction in activating the various cognitive processes. As can be seen, SLC only requires activation on goal maintenance and interference control (ignoring distractions) because bilinguals must constrain influence from non-intended information. The DLC contexts place a greater demand on the SLC-activated processes as well as all other cognitive processes with the exclusion of opportunistic planning. The DLC places demands on the opportunistic planning while other processes return to a neutral state. Recently, Abutalebi and Green (2016, see figure 10) found that the contextual language demands require the neural network to adapt depending on the context of language use.

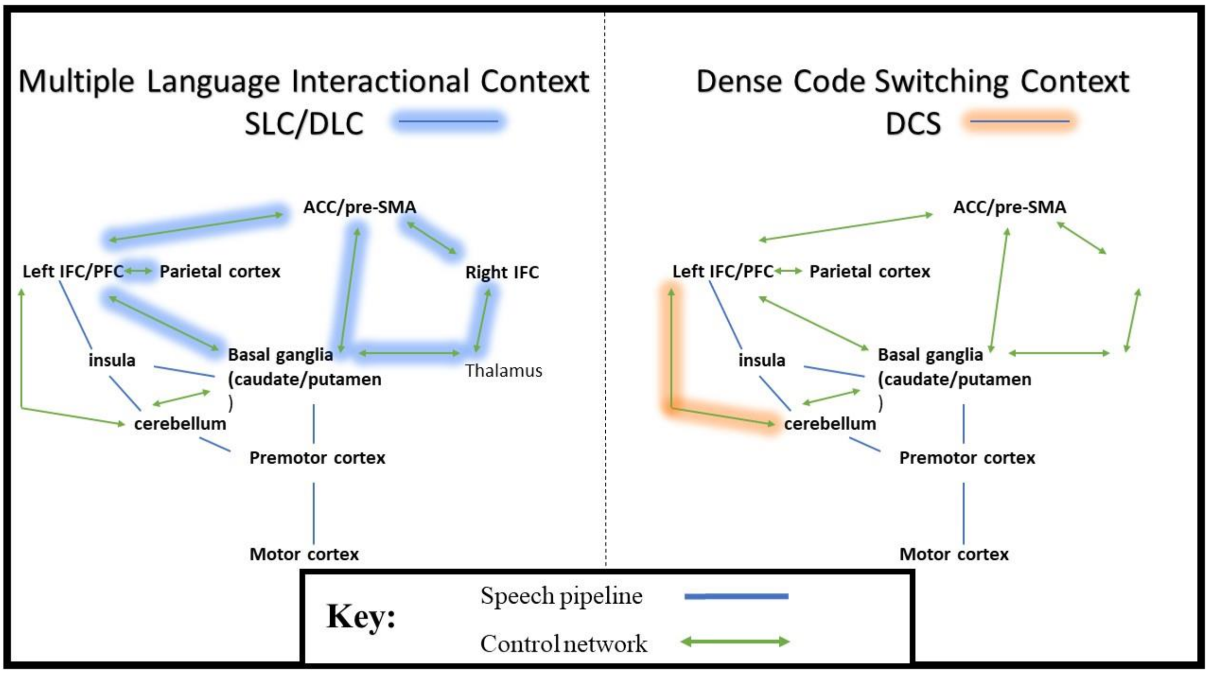

Figure 10: Adapted from Abutalebi & Green (2016: 691): Two situational contexts are outlined through the ACH. The left side indicates SLC and DLC, while the right side indicates activation through DCS.

Nevertheless, the demands on cognitive control and the subsequent effects there is, or lack thereof, on a speaker have been the fuel for another controversy, namely advantages versus disadvantages of bilingualism.

5. Advantages and Disadvantages of bilingualism

The advantage/disadvantages of bilingualism debate concerns empirical findings in which bilinguals appear to be at an advantage compared to monolinguals in, amongst others, executive functions (Bialystok & Craik, 2010; Lehtonen et al., 2018), metalinguistic awareness (Cenoz 2003; Cummins, 1978,), phonological comprehension (Antoniou et al., 2015), and cognitive flexibility (Adi-Japha et al., 2010, but see Antoniou, 2019 for a meta-analysis of cognitive advantages and disadvantages). However, most pressingly is that mounting evidence in our focus field, namely spoken language production, is leaning towards the disadvantages with evidence from TOT retrieval failures (e.g., Gollan & Silverberg, 2001; Gollan & Acenas, 2004; Gollan et al., 2005a;) and picture naming (e.g., Gollan et al., 2005b; Gollan et al., 2007; Ivanova & Costa, 2008) which are described in more detail below, but also in category fluency (e.g., Gollan et al., 2002; Portocarrero et al., 2007), and more recently in production of full sentences (Runnqvist et al., 2013).

Two different hypotheses have been proposed to explain the disadvantages, namely the Weaker Links Hypothesis (Gollan & Acenas, 2004; Gollan et al., 2005a; Gollan & Silverberg, 2001, Gollan et al., 2008) and the Competition Model (Green, 1998; Kroll et al., 2006; Kroll et al., 2008). The weaker links hypothesis explains the disadvantages as follows. Since bilinguals speak two languages, and only one at a time, it follows that the bilinguals use each language less often than a monolingual who always speak one language. As such, lexical representations in both languages in the bilingual’s system have accumulated less practice due to less frequent use of each language. With time the less frequent use of each language, results in *weaker links* between the semantic level and the phonological level and, subsequently, poorer accessibility for lexical representations in both languages. Given that bilinguals use each language less frequently than a monolingual, it follows that bilinguals will be at a disadvantage producing speech in the L2 but also in the L1 when they are compared to monolinguals. One interesting aspect of this hypothesis is that the consequences of bilingualism is assigned the same mechanism that influences accessibility in *all* speakers rather than a specific mechanism deriving from being a bilingual speaker. Weaker links should, thus, occur where frequency of language use is smaller in *all* speakers, including monolinguals.

Gollan and Acenas (2004) conducted a study to elicit TOTs for English words in monolinguals, Spanish-English, and Tagalog-English bilinguals using pictures. The pictures were designed to relate to cognates (e.g., *vampire* in English and *vampiro* in Spanish and non-cognates (e.g., *funnel* in English and the corresponding word *embudo* in Spanish). They found that bilinguals produced TOTs in their dominant English at a higher frequency than English-speaking monolinguals in the non-cognate pictures, and less so in the pictures with cognate translation equivalents, especially when participants were able to translate the cognates after the picture naming task was complete. Gollan and Acenas (2004) mention how prior evidence (Burke et al., 1991; Dijkstra et al., 1999), propose that TOTs are a result of failed phonological retrieval during activation from the semantic to the phonological level. Based on Burke et al.s proposal, Gollan and Acenas (2004) argue that their evidence (mentioned above) provide support for the weaker links hypothesis as cognates have the benefit of stronger links owing to dual activation. As mentioned, the bilinguals experienced less TOTs when they were able to name the cognate translations afterward, which Gollan and Acenas took to suggest that translation equivalents are not inhibited through competition, as is predicted by a competition model. A subsequent study by Gollan, Bonanni and Montoya (2005a) again found that Spanish-English bilinguals produce more TOTs overall than monolinguals. However, in the latter study, Gollan et al. (2005a) found that bilinguals experienced *less* TOTs with proper nouns than the monolinguals. Building on prior research (Burke et al., 1991; Cohen & Faulkner, 1986) that found bilinguals performed better than monolinguals on retrieving proper names, whereas monolinguals better retrieved nouns, verbs, and adjectives, Gollan et al. (2005a) propose that their results to support the weaker links hypothesis given the language duality of proper noun features.

The competition model (Green, 1998; Kroll et al., 2006; Kroll et al., 2008) attributes the bilingual disadvantage to competition between translation equivalent candidates. Simply put, bilinguals face dealing with competition for any semantic concept they know several related words for. The competition may be compared to that proposed in the cascaded processing models proposed for monolinguals (e.g., Peterson & Savoy, 1998). Furthermore, the hypothesis that bilinguals experience cross-language interference may also be compared to many of the bilingual models of spoken production mentioned earlier (e.g., Costa, 2005; Hermans, 2000; Poulisse & Bongaert, 1994). A possible solution for the competition was related to inhibition (e.g., Green,1998a, 1998b). Indeed, the evidence supporting Greens' IC model (Meuter & Allport, 1999; Jackson et al., 2001; Costa & Santesteban, 2004; Schwieter & Sunderman, 2008; Schwieter, 2010, 2013; Linck et al., 2012) also support the competition model.

Critically, the weaker links and competition models are not mutually incongruent, instead, both mechanisms can explain different types of effects found in bilingual language processing. Instead, an interesting aspect may be which of these mechanisms is more effective at explaining the bilingual disadvantage. For example, some of the disadvantages are addressed best through the weaker links hypothesis in the studies mentioned earlier, including picture naming (Gollan et al., 2005) and TOT rates (Gollan & Acenas, 2004). The results cannot with ease by explained by a competition model as bilinguals are, in many regards like monolinguals, at least for the words they only know in one language, which cannot compete and thus invoke an interference. Furthermore, bilinguals were disadvantaged, in contrast to monolinguals, when speaking in their dominant language, which is best explained by the weaker links hypothesis as little evidence has been found to suggest that language production in a dominant language encounters interference from a less dominant language (see Jared & Kroll, 2001; Kroll et al., 2006 for exceptions). While the weaker links suggests that the links between the less dominant L2 and phonological representations are weaker (much like how the L2 words are L1 mediated cause of stronger L1 links with the concept level in the RHM) due to frequency of use, the competition model would suggest that the less dominant language result in weaker competition with a dominant language in production than would be the case if it was less dominant-language production.

Thus far, the litterature suggests that bilingual spoken production is influenced by a number of aspects of a bilingual's profile. As such the following chapter will discuss how one might accurately measure important aspects of bilingual language profiles.

6. Valid Bilingual Profiles

As the current study aims to investigate factors of bilingual profile and relationships between these and proficiency in speaking, it follows that the study requires a corpus of reliable linguistic profiles of the Norwegian-English participants. An early methodological inquiry by Grosjean (1998) distinguishes a few aspects that research on bilinguals may consider. These include *language history and language relationship, language stability, function of languages, language proficiency, language modes*, and *biographical data*.

"Language history and language relationship" refer to the acquisition context in terms of location and by what means languages were acquired. "Language stability" refers to ongoing developments in each language in terms of whether they are still being acquired, being lost, or changed due to changes in the access to the language. "Function of languages" refers to in which context the languages are being used in, for example home, work, school and so forth. "Language proficiency" refers to the skill level in each of the four basic skills, writing, reading, listening, and speaking. "Language mode" refers to how long and often a bilingual is only in monolingual mode (i.e., one active language) or bilingual mode (i.e., both languages active) and how often/how much code-switching takes place. Finally, "biographical data" refers to age, sex, socio-economic and educational status and so forth. These aspects show how not only language proficiency, but rather a range of other aspects and corresponding factors are necessary for research on bilinguals. Given that the goal of the current study is to find how bilingual profiles influence spoken proficiency it is, consequently, important to regard these multifaceted aspects, and subsequentially comprehensive number of factors.

Furthermore, studies suggest that bilinguals are able to consistently assess themselves and their language proficiency as well as report their language history in a consistent manner with behavioral performance (e.g., Chincotta & Underwood, 1998; Flege et al., 1999; Jia, Aaronson & Wu, 2002). Indeed, a range of studies have examined bilinguals, to examine different domains of inquiry, such as grammatical ability (e.g., Jia et al., 2002), degree of foreign accent (e.g., Flege et al., 1999), and computational language (e.g., Vaid & Menon, 2000). However, there is an absence of valid and uniform assessments of bilinguals that may be used to interpret, generalize, and replicate the mentioned studies, and others.

However, a questionnaire constructed by Marian, Blumenfeld and Kaushanskaya (2007) set out to create a reliable and valid questionnaire that may be used for efficient assessment of bilinguals' linguistic profiles. Marian et al. (2007) explain how previous research, although employing differentiated methods to establish bilingual profiles, resulting in contradicting conclusions, does indeed in totality support a theoretical framework that may suggest that language proficiency, experience, *and* language history all are important attributes of bilingual profiles. However, the previous research (mentioned above, e.g., Jia et al., 2002; Flege et al, 1999; Vaid & Menon, 2000 amongst others) had non-uniform methods and questionnaires for bilingual profiling, making a comparison of the findings problematic. As such, the language Experience and Proficiency Questionnaire (LEAP-Q, see Appendix A), was designed to provide a valid assessment of factors that contribute to bilingual status. These factors are critically identified as language competence (including proficiency, dominance, and preference ratings), the age of language acquisition; prior language exposure; and current language use in life.

Marian et al. (2007) conducted two studies to maximize and test the validity and reliability of the LEAP-Q. The first of these two aimed to establish internal validity through factor analysis and multiple regression analyses. They asked 52 bilinguals to answer a preliminary edition of the questionnaire. The results from their first study led the authors to omit some measures such as writing proficiency, current classroom exposure, amongst others. These measures were omitted as they correlated with other measures or did not predict any other measure of proficiency. Study 2 aimed to establish criterion-referenced validity by comparing self-assess scores and standardized proficiency measures and then conducting a correlation and regression analysis. They asked 50 bilinguals to answer a revised LEAP-Q first, after which they were asked to perform a range of standardized proficiency tests. Results from study 2 found a positive correlation between the standardized tests and self-rating on some proficiency factors. For example, high self-rating in the L2 reading proficiency was reflected in behavioral measures. Another finding suggests that objective measures of comprehension and grammatical judgment were good predictors of proficiency self-rating in bilinguals' languages (i.e., participants' rating in terms of proficiency are reflected in grammatical judgement tasks). One significant finding given our current study is that the multiple regression analysis suggests that self-rated language proficiency, to a certain degree, predicted linguistic performance.

Evidently, the LEAP-Q was shown to be reliable and valid and has, in the 15 years since its creation, been used in a multitude of studies (see Marian et al., 2019 for a review). Nevertheless, since the questionnaire helps in creating valid bilingual profiles, as well as predict behavioral performance through self-assessment, it is, indeed, valuable for the current study which does exactly that, studying bilingual factors' influence on behavioral performance, namely language production. Critically, Marian et al. (2019) emphasize that additional objective measures of proficiency in addition to the LEAP-Q best provide a valid bilingual profile. The next chapter will address how one might measure proficiency, specifically through language spoken production.

7. Measuring Proficiency

Before starting any inquiry into measurements of L2 proficiency, it is essential to address the *concept* of proficiency. Leclerq et al. point out that the European educational policy texts (2001) view proficiency, in language education, in terms of adequate language communication abilities instead, a perspective shared by and Norwegian EFL texts (Norwegian Department of Education, 2022). The latter approach to how proficiency is understood is reflected in that studies and models that focus on proficiency as a variable have considered proficiency in terms of *levels* of achievement (e.g., Ginther et al., 2010; Hulstijn, 2011; de Jong et al., 2015; Nadri, Baghaei & Zohoorian, 2019; Duran Karaoz, 2019; Flognfeldt et al., 2020). A definition proposed by Hulstijn (2015, p. 21), "knowledge of language and the ability to access, retrieve and use that knowledge in listening, speaking, reading and writing," reflects the perspective of various levels. As such Hustijn's definition is helpful in the present study as the interest lies in knowledge and ability to various degrees based, and because it adheres to how the literature address proficiency.

Albeit a definition has been proposed, measuring oral proficiency-based on these alone is further exacerbated as studies have examined differentiated possible indicators of proficiency such as *fluency complexity*, and *accuracy* which are often used together to assess overall proficiency (see Norris & Ortega, 2009 for review).

Fluency is used as an indicator of proficiency, specifically oral proficiency because it broadly refers to "a cover term for oral proficiency" (Lennon, 1990 in Duran-Karaoz, 2019). However, it is also used to refer to a person's overall proficiency (see Segalowitz, 2016 for a review). Duran-Karaoz (2019) highlights a narrower sense of fluency which refers to the quantity, rate, and pausing measures both in a temporal and non-temporal aspect employed to assess oral communicative abilities. Additionally, Fluency is also a multifarious concept that entails what Segalowitz (2011; 2016), in a tripart framework, coin as three aspects of L2 fluency, namely utterance fluency, cognitive fluency and perceived fluency. *Utterance fluency*, Segalowitz (2016, p. 5-13) explains, refers to "the fluidity of the observable speech as characterized by measurable temporal features" (e.g., measures of speed, hesitations, and pauses); *cognitive fluency* is "the fluid operation (speed, efficiency) of the cognitive processes responsible for performing L2 speech acts" (i.e., measures of L2 speech acts including all the processes from semantic retrieval to articulation); *perceived fluency* refers to "subjective judgments of L2 speakers' oral fluency" (i.e., rating fluency based on speech sample). Although Segalowitz (2016) explains how despite utterance, cognitive and perceivd fluency are interrelated and equally important, it is nevertheless the case that a myriad of research involving L2 proficiency (Wright, 2013; Kormos & Denes, 2004; De Jong et al., 2015; Huensch & Tracy-Ventura, 2017; Duran-Karaoz, 2019) choose to focus on utterance fluency. One proposed reason is the concreteness and measurable nature of utterance fluency (Duran-Karaoz, 2019) allows for objective measures and subsequently systematic analysis.

The measurable nature of utterance fluency is approached through three categories: *breakdown*, *speed*, and *repair fluency* (Skehan, 2003 Tavakoli & Skehan, 2005). Breakdown fluency refers to the pauses both in terms of the amount, type (filled or silent), and location (where it is in relation to a clause or meaning bearing utterance). Speed fluency is measuring the speed at which speech is being produced, for example, by measuring syllables per second.

Repair fluency is measured by "repair" strategies employed by speakers to modify or correct their speech (i.e., corrections, repetitions, false starts). Furthermore, these measures can be either *composite* or *pure* (e.g., de Jong et al., 2015; Tavakoli & Hunter, 2017). Composite measures combine categories of measurements such as speed and breakdown fluency, which evidently link with measures of perceived fluency (Kormon & Denes, 2004). Pure measures, however, focus on only one category at a time (e.g., speed alone) and can provide insight into speech production processes as it is believed to reveal underlying processes in production (see Huensch & Tracy-Ventura, 2017, p. 277).

In an overview of fluency measures, de Jong (2016) lists typical temporal measurements of fluency that have been employed in the research field (Table 2). However, various researchers have varied in whether they use all, or only some of these measures, depending on what they find to be crucial for measuring proficiency (e.g., Kormos & Denes, 2004; Ginther et al., 2010; De Jong et al., 2015; Duran-Karaoz, 2019; Huensch & Tracey-Ventura, 2017).

Table 2: Global measures of fluency (adopted from De Jong, 2016 p. 212)

Measure	Formula
Speech rate	Number of syllables / total time
Pruned speech rate	(Number of syllables – number of disfluent syllables) / total time
Articulation rate	Number of syllables / speaking time*
Pace	Number of stressed syllables / total time
Mean length of utterance	Total speaking time / number of utterances# or Number of syllables / number of utterances#
Number of silent pauses	Number of silent pauses / total time or speaking time*
Duration of silent pauses	Pausing time / number of silent pauses
Phonation time ratio	Speaking time / total time
Number of filled pauses	Number of filled pauses / total time or speaking time*
Number of repetitions	Number of repetitions / total time or speaking time*
Number of repairs	Number of repairs and restarts / total time or speaking time*

* Speaking time is equal to total time minus silent pausing time.

#Number of utterances is equal to the number of silent pauses *plus* 1.

For example, an early study by Kormos and Denes (2004) analyzes temporal measures of fluency from 16 (8 fluent and 8 non-fluent) L2 speakers. The aim of their paper is to investigate whether the perception of fluency (i.e., *perceived fluency*) from both native and non-native teachers (often the judges of speaker's fluency) could be predicted by speech-sample fluency measures (i.e., utterance fluency measures). Kormos and Denes explain how speakers' fluency is frequently judged although research on what underlies listeners' perception is scarce. As such, they aim to investigate the relationship to develop more reliable criteria for fluency assessment.

Kormos and Denes (2004, pp. 148-152) list a range of fluency measures they will extract. First, they explain that the supposed best predictors of fluency are *speech rate* (same formula as in Table 2) and *mean length of runs* which is the average number of syllables produced in utterances between pauses of 0.25 seconds and above (identical to mean length of utterances in Table 2). Another predictor *Phonation-time ratio* is the percentage of time spent speaking as a percentage proportion of the total time (including pauses). Moreover, they measured *articulation rate*, *number of silent pauses* per minute, *number of filled pauses* per minute, and pace (all share the same formula as in Table 2). Additionally, the *number of disfluencies* per minute was counted, including repetitions and repairs (unlike the formula in Table 2), and space, which is the proportions of stressed words in the total number of words.

Kormos and Denes administered a monologic (monolog speech) narrative task based on three fixed-narrative lines of cartoons to observe the selected measures. There were two motivations for the monologic fixed narrative task. First, an interactive task would be difficult when analyzing the speech sample as both speakers may talk at the same time and because a fixed narrative provides some control of the output. Second, a fixed narrative prevents differentiated cognitive loads (because participants can produce theoretically infinite variations), which may, in turn, interfere with fluency (i.e., the influence of choosing different content is limited).

Results showed that there were significant differences in speech rate, phonation time ratio, mean length of run, and mean length of pause-measures between the non-fluent and fluent participants. All these measures showed higher and more extended production levels for the fluent than the non-fluent group, meaning the more fluent participants produced longer utterances, more syllables, and spoke faster/compact. Moreover, Kormos and Denes found *strong* correlations between teachers’ rating and phonation time ratio (r=0,74), mean length of pauses (r=-0.62) and accuracy (r=0.76). Although they measured pauses, Kormos and Denes underline the uncertainty measuring pausing has as research is equivocal on whether filled/unfilled pause measures relate to fluency rates (2004). It is noteworthy to mention that Kormos and Denes also measures non-temporal variables not listed here for lack of relevance.

Another study conducted by Ginther et al., (2010) also examined temporal fluency measures. Ginther et al. (2010) describes how fluency is considered an important component of speaking proficiency and how temporal measures of fluency are expected to relate to the holistic rating of speech quality strongly. As such they aim to examine the relationship between temporal measures of fluency and holistic scores. In their study Ginther et al. (2010) examined a larger sample of 150 participants. The participants had various language backgrounds, and the most represented language backgrounds were L1 Chinese or Hindi. Additionally, the participants had various levels of English proficiency. The latter was calculated through the Oral English Proficiency Test (OEPT), which Ginther et al. explain is administered to screen hundreds of international teaching assistant subjects. The OEPT consists of 8 differentiated task items which were recorded. These speech samples were rated by trained raters who used a holistic scale ranging from 3 to 6, where the higher numbers refer to increased oral proficiency.

Ginter et al. (2010) categorized their measures in terms of the mentioned categories; quantity, rate, and pausing. In Table 3, the temporal measures are placed in each of these categories. These measurements were analyzed in the OEPT participant responses in their L2 English. While the OEPT consisted of 8 task items, Ginter et al. only analyzed the response to one task item, namely, a news item (NI). The NI task entails participants giving an opinion about campus issues.

Table 3: Categorical listing of temporal measurements adapted from Ginter et al., (2010, p. 387)

Category	Temporal measurements	Formulation
Quantity	Total response time	Speaking time including silent pauses and filled pauses.
	Speech time	Speaking time, excluding silent and filled pauses.
	Speech time ratio	Speech time/total response time.
	Number of Syllables	Total number of syllables in a speech sample.
Rate	Speech rate	Total number of syllables/total response time .
	Articulation rate	Total number of syllables/total speech time.
	Mean syllable per run	Total number of syllables/number of runs.*

Pauses (silent)	Silent pause time	Total time of silent pauses (250ms and above).
	Number of silent pauses	Total number of silent pauses.
	Mean silent pause time	Silent pauses time/number of silent pauses.
	Silent pause total pause ratio	Silent pauses time/total pause time
	Silent pause total response ratio	Silent pauses time/total response time
Pauses (filled)	Filled pause time	Total time of filled pauses (non-lexical sounds).
	Number of filled pauses	Total number of filled pauses.
	Mean filled pause time	Filled pauses time/number of filled pauses.
	Filled pause total pause ratio	Filled pauses time/total pause time.
	Filled pause total response ratio	Filled pauses time/total response time.

* Ginter et al., refer to a Run as utterances between pauses of 250ms and above (Kormos & Denes, 2004 in Ginter et al., 2010)

Results showed that participants that were rated highly in the OEPT (high level of proficiency) related strongly to measures of speech rate, articulation rate, and mean syllables per run. Furthermore, the relation was strongly negative to measures of silent pause time and slightly negative to total response time. When calculating the latter measures (pure measures) in relation to other measures (composite) the speech time ratio scored rather positively (r=0,57) while the silent pause ratio, again, scored negatively. They also found lower proficiency speakers' speech to be mainly composed of 60% filled utterances and 40% pausing, while the more proficient speakers scored, respectively, 80% and 20%.

In other words, with an increased OEPT score, participants have an increased speaking rate, articulation rate, percentage of time spent producing speech, and the mean syllables per run. However, both silent pause and total response time measures show no significant relation with the OEPT scores unless used together with other measures to elicit ratios in which speech time ratio correlated, while silent pause ratio still proved as an inaccurate measure across proficiency levels (OEPT scores).

Ginther et al. (2010) conclude that the most accurate measures of the OEPT, and by extension proficiency, were speech rate, articulation rate, and mean syllables per run. Moreover, highly rated speakers in terms of their OEPT scores (English proficiency) produced more syllables per minute. They take that to mean that a higher proficiency equals quick and continuous speech stream contrasted to lower proficient speakers. As mentioned, there was no significant relation between pause ratios across proficiency levels. Despite the negative relation, and because literature evidently shows interest in such measures leads

Ginter et al. (2010, p.394) to suggest "careful consideration" of measuring pauses as indicators of proficiency.

A third relevant body of research, de Jong et al. (2015), investigate if L2 fluency is a *valid* indicator of proficiency or of speaking style. De Jong et al. describe how a study conducted by Derwing et al. (2009 in De Jong et al., 2015) suggests correlations across a bilingual's languages in terms of measures such as number of pauses per second, speech rate, and pruned syllables per second (total number of syllables produced excluding repetitions, repairs, and restarts/minute). Such correlations indicate that non-fluent speakers may carry other L1 characteristics (speaking style) to L2 speech production. If the latter is true, the validity of using fluency measurements to assess L2 proficiency may be uncertain. As such, De Jong et al. (2015) set out to investigate three questions:

1. To what extent can different measures of L2 fluency (e.g., length of pauses or speed of speech) be predicted from the equivalent measures in L1?
2. Are L2 fluency measures that are corrected for L1 fluency behavior better predictors of overall L2 proficiency than are uncorrected L2 measures?
3. Is the predictive value of (corrected) measures of L2 fluency dependent on typological distance between L1 and L2?

In de Jong et al. (2015) 53 participants with either L1 English or L1 Turkish speakers were examined. Critically, they all shared a common L2, Dutch. The participants' L1 was chosen to avoid possible cross-linguistic difference effects and gain a more distributed result later as English and Dutch are more typologically similar than the agglutinative nature of Turkish.

The researchers collected proficiency data based on utterance fluency and objective proficiency measures. They operationalized the fluency measures in two ways: uncorrected measures (only relating to L2 speech production) and corrected measures (adjusted for L1 behavior). The objective measure (vocabulary knowledge) collected was then related to both (un)corrected measures to find which measures *better* reflected L2 cognitive fluency as an indicator of L2 speaking proficiency. The utterance measures (see Table 4) include speed fluency through mean syllable duration. Breakdown fluency in both mean length of silent/filled pauses and mean length of pauses between analyzed speech units (ASU). ASU refers to utterances that consist of an independent clause, or subclause unit, together with any subordinate clause(s) that may associate with either. Finally, repair fluency was measured in repetitions and corrections per second. All measures were collected for participants both L1 and L2.

Table 4 Measures of fluency adapted from De Jong et al., (2015, p. 9)

Measure	Formula
Mean Syllable Duration (ms)	Speaking time / Number of syllables
Mean Silent Pause Duration (ms)	
Within ASU (ms)	Mean length of pauses within ASU (lower bound of 250ms)
Between ASU (ms)	Mean length of pauses between ASU (lower bound of 250ms)
Number of	
Silent pauses/second	Silent pauses / Speaking time
Filled pauses/second	Filled pauses / Speaking time
Repetitions/second	Repetitions / Speaking time
Corrections/second	Corrections / Speaking time

Note: De Jong et al., describes how each measure of speaking time is excluding pauses to avoid confounding (2015, p. 8).

To measure fluency, De Jong et al. used eight speaking tasks from a previous study by de Jong et al. (2012). The tasks, in each language, were matched for three parameters, complexity, formality, and discourse mode. These were in turn operationalized in terms of the content in each task and instruction given. For example, when presenting a task in participants L1, a mirrored task would be provided in L2. Furthermore, the tasks in English were translated into Turkish. First, participants completed the vocabulary task, followed by the speaking tasks performed in Dutch and finally, between one to four weeks later, the final speaking task was set in their L1 (see Table 5 for speaking tasks).

Table 5 Speaking tasks adapted from Appendix A in De Jong et al., (2015, p.18)

Task	Parameters	Description
1	Simple, informal, descriptive	Participant speaks on the phone to a friend, describing the apartment of friend who have recently moved house
2	Simple, formal, descriptive	Participant, who has witnessed a road accident some time ago, is in a courtroom, describing to the judge what had happened
3	Simple, informal, argumentative	Participant advises his/her sister on how to choose between (or combine) childcare, further education, and paid work.

4	Simple, formal, argumentative	Participant is present at a neighborhood meeting in which an official has just proposed to build a school playground, separated by a road from the school building. Participant gets up to speak, takes the floor, and argues against the planned location of the playground.
5	Complex, informal, description	Participant tells a friend about the development of unemployment among women and men over the last ten years.
6	Complex, informal, argumentative	Participant discusses the pros and cons of three means of transportation (public transportation, bicycle, automobile) on how to solve the problem of traffic congestions.
7	Complex, formal, descriptive	Participant works at the employment office of a hospital and tells a candidate for a nurse position what the main tasks in the vacant position are.
8	Complex, formal, argumentative	Participant, who is the manager of a supermarket, addresses a neighborhood meeting and argues which one of three alternative plans for building a car park he/she prefers.

Results show that both language groups performed equally in their L2 vocabulary task, which was subsequently checked for normality distribution (De Jong et al., 2015, p. 8). Furthermore, both groups were more fluent in their L1 than their L2, with the most significant differences in mean syllable duration and a small, "yet large" (De Jong et al., 2015, p. 15) difference in number of repetitions. In respect to their first research question, variances based on L1 on the L2 speech production accounted for 21% of the syllable duration measure (speed fluency) and 57% of the mean length of pauses between ASU (breakdown fluency). Regarding the second research question, both corrected and uncorrected L2 measures, mostly, related equally to the vocabulary knowledge scores. However, for the syllable duration measure, the corrected measure corresponded more to the vocabulary score. Since both groups measurements of fluency were the same across language groups, De Jong et al. (2015) determined that the results from the previous research questions may be generalized to other language groups.

de Jong et al., (2015) conclude that research in L2 speech production may benefit from corrected measures by taking subjective L1 speaking style into account. However, their results only suggest that *some* measures (syllable duration, mean length of pauses between ASU), when corrected, to an extent, better predict proficiency. Repair measures correlated equally across both (un)corrected measures. Moreover, both (un)corrected measures equally predicted objective proficiency measures, except for the syllable duration measure.

In regard to complexity, despite being widely researched, it seems to be differentially interpreted by different researchers. For example, Bulté andHousen (2012) describe measures taken in forty studies and found that forty different measures were taken. Given the scope of the current study these will not be described or applied in their entirety. Likewise, definitions of the measures vary, however most indicate some sort of language use associated with increased difficult constructions of language (e.g., Foster & Skehan, 1996; Bygate, 1999; Ellis & Barkhuizen, 2005).

As mentioned previously, various measures have been considered, for example some studies have considered T-units (Halleck, 1995; Iwashita, 2006). These T-units refer to as grammatically defined structures as consisting of independent finite clauses plus any finite/nonfinite clauses depending on it. Furthermore, others (e.g., Iwashita, Mcnamara & Elder, 2001; Skehan & Foster, 1997) measured complexity on the basis of C-units, which differ from T-units as these are simple clause, or independent subclausal units together with the subordinate clauses associated with them. Furthermore, more recent research, (e.g., Michel, Kuiken & Vedder, 2007; Tavakoli & Foster, 2008) have measured complexity in terms of analysis of speech units (AS-units). AS-units refer to a single speakers' utterance consisting of an independent clause, or subclausal unit, together with any subordinate clause(s) associated with either (see Foster et al., 2000). In totality, the mentioned studies focus on two distinct complexity measure typologies, which include complexity ratio measures (i.e., mean length of run/utterances) or complexity frequency of frequency counts of particular production units such as morphemes (Gan, 2012).

The studies described above inform the present study in various ways. First, Kormos and Denes, despite the low participant pool, found significant correlations between temporal fluency measures and the perceived fluency of trained raters. As such, measuring the fluency of the present studies' smaller participant pool is to a larger extent legitimized. Moreover, the current study also profits from using the same measures Kormos and Denes found to correlate mostly, namely, phonation time ratio, mean length of pauses as well as accuracy. Critically, one of these measures, namely the mean length of runs is also taken as a measure of syntactic complexity in a range of studies (e.g., Bulté et al., 2012 for meta-analysis). Meaning, not only does it indicate fluency, but also complexity.

An additional gain for the current study is the fixed narrative task employed (see Kormos & Denes, 2004, p. 151). Its controlled narrative and its monologic nature avoid difficulties, which is useful for the current study. A final gain from Kormos and Denes' study is their remark on how research on applying repair measures as proficiency indicators is equivocal, which grants the present study reason to omit that measure. Critically, these measures are often referred to as belonging to *accuracy* measures (Norris & Ortega, 2009). Given the scope of the current study, and how accuracy measures may result in ambiguity, these measures will not be considered.

Ginther et al. (2010) found that accurate fluency measures of proficiency may manifest in speech rate, articulation rate, and mean syllables per run. Additionally, in their large sample bilinguals with a higher proficiency level produced more syllables per minute than lower proficient bilinguals. Finally, acknowledging the suggestion made in Ginther et al. (2010) to carefully consider the use of filled/unfilled measures provides a reason for omitting these measures in the current study. Omitting the nature of pauses as filled or unfilled is also recognized in the literature (see Kormos & Denes, 2004; De Jong et al., 2015; Huensch & Tracey Ventura, 2017, p. 286 for details).

Additionally, de Jong et al. (2015) found that some measures may be better indicators of proficiency when corrected for subjective L1 speaking style. Keeping that in mind, measuring these fluency indicators is nevertheless legitimate when recognizing their limited power. Moreover, corresponding with Kormos and Denes (2004) and Ginther et al., (2010), de Jong et al., (2015) colleagues find repair fluency a redundant and at least uncertain measure of proficiency.

de Jong et al. (2015) found that objective measures in vocabulary knowledge correlated equally across both (un)correction measures (except for syllable duration) dimensions, providing legitimacy to their accuracy, despite the influence of L1 speaking style, a legitimacy this study aims to utilize given Marian et al. (2007) suggested such measures as an addition to the LEAP-Q.

Finally, we saw that complexity can be measured in terms of mean length of utterance, mean length utterance duration (ratio measures), as well as frequency counts, of which morphemes are feasible for the current study. Surprisingly, despite cognates being linked with proficiency (e.g., Gollan & Acenas, 2000, 2004 and others), no studies found or reported in Bulté and Housen (2012) used it as a measure of proficiency. An absence which the current study can address. Having reviewed some approaches to how proficiency may be measured, the following chapter will address the aim of the study.

8. Aim of the present study

As discussed above, a range of studies have examined various factors that may influence proficiency in bilinguals. The current study aims to expand our understanding of the factors influencing L2 spoken production by investigating how aspects of bilingual profile relate to proficiency in speaking L2 English. The study has three components. First, we collected data on aspects of bilingual profile derived from a detailed questionnaire. In particular, Norwegian-English bilinguals were asked to rate themselves on a number of factors including the nature of their instruction and exposure to English. Second, we collected objective language proficiency data using a vocabulary and a sentence comprehension tasks. Finally, we employed a novel methodology to gather measures of spoken proficiency, using a partially controlled free picture description task to elicit natural speech production. From this speech we derived a number of measures including measures of vocabulary (e.g., no. of content words, cognates, non-cognates), measures of utterance complexity (e.g., utterance duration, MLU) and measures of speech fluency (e.g., total response time, speech rate, articulation rate and pausing). These are described in detail below.

The study is designed to address a number of questions about the relationship between bilingual profile and L2 speech production.

1. Which aspects of bilingual profile relate to self-rated L2 English proficiency
2. Which aspects of bilingual profile relate to performance in objective tests L2 English vocabulary and sentence comprehension?
3. Which aspects of measured L2 proficiency relate to aspects of proficiency of L2 English speech production?
4. Which aspects of bilingual profile relate to measures of proficiency of L2 English speech production?

Of particular interest are the effects of formal and informal exposure to English. Here, the study aims to test the hypothesis that different exposure and learning factors may contribute to utterance complexity and utterance fluency. Utterance complexity will be measured in terms of vocabulary (e.g., cognates, non-cognates) and key composite measures (e.g., mean utterance duration, MLU) while utterance fluency is measured in terms of key composite measures (e.g., total response time, speech rate, articulation rate, mean syllables per run, phonation time ratio, and pausing). All described in detail below. Some possible hypotheses may be considered.

One possible hypothesis is that increased contribution of formal learning and exposure to English, for example school will enhance the bilingual's production of more compact fluency, and extended complexity measures. If not, it may be the case that the opposite is true, in that informal learning and exposure, for example through friends and family better supports these measures. Another hypothesis is that language use domain aspects which are closely related to formal instruction without necessarily being so, such as reading, will relate to increased fluency and complexity measures in contrast to informal usage domains such as gaming. Specifically, gaming will hinder lengthier speaking time (total speaking time) while enhancing compact speech (mean syllable per run) given its natural use of fast-paced communication.

9. Method and procedure

The aim of the tasks and experiment in this study were as follows. To establish bilingual linguistic profiles on Norwegian-English bilinguals and to obtain self-ratings on their proficiency through a questionnaire. Second, to obtain objective measures of their proficiency through vocabulary and sentence comprehension tasks. Finally, to obtain measures of proficiency from a partially controlled novel free picture description experiment.

9.1 General Procedure

The study comprised five tasks which were administered in four steps (see Table 6). Step 1 included a participation questionnaire (see Appendix A) with inclusion and exclusion criteria (see *participants*) accompanied by an adapted version of the LEAP-Q (see Appendix B), which was e-mailed to participants. Step 2 was a vocabulary task, followed by step 3, a comprehension task. Steps 2 and 3 were administered at the same time by e-mail. The final step required participants to complete a spoken picture description task at the University of Agder's experimental linguistics lab. All the tasks took place within a few days to a maximum of four weeks.

Table 6 General procedure steps

Task/Experiment	Inquiry results	Duration	step
Adapted LEAP-Q	Self-ratings & Language background	15 minutes	1
Vocabulary Task	Vocabulary depth	7.5 minutes	2
Comprehension Task	Comprehension depth	7.5 minutes	3
Free Picture Description	Spoken language variables	3 minutes	Final

9.2 Participants

Twenty Norwegian-English bilinguals aged 18-35 participated in this study. These participants were recruited in connection with a larger study, the FAB project (FAB, n.d.).All participants took part in all stages of the study. Participants were native Norwegian speakers with English as a second language with little to no additional language proficiency beyond these two languages. All participants had normal or corrected to normal hearing and vision, and no other known language impairments such as dyslexia or stuttering. Participants varied in their level of education from first-year university students to postgraduates. The majority of participants were recruited at the University of Agder, Kristiansand by targeted recruitment and pamphlet handouts. After completing all test sessions, participants received monetary compensation in the form of a bookstore gift card. All procedures were approved by the university ethics committee and the Norwegian Centre for Research Data (NSD nr. 239577).

9.3 Leap-Q Adapted

Design. The adapted LEAP-Q task collected quantitative within-subjects data set on aspects of participants' bilingual profiles. To gather the data, an **a**dapted version, LEAP-Q**A** was administered to participants (see Appendix B)

Materials. For this study, the original LEAP-Q questionnaire was adapted into an online survey LEAP-QA. The adapted version was extended and altered in a number of ways to make it appropriate for the language use of our participant population. These adaptations are as follows. The adapted version asks about the time spent using language in different contexts. Another change is that the adapted version also asks about perceived changes in fluency and which preferences participants have in cognitive tasks such as simple maths, dreaming, expressing anger or affection, and talking to oneself. The LEAP-QA also includes additional questions on language switching, whether it is accidental/intentional and to which extent. The items regarding years of education and cultural identity were omitted in the adapted version, instead finding place in the participation questionnaire.

Procedure. The apparatus used was a personal PC/Laptop as no other device would support the task program. The program used in this task was implemented in Qualtrics. Participants received the LEAP-QA in the same email as the inclusion and exclusion criteria. The email stated that the questionnaire was regarding the language participants speak and how often the different languages are used. After entering the LEAP-QA, the participants were informed that it would take about eight minutes to complete and that if they were unsure about the question answers, they were to provide an estimate.

9.4 The Vocabulary depth Task

Design and Apparatus. The vocabulary task was collected to ascertain participants' vocabulary depth. To gather the data, a task was administered that required participants to choose synonyms or antonyms of ~~criteria-selected~~ English words. The percentage correct was recorded for each participant.

Materials. For the vocabulary task, two word sets with 20 words in each were constructed. One set consisted of words matched with a synonym of the said word (e.g., English *caprice* with correct synonym *whim*) and the other set matching words with their antonyms (e.g., English *concerned* with correct antonym *uncaring*). Additionally, all test words were controlled for being non-cognates of the two languages examined in the study. In both sets, the targets and the correct synonyms/antonyms were matched for the frequency with three additional non-correct words/foils (e.g., *caprice* with correct synonym *whim* was frequency matched with incorrect words *cattle, brute, and lounge*) (see Appendix C, table 18 for antonyms and table 19 for synonyms list).

Procedure. Participants were quires to use a PC/Laptop as other devices (smartphones etc) was not supported in the task program.The program used in this task is based on the platform Pavlovia by PsychoPy. The vocabulary task was handed out in form of a link in an e-mail along with the task instructions (see Appendix D for full text). After entering the task web link, participants were instructed on how to proceed in the task and that each word set is administered separately on the monitor (see Appendix E). The first test block was always the synonyms followed by the second test block of antonyms. After both test blocks are completed, the program shuts off automatically. The series of events on each trial was as follows. First, participants received the web link. Secondly, participants saw the information screens and proceeded with the task. Third, participants were either presented the synonym task and asked to find the similar word or the antonym task and asked to find the opposite word, amongst the word sets. For example, participants were presented the word *caprice* and aske to choose between *whim, cattle, brute* and *lounge*.

9.5 Text Comprehension Task

Design. The comprehension task required participants to read a sentence with differing levels of syntactic complexity and then answer a follow-up question to test their understanding. There were four levels if syntactic complexity and examples of each along with a question are shown in table 7.

Table 7: Syntactic structures in the comprehension task

Syntactic structures	Formulation	Example
1	Independent clause conjoined by coordinate conjunction	The boy is examining he girl *and* he is toasting a stale sandwich
2	Relative clause referring to the clause behaving as an adjective	The boy who is examining the girl *is* toasting a stale sandwich
3	A clause describing the animate transitive verb event using the passive that is then modified by a late relative clause	The girl is being examined by the boy who is toasting a stale sandwich
4	A noun-adjective clause describing the inanimate noun is modified by two embedded relative clauses	The sandwich which the boy who is examining the girl is toasting is stale

Materials. Considering word selection; first, two lists of fifty words, as described in Table 8 was created so that both word lists could create sentences with the same event. An additional set was created where the words were controlled to be non-cognates. Next, all the transitive verbs were controlled to be regular so that they could be derived to make the active form passive. Step 2 paired verbs as described to make fifty different sets of semantic events (e.g., animate "*teased*" and inanimate "*colouring*" → The boy is being *teased* by the girl who is *colouring* a silly picture), followed by steps 3 and 4. The sets were designed so that the inanimate nouns could with plausibility, occur with the matching transitive verb in everyday life (e.g., the man is cooking a wooly hat, although providing a laugh, is *not* likely in everyday life). In total, fifty (50) semantic groups were created- 25 in English and 25 in

Norwegian (24 experimental items plus 1 practice item). Of these, the English ones are of interest in this study.

Table 8: Word selection steps

Initial selection steps	Text Comprehension Word Selection	Example
1	Created a list of 50 (fifty) animate transitive verbs	
2	Created a list of 50 (fifty) inanimate transitive verbs	
3	Separate words into Norwegian and English sets.	
Matching criteria steps		
1	All transitive verbs were regular, adding “ing” affix makes active “ed” affix makes passive	base form *tease* → *-Ing* to make active form *teasing* *-ed* affix to make passive *teased*
2	Animate transitive verbs were paired with inanimate transitive verbs	
3	The lead noun and animate noun in matches were always either boy or girl For semantic sets an inanimate noun paired with a plausible adjective was added to complete inanimate transitive verb event	

These semantic groups laid the foundation for constructing the test sentences (see Appendix F). For each semantic group, 4 different sentences that described events were created (see Table 8 above). Critically, each of these sentences contained *the same* semantic information just expressed using different syntactic structures.

The semantic groups were also the foundation for generating two questions using the if-conjunction, with each question’s correct answer being either the lead or animate noun (e.g., either *boy* or *girl*). Asking for either avoids participatory strategic planning. Question formulation 1 focused on the agent of the animate transitive verb event (e.g., for the structure one example above, question one would be “Who is toasting the stale sandwich?” answer being the boy). Question formulation 2 focused on the patient of the animate transitive verb event (e.g., for the structure one example above, question two would be “Who is being examined?”, answer being *girl*) (see Appendix F). The sentences with corresponding answers were used to create different lists. Each list had one version of a sentence in a different condition so that every participant saw the same sentences but once only and each list had a similar number of sentences in each condition. They were presented in a pseudo random order with no immediate repetition of the same syntactic structure.

Procedure. The sentence comprehension task was administered alongside but not simultaneously with the vocabulary task. The task was issued as a link in an email with some additional instructions. In the e-mail, the participants were informed about what the task entails (see Appendix D for full text). Moreover, participants were told, "*it is important that you press the space key as soon as you have* finished reading the sentence" and explained that data was collected on the speed at which they processed the information (comprehend the sentences). After starting the task an introductory page is followed by a practice block. Having finished the task, the program shut off automatically. The series of events on each trial was as follows. First, participants enter the web link and were welcomed with information regarding the task (see Appendix G). Second, the participants were given four trial runs with each level of syntactic complexity. Third, participants were then informed the trial was over and that the real experiment was to begin. Finally, participants answered questions with differentiated syntactic complexity where they answer either boy or girl.

9.6 Free Picture Description

Design. The free picture description task collected a sound recording of participants describing one of the four pictured scenes in English.

Materials. For the free picture description task, pictures were created based on the widely-used (Goodglass et al., 2001) "Cookie Theft" picture. One of the pictures employed also portrays a cookie theft scenario (see Appendix H, figure 15) but in an updated up-to-time diverse, and non-stereotypical fashion (Heyselaar, Wheeldon, & Segaert, 2021). Moreover, 3 additional pictures were created which share the same characteristics. They were match closely for complexity so that they comprise the same amount of animates (e.g., *bird* or *cat*), verbs (e.g., *flying* or *sitting*), and objects (e.g., *stool* or *speaker*) (see Appendix I for lists)

Procedure. Recording took place at the Experimental Linguistics Lab at the University of Agder. Specifically, participants were seated in a sound-attenuated booth in front of a Lenovo ThinkVision T2454p monitor, Logitech K120 USB Wired keyboard, and for recording speech a Røde NT-USB microphone. The task proceeded as follows. First, participants are seated in and informed that their task is to describe the picture in English. Third, participants were given task information visually on the monitor, as seen in Appendix H, table 19. Next, the participants started the task by pressing the spacebar on the keyboard.

After five seconds the picture to be described appeared on the monitor. After finishing their task, the participants ended the recording by pressing the spacebar again.

9.6.1 Analysis of speech samples

In order to obtain proficiency measures, the speech samples were transcribed orthographically using the speech analysis program PRAAT developed by Paul Boersma and David Weenik (2013). Every speech sample was transcribed by the author, and the following eight (8) variables were measured and marked using Praat as follows. (1) *Sentence* (2) *Syllables* (3) Outside Utterance (4) *In Utterance Pause* (5) *Content words* (6) *Morphemes* (7) *Cognate(s)* (8)*non-Cognate(s).* Each of these measures were understood as follows.

1. All utterances that could bear any meaning were transcribed orthographically. The choice to limit these to utterances were motivated by De Jong's' (2015) ASU.
2. Syllables were counted based on the orthographically transcribed utterances in measure 1.
3. Pauses outside/between utterances were marked beginning at 250 ms and above in accordance with Ginter et al., (2010). Pauses were marked regardless of whether they were filled or unfilled.
4. Pauses within utterances were also marked regardless of whether participants had them filled/unfilled.
5. Content words were counted on the orthographically transcribed utterances in measure 1.
6. Morphemes were counted on the orthographically transcribed utterances in measure 1.
7. Cognates were counted on the orthographically transcribed utterances in measure 1.
8. Non-Cognates were counted on the orthographically transcribed utterances in measure 1.

10. Results

10.1 Participants

18 participants took part in the study, 12 female and 6 males. Participants' ages ranged from 19-30, with an average age of 24.3. Additionally, all were right-handed. Seventeen

participants were born in Norway, while one was born in China and they all reported residing most of their living years in Norway except for a few weeks to a month in an English-speaking environment. An average of 16.2 years of formal education was reported, with most participants' level of education being an undergraduate degree. However, one participant's highest level was upper secondary.

10.2 Leap-Q Results

All participants reported Norwegian as both the first acquired and most dominant language, followed by English. Four participants had additional language experience, which included German, French, and Danish. Critically, these were reported to be less dominant languages for them than Norwegian and English. Rated dominance for Norwegian was confirmed by usage preferences. For language use in other contexts, all participants reported using Norwegian for doing mathematics. 16 reported dreaming in Norwegian except for 2 who dreamed in both languages. Fifteen of the participants preferred Norwegian when expressing anger or affection, with 1 participant reporting equal preference and 2 participants reporting both, but *primarily* Norwegian. In terms of the language used when talking to oneself, 11 participants reported Norwegian while 4 reported English, and 3 reported an equal use of both Norwegian and English. Finally, participants reported they were less fluent in one or both of their languages, where 3 reported being less fluent in English, and 1 reported being less fluent Norwegian.

L2 English exposure

Participants reported an average age of first hearing English of 5.2 years (range 0-7), of speaking with an average of 6.4 years, with 1 participant reporting the age of 12 (range 5-12) and reading with an average of 6.2 years (range 5-12).

Table 9: Exposure in English

*Overall percent exposure in			**Current Exposure interacting with		
	Mean	Range		Mean	Range
General	39,4	5-70	friends / colleagues	3,6	0-8
Speaking	18,4	1-45	Family	0,2	0-1
Reading	48,7	10-90	Reading	5,1	0-10
			School	5,6	0-10
			Self-instruction	1,3	0-10
			Media	7,6	3-10
			Gaming	4,3	0-10

Note: * Is given in total percent of overall exposure in all languages
** Current exposure rates following 0 = Not a contributor, 5 = Moderate contributor and 10 = Most important contributor

Table 9 shows the participants' self-reporting regarding their exposure to a language (see Appendix J, table 21 for raw data). Unsurprisingly, since all participants are native speakers of Norwegian, they all reported being exposed to their L2 English than L1 Norwegian. Interestingly, of the general exposure percent in L2 English, most of the exposure derives from reading rather than speaking. Furthermore, it seems that entertainment/interactive platforms (Media and Gaming) and formal education in School are amongst the most important contributors. Reading also scores well in this regard, while friends/colleagues/family and self-instruction seems to be less of contributors in these speakers' everyday lives.

Factors contributing to learning English

Table 10: Participants rate of contributing factors to their learning of English

	Mean	Range
Friends / colleagues	4,5	0-8
Family	1,4	0-8
Reading	6,9	0-10
School	8,9	5-10
Self-instruction	2,0	0-7
Media	9,0	7-10
Gaming	4,3	0-10

Note: Factors contributing are rated following 0 = Not a contributor, 5 = Moderate contributor and 10 = Most important contributor

What participants' rate as most contributing factor to their learning of English is shown in Table 10 (see Appendix J, table 22 for raw data). Three factors, that is, Reading, School, and Media, are deemed close to the most important contributors. Additionally, it seems that Gaming and Friends/colleagues are viewed as slightly less than moderate contributors. Both Family and Self-instruction lean towards being non-contributing when the participants are viewed as a group, but it is evident that some find both more than moderate contributors.

Participants' Proficiency ratings

Participants' self-reported ratings of language proficiency in English are shown in table 11 (see Appendix J, table 23 for raw data). Interestingly, participants report almost similar proficiency ratings across all listed factors ranging from adequate proficiency to perfection. One factor, Accent, differs in that some participant(s) report almost no proficiency on this measure.

Table 11: Proficiency Self-Ratings

	Mean	Range
Speaking	7,1	5-9
Writing	7,4	5-10
Reading	7,6	5-10
Listening	7,9	5-10
Accent	6,5	2-8
Vocabulary	7,1	4-10

Note: Proficiency rates as 0 = None, 5 = adequate and 10 = perfect

Language switching

Table 12: Self-reported Language Switching

	Mean	**Range**
Accidental Intrusions		
ENG in NOR	2,4	0-7
NOR in ENG	2,1	0-8
Intentional switching		
ENG in NOR	1.6	0-6
NOR in ENG	1,6	0-6
Switching Frequency	1,9	0-3
Switching Proficiency	6,8	3-9

Note: Switching and frequency rates follow 0 = never, 5 = half of the time and 10 = all the time, while proficiency rates follow 0 = None, 5 = adequate and 10 = perfect

Numerical means for accidental language switching intrusions in participants, both L1/L2 as well as intentional switching, are shown in table 12 (see Appendix J, table 24 for raw data). Additionally, frequency of the switching is listed, and participants' self-reported proficiency in switching. Interestingly, the accidental intrusions both seem to be very rare, while intentional switching is even more scarce given the almost infrequent number of switches. Nevertheless, participants report they are beyond adequate but have not perfected switching.

Correlation matrix of the exposure, learning, proficiency and switching variables

A correlation matrix was calculated for all the LEAP-Q variables described in tables 9-12. The results are shown in figure 11. Due to the large number of correlations involved, only results with a significant level of p>.001 are shown.

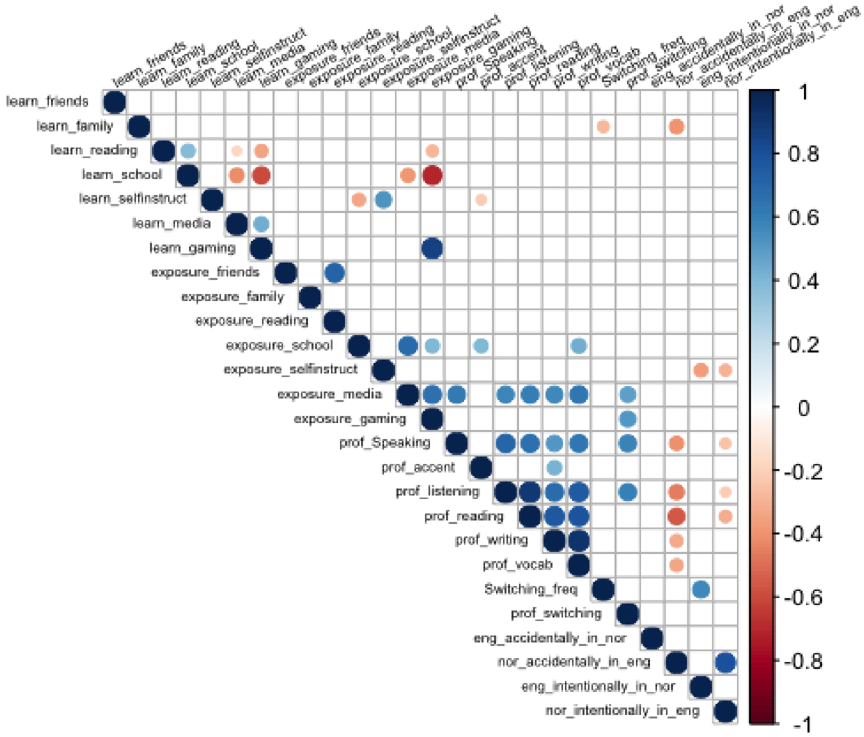

Figure 11: Correlations significant above p>.001 are shown for the learning, exposure, and proficiency variables. Positive correlations are shown in blue and negative correlations in red.

As can be seen, learning English through reading and school correlates negatively with learning through media or gaming. Additionally, learning through reading relates positively to learning in school. Furthermore, learning by gaming hinders, or at least does not coincide with, learning through school, at least in these participant pools. At the same time, we can see a slight positive correlation between learning by reading and through school. The same can be seen regarding learning through media and gaming. Thus, although learning through reading and school does not pattern with learning by media/gaming, these variables complement each other in learning English. The same results can be seen in terms of learning through reading/school and the exposure to media/gaming negatively correlating. Moreover, when exposed to English in school, participants are also often exposed through media and gaming despite the negative learning relationship. Finally, exposure in school also relates to vocabulary proficiency.

Furthermore, it seems there is a positive correlation to suggest that exposure to English reading occurs when there is an exposure to English through friends/colleagues. Additionally, it seems that when learning English through family, there is no language switching, and fewer accidental intrusions from the L1 Norwegian. Moreover, it seems that the accidental intrusions from L1 Norwegian seem to occur regardless of proficiency levels (except for accent proficiency). Regarding switching, there is a negative correlation between intentional switching and exposure to self-instruction meaning that intentional switching proceeds despite increased appearance of self-instruction in everyday life. Additionally switching frequency corelated positively with using L2 English intentionally when speaking L1 Norwegian.

The proficiency measures, excluding accent proficiency, are highly correlated. Thus, it motivated the calculation of a mean proficiency score across these measures. Overall proficiency ranged from 7,1 to 7,9, with a mean of 7,42.

10.3 Objective Language Proficiency Results

For each participant, their correct percentage score for the vocabulary test was calculated. These ranged from 25 to 79.6, with a mean percentage correct of 48.76 for the whole group. Mean percent correct was also calculated for each participant for the text comprehension task (mean=93.3, range 79.2-100) as well as their mean reading time for correct trials (mean=5315, SD= 1680) (see Appendix K).

Correlations between objective measures, self-rate scores and mean proficiency

Kendall rank correlations were run between these objective language proficiency measures and both relevant self-rating scores and the mean proficiency score. None of these correlations were significant (see Appendix L). The strongest observed relationship was between mean reading duration in the text comprehension task and rated proficiency in reading ($z = -1.4944$, p-value = 0.13). As can be seen in Figure 12, higher self-rated reading proficiency was associated with faster reading times in the text comprehension task.

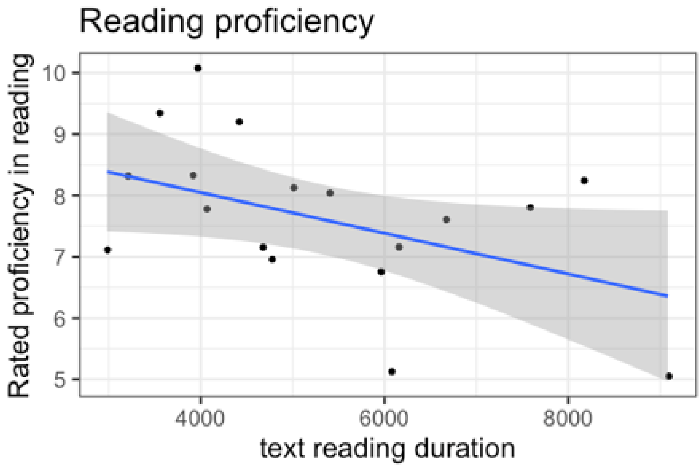

Figure 12: Correlation between mean reading duration in text comprehension and participants rated proficiency.

Linear regression for objective measures and learning and exposure variables

Linear regression models were run for each objective proficiency measure against the learning and exposure variable from the questionnaire. All variables were centered. A summary of the significant relationships is given in Table 13 (see Appendix M for full model output). As can be seen, percent correct in the vocabulary task is positively related to the rated degree of learning through reading but negatively related to the degree of learning through school and media. The percent correct in the text comprehension task is positively related to the rated degree of learning through friends. This measure also showed two borderline negative relationships with a rated degree of learning through self-instruction and gaming. The comprehension means of reading duration is positively related to learning through friends but negatively related to learning by gaming as well as exposure to English by media.

Table 13: Significant variable relationship between objective measures and learning and exposure variables

		Learning		Exposure	
Objective measures	Learning/Exposure	t value	p value	t value	p value
Vocabulary correct percent	Reading	2.307	0.044	n/s	n/s
	School	-2.240	0.049	n/s	n/s
	Media	-2.230	0.050	n/s	n/s
Text comprehension correct percent	Friends	2.208	0.052	n/s	n/s
	Self-instruction	-2.084	0.063	n/s	n/s
	Gaming	-2.897	0.016	n/s	n/s
Text comprehension mean reading duration	Friends	2.312	0.044	n/s	n/s
	Gaming	-2.946	0.014	n/s	n/s

Media	n/s	n/s	-2.559	0.028

Note: area marked n/s indicates no significant relationship

10.4 Free picture description results

The means data for the measures taken from the free picture description task recordings are summarized in Table 14 (see Appendix N for raw data). Based on these data some variables of interest were calculated for each participant. These calculations include measures of fluency such as total response time, speech rate, articulation rate, mean number of syllables, phonation time ratio and pausing per run in addition to measures of speech complexity such as mean utterance duration and MLU as shown in Table 15.

Table 14: Free picture description measures summary

Variable	Measure	Mean	SD
Utterance	Occurrence(s)	13,78	4,76
	Total length	67931,17	19530,52
	Length mean	5138,09	1224,51
Pause outside utterance	Occurrence(s)	9,56	3,42
	Pause length	13696,38	6851,89
	Length mean	1406,81	390,51
Pause within utterance	Occurrence(s)	12,44	6,15
	Pause length	8675,27	4163,67
	Length mean	734,48	169,82
Syllables	Total	194,72	70,39
	Mean	14,25	2,76
Content words	Total	62,78	18,61
	Mean	4,66	0,82
Morphemes	Total	181,44	69,36
	Mean	13,15	2,47
Cognates	Total	30,50	10,05
	Mean	2,26	0,65
Non-Cognates	Total	32,11	17,67
	Mean	2,26	1,023

Note: All numbers rounded up for two decimals.

Table 15: Formulation of key measures by speech data

Category	Temporal measurements	Formulation
Quantity	Total response time	Total utterance time including pauses (within and between)
	Mean Length of Utterance	Total number of morphemes/number of utterances
	Mean Utterance Duration	Total utterance duration/Number of utterances
Rate	Speech rate	Total number of syllables/total response time .
	Articulation rate	Total number of syllables/total utterance time
	Mean syllable per run	Total number of syllables/number of utterances.
	Phonation time ratio	Total utterance time/total response time

Pauses	Total pause within utterance	Total pause duration within utterance/number of utterances
	Total pause between utterance	Total pause duration between utterance/number of utterances

Results of the key measures of utterance fluency and utterance complexity are shown in Table 16 (see Appendix O for raw data).

Table 16: Key measures results summary

	Mean	SD
Total response time	90302	23214
Mean Length of Utterance	13,15	2,54
Mean Utterance Duration	6905	1642
Speech rate	0,0021	0,00042
Articulation rate	0,0028	0,00047
Mean syllable per run	14,25	2,83
Phonation time	0,75	0,080
Total pause within utterance	734	165
Total pause between utterance	1406	401

Correlation matrix across the speech variables

A correlation matrix was calculated for all the speech variables summarized in table 14. Figure 13 shows a correlation with a significance level of p>.05. As can be seen, a correlation matrix was calculated for all the speech variables summarized in table 16. Figure 13 shows a correlation with a significance level of p>.05. As can be seen, utterance mean correlates positively with mean syllables, morphemes, and cognates. Moreover, the mean syllable count relates to both the content word and morphemes mean. The content word mean and morpheme mean also positively correlate.

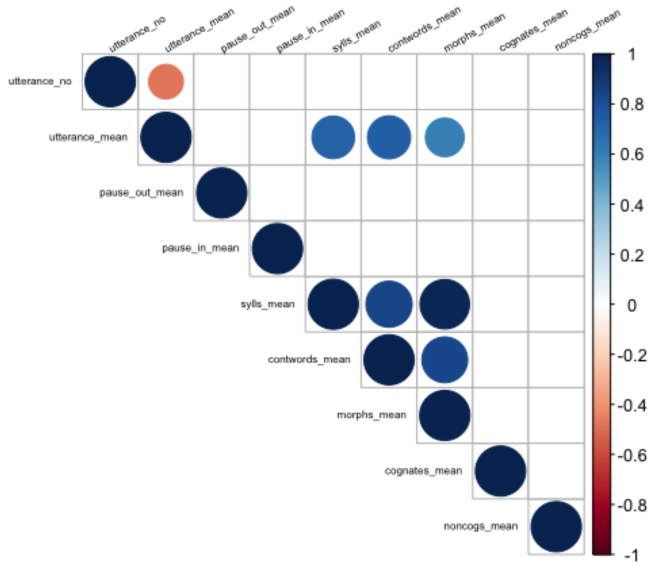

Figure 13: Correlation significance level p>.05 are shown for the speech variables, Positive correlations are shown in blue and negative correlations in red.

Correlation matrix of the speech variables, rated proficiency, and objective proficiency

A correlation matrix was calculated for all the speech variables summarized in table 16, the mean of rated proficiency and objective proficiency variables. Figure 14 shows a correlation with a significance level of p>.05.

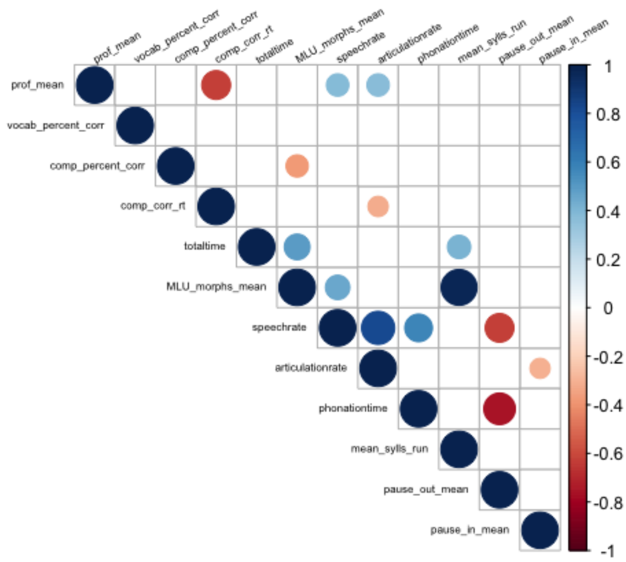

Figure 14: Correlation significance level p>.05 are shown for the speech variables, rated proficiency, and objective proficiency. Positive correlations are shown in blue and negative correlations in red.

As can be seen, the rated proficiency correlates negatively with the comprehension reading time. However, it correlates positively to both speech and articulation rate (both rates). The comprehension correct negatively correlates to the mean length of utterance.

Correct reading times do the same regarding articulation rate. Unsurprisingly, the mean length of utterance correlated positively with the mean number of syllables.

Linear regression for key speech data measures and learning and exposure variables

Linear regression models were run for each key speech data measure against the learning and exposure variable from the questionnaire. All variables were centered. A summary of the significant relationships is given in Table 17 (see Appendix P for full model outputs).

As can be seen, the total response time in speech is predicted by learning English through gaming. The mean length of utterance is positively related to learning by reading and exposure to reading. Furthermore, the mean utterance duration shows a negative relationship with self-instruction exposure. Regarding speech rate, it relates only to exposure in both reading and self-instruction. Likewise, articulation rate relates positively to self-instruction. The most significant relations derive from mean syllables per run, which is positively predicted by exposure through family and reading, but negatively by friends. Like other measures, phonation time relates to reading exposure positively. Meanwhile, learning English through school relates negatively to the mean total pause within utterances, while reading exposure relates negatively to the number of pauses outside utterances.

Table 17: Significant relationships between speech data key measures and learning and exposure variables

		Learning		Exposure	
Key speech data measure	Learning/Exposure variables	t value	p value	t value	p value
Total response time	Gaming	2.067	0.0656	n/s	n/s
Mean Length of Utterance	Reading	2.053	0.0672 .	2.260	0.0473
Mean utterance duration	Self-instruction			-1.997	0.0737
Speech rate	Reading	n/s	n/s	2.764	0.0200
	Self-instruction	n/s	n/s	2.226	0.0502
Articulation rate	Self-instruction	n/s	n/s	2.606	0.0262
Mean syllable per run	Reading	2.028	0.070	2.130	0.0590
	Friends	n/s	n/s	-2.162	0.0559
	Family	n/s	n/s	1.970	0.0771
Phonation time ratio	Reading	n/s	n/s	1.907	0.0856
Mean Total pause within utterance	School	-2.310	0.0435	n/s	n/s

Mean Total Pause outside utterance	Reading	n/s	n/s	-1.888	0.0884

11. Discussion

The aim of the current study was to expand our understanding of the factors that influence L2 spoken production by investigating how aspects of bilingual profile, and in particular second language learning and exposure, relate to spoken proficiency in L2 English. The study was approached through various means. First, Norwegian-English bilinguals were asked to complete a detailed questionnaire to collect data on aspects of their language profile. Specifically, they were asked to rate themselves on a number of factors and report the nature of their exposure and instruction in English. Second, the bilinguals were asked to complete a vocabulary and a sentence comprehension task to collect objective language proficiency data. Finally, a novel experiment using a partially controlled picture description task was employed to elicit speech production to derive a number of measures (e.g., length or no of utterance, syllables, both within and outside utterance pauses, content words, morphemes, cognates and, non-cognates). The latter measures were also used to calculate key composite measures of spoken proficiency (e.g., utterance fluency and utterance complexity). Moreover, the study was designed to address a number of questions about the relationship between bilingual profile and L2 speech production. This chapter will address the results in the current study, to begin with, in relation to these questions.

The first question concerned which aspects of bilingual profile relate to aspects of L2 English proficiency self-ratings. The correlation matrix across these aspects revealed no significant results in either learning or exposure aspects. While it is surprising at first glance that none of these aspects of the bilingual profile related to self-ratings of L2 English proficiency when considering the small participant pool, it is more sensible. Because the current study only examined a small group of bilinguals, and that this group had highly correlated proficiency would suggest the group is not varied enough. This provides ample possibility for future research to revisit these aspects of bilingual profile in a more varied and larger group of participants to relate aspects of their bilingual profile with the proficiency self-ratings.

Notably, however, there were some relationships between proficiency self-ratings and language use, specifically in switching. Results showed that all proficiency measures, except for accent, related negatively to Norwegian being accidentally switched in English language production. Moreover, some of the proficiency self-ratings were also negatively related, albeit to an even more minor degree, to intentional switching from Norwegian in English production. Since the bilinguals rated themselves beyond adequate in all proficiency measures and that switching frequency was low, the negative relationship is surprising. Following the suggestions set forth by the weaker links theory (e.g., Gollan & Acenas, 2004; Gollan et al., 2005a; Gollan & Silverberg, 2001, Gollan et al., 2008) and the RHM model (Kroll & Stewart, 1994) one would expect that higher proficiency means stronger links and therefore a much weaker dual activation of words which subsequently would result in fewer accidental intrusions. The latter would lead to a positive relationship between the high proficiency ratings and the minor frequency of accidental switching, which is not the case. The unexpected result and that there were no significant relationships across the remaining aspects of bilingual profile and proficiency self-rating is also an item to revisit in future research.

Regarding which aspects of bilingual profile relate to performance in the objective L2 English tasks, the results showed relationships between an aspect of the bilinguals' self-rated proficiency and a number of aspects of bilinguals' L2 learning and exposure. The most robust observed relationship with reference to self-rated proficiency is observed between mean reading duration in the text comprehension task and rated proficiency in reading. Results showed that higher self-ratings were associated with faster reading times. Additionally, the participants' self-ratings in reading proficiency were beyond adequate, with some participants rated their reading proficiency as perfect (mean=7.6, range 5-10). Together, this means that participants that have a higher self-rated reading proficiency participants are good predictors of their reading abilities. This result echoes that of Marian et al. (2007), in which they found that self-rates in L2 reading proficiency reflect objective measures of comprehension. However, this only applies to the bilinguals with a higher self-rating in reading and not necessarily in general to all bilinguals or in other measures. With the mentioned smaller participant pool and that the participants had a limited variation in their proficiency self-rates with only using the higher end of the scale, it is not viable to question whether self-ratings in reading and in the other measures of proficiency relate to all bilinguals', including those who rate themselves on the lower end of the scale, in objective comprehension abilities.

Notably, the mean proficiency across all the self-rate proficiency measures negatively correlated with the comprehension correct (see figure 17), suggesting that the bilinguals in this study rated themselves higher across all self-rating variables compared to how they performed in sentence comprehension. As such, and despite the results in Marian et al. (2007)

echoing our results, it will nevertheless be an interesting relation to revisit in future research with a more extensive and more varied participant pool to see whether L2 reading rates relate to comprehension beyond high-self ratings and whether the overall proficiency relation changes.

Results further showed that aspects of L2 learning and exposure are related to objective proficiency. Interestingly, most of the significant relations are found in terms of L2 learning. First, participants learning L2 English through reading related to the objective vocabulary task. This suggests that participants that learn L2 English by reading have an enhanced vocabulary depth. Surprisingly, however, the learning by reading negatively relates to school, meaning that the enhanced vocabulary depth is not attained in formal instruction but rather in reading in other domains. Unsurprisingly, it also relates negatively to learning through media. While the objective vocabulary does not significantly relate to other learning aspects, there was indeed, a smaller but present positive relationship in the correlation between learning through reading and learning through school, as seen in figure 11. As such, the results seem to be conflicting in that learning through reading and school related, while the learning English through school negatively relates to objective vocabulary performance. This could be understood as that while reading does occur, often, in formal instruction, it is not in this reading domain that the bilinguals acquire enhanced vocabulary depth. Moreover, it is not possible to determine more closely where the reading in L2 English occurs, as there are no relations across either bilingual profile aspects and between these aspects and objective vocabulary proficiency. Nevertheless, bilinguals’ learning L2 English through reading does seem to eventuate obtaining vocabulary depth.

There were also significant relations between aspects of bilingual profile and text comprehension. The comprehension correct answers relate positively to learning English through friends. This would suggest that learning with friends enhances comprehension proficiency. Interestingly, text comprehension is negatively related to self-instruction and gaming. Taken with the positive relationship between learning through friends, this would suggest that friends do not contribute to each other’s English when gaming. Indeed, as we can see in figure 11, exposure to English with friends relates to exposure to reading, which may suggest that friends carry out reading activities together. Moreover, reading duration in the comprehension task also positively relates to learning through friends. Unsurprisingly, reading duration is negatively related to learning by gaming, while exposure to English through media negatively affects comprehension abilities. The latter is unsurprising given how gaming and media share many components. With the high correct percent in comprehension, it seems to be the case that bilinguals' learning with friends contributes positively to comprehension depth and is often associated with being exposed to reading. Meanwhile, learning through gaming and self-instruction and exposure to media negatively affect enhanced comprehension ability.

Considering which aspects of measured L2 proficiency relate to aspects of proficiency of L2 English speech production, one significant result is a positive relationship between mean self-rating proficiency and both speech rate and articulation rate. This would suggest that how quick the bilinguals speak is to an extent predicted by higher proficiency, at least in the case of self-rated proficiency. The result does match Marian et al. (2007) finding that L2 reading proficiency relates to behavioral measures. While interesting on its own, the result is further fascinating since Ginther et al. (2010) concluded that bilinguals with higher oral language production proficiency indeed speak more quickly and more accurately. If bilinguals' higher self-rated proficiency is a legitimate, accurate measure of how quickly they speak, and higher proficient bilinguals indeed speak quicker, one might expect that these bilinguals indeed produce faster speech than participants with lower self-rated proficiency. However, since the mean proficiency is based on almost no variation in a set of self-ratings that were all beyond adequate, it is not possible to determine whether there is a difference between low and high proficient bilinguals and how they perform in their L2 speech production.

However, while self-rating proficiency related positively to speech and articulation rate, comprehension reading time and comprehension correct negatively related to articulation rate and MLU, respectively. This suggests that the objective comprehension measures contradict the proficiency self-ratings for the same behavioral measures. This contradiction is surprising considering Marian et al. (2007) findings that suggested that objective measures of proficiency were instead good predictors of self-ratings. Based on Marian et al. findings, one expectation would be that since objective measures and self-rating measures predict each other, they should also predict behavioral measures equally. While the results contradict such an expectation, it is nevertheless explainable since the bilinguals' objective measures and self-ratings in the current study only related to comprehension reading duration and self-rated reading proficiency.

Regarding which aspects of bilingual profile relate to measures of proficiency of L2 speech production, results showed a number of relationships. The bilingual profile aspects with a significant relationship with proficiency in L2 speech production include learning L2 English by gaming, reading, and school , and to a larger extent, exposure through reading, self-instruction, friends, and family to L2 English. First, gaming, and total response time related positively, despite the varied reports of gaming as a contributor to learning English (mean=4.3, range 0-10). Following that, gaming was rated across the participants, as just below moderate contributor to learning English, it follows that learning by gaming does not increase total speaking time, perhaps owing to gaming usually employing fast and brief communication as hypothesized.

Also, in congruent with a suggested hypothesis, learning through reading related positively to both MLU and mean syllables per run. Moreover, exposure to reading also related to MLU, mean syllables per run, measures of speech rate, and phonation time ratio. Since our participants rated themselves beyond adequate in overall proficiency, and specifically in reading proficiency, the relationship between reading and measures of both complexity and fluency is consistent with the findings in Ginther et al. (2010) and Kormos and Denes' (2004) studies. Ginter et al. found more proficient bilinguals to relate strongly with speech rate, and mean syllables per run, while Kormos and Denes found that more fluent bilinguals produced longer utterances (i.e., MLU), more syllables, and spoke faster/more compact. This would suggest that not only does learning through and exposure to L2 English by reading relate to aspects of spoken proficiency, but that more proficient bilinguals in the current study produce these measures of spoken proficiency to extended production levels in congruent with other L2 learners found in Kormos and Denes and Ginther et al. Furthermore, exposure to reading in L2 English negatively relating to total pause time outside of utterance, which is expected given the speech is quicker and more compact in this group.

Another aspect of the bilingual profile, exposure to self-instruction, negatively relates to complexity measures (mean utterance duration) but positively to fluency measures (speech rate, articulation rate). The participants deemed self-instruction as almost not a contributor to L2 English exposure; taken together with the positive relation between exposure through self-instruction and fluency measures, it would suggest that eluding self-instruction is a legitimate approach to speaking faster. While self-instruction specifically was not mentioned in the hypotheses, it can be classified as informal exposure, thus coinciding with other informal learning and exposure in that it is to a disadvantage for producing enhanced fluency. The opposite is true for producing longer utterances, in that low exposure to self-instruction negatively relates to mean utterance duration.

Furthermore, exposure to L2 English through friends is negatively related to mean syllable per run. Considering that exposure to friends was rated as less than a moderate contributor, it would suggest that the lack of exposure through friends negatively impacts compact speech. As such, to produce more compact speech, which is associated with higher proficiency (Kormos & Denes, 2004), exposure to English through friends is to the bilinguals' benefit. The opposite is true in the case of exposure through family, which related positively to the same measure of mean syllable per run. Exposure through family was rated as almost not a contributor, which suggests that the lack of exposure in this domain is a positive contributor to more compact speech. While both can be considered informal domains, they contradict in that exposure through friends enhances fluency while family does not. A final significant aspect of the bilingual profile is how learning through school negatively relates to mean total pause within utterance. School was deemed as a very important contributor to learning L2 English; therefore, this suggests that learning in school does not help reducing pauses within utterances. The latter can be explained as participants may take their time to think through what is being said while the utterance is being produced with formal instruction. Since friends contribute to fluency, while family and school does not, our hypothesis that formal instruction but not informal contribute to fluency and complexity is wrong. Indeed, some aspects, namely exposure through friends is to bilinguals' advance in acquiring more compact fluency.

To summarize, there were, except for the negative relationship between high proficiency and low frequency of accidental L1 Norwegian to L2 English, no other bilingual aspects that related to L2 proficiency self-ratings. This is suggested to be due to a lack of a varied, and larger participant pool. Next, reading proficiency self-ratings related to objective comprehension proficiency, as found in Marian et al. (2007), however, it is not possible based on our results to determine whether this applies to other proficiency levels. Nevertheless, the mean of the high self-rating proficiency is not an accurate measure of comprehension ability. Furthermore, bilinguals acquire more vocabulary depth in learning by reading, and while reading does occur in relation to school, it is not necessarily in school that the enhanced ability is acquired. Likewise, comprehension ability increases when the bilinguals learn English through friends, not in gaming, but rather when reading. Our bilinguals' participants high proficiency self-ratings also predicted faster speaking, which following Marian et al. (2007) and Ginther et al. (2010), would by extension, mean that these bilinguals, given their high proficiency outperform bilinguals with low proficiency, an expectation not feasible to investigate further with the results in this study given there was barely variation in proficiency self-ratings. Moreover, objective measures, specifically comprehension reading duration and correct percent, contradicted self-ratings across the same behavioral measures, an unexpected outcome given Marian et al. (2007) finding suggested objective measures were good predictors of self-ratings, a suggestion not applicable to the current study since comprehension reading duration only related to self-rated reading proficiency. Regarding bilingual profile aspects and measures of L2 speech production, results showed that bilinguals that learn and are exposed to English in more formal usage domains, specifically reading,, perform in both fluency and complexity like highly proficient bilinguals in other studies (e.g., Kormos & Denes, 2004; Ginther et al. 2010). In contrast, informal learning and exposure, for example, in gaming, only enhanced compact speech, probably due to the compact and fast-paced nature of the communication employed. Unlike gaming and reading, learning and exposure in formal instruction (i.e., school) did not enhance fluency, while informal exposure through friends did. However, some informal exposure also hindered more proficient fluency (i.e., exposure through family).

Beyond the implications of bilingual profile on speech production, these results have pedagogical implications that educators should be alerted to. For example, formal instruction through school does not necessarily predict enhanced fluency. Moreover, it does seem that reading is an essential item in acquiring more compact and faster speech, both indicators of highly proficient speakers. Moreover, the reading does not need to occur in a formal situation but rather with friends.

However, a critical limitation of the current study is the small participant pool, which also reports minimal variation in proficiency self-ratings. For example, variation probably hindered determining what aspects of proficiency relate to aspects of bilingual profile. A larger participant pool would also be able to address the differences between low-proficient bilinguals and highly-proficient bilinguals. Future research may also gain from investigating the bilinguals' L1 as well as their L2 following de Jong (2015) to determine possible relationships across language performance. Moreover, while the current study used a set amount of complexity measures, it would be interesting to measure a broader range of complexity measures reported in Bulté and Housen (2012). Finally, accuracy measures were omitted given the scope of the current study but would nonetheless provide an interesting additional proficiency measure to investigate.

12. Conclusion

This study aimed to investigate how different aspects of L2 English learning and exposure relate to different aspects of proficiency in speaking. The results indicate, first, that L2 English learning by reading increases speed and compactness aspects of proficiency in speaking. Second, learning by gaming influences fluency in that it induces shorter speaking time, but also increases compactness in fluency. Third, L2 English learning and exposure through formal instruction, particularly school, does not enhance fluency in spoken production, while informal exposure to English through friends does. However, informal exposure in another domain, namely family, hindered more proficient fluency. Furthermore, the current study investigated how other aspects of bilingual profile relate to each other and to speaking L2 English. Results here show that high proficiency self-ratings relate to faster speaking, in congruence with Marian et al. (2007).

By employing a novel, and only partially controlled production methodology, the study expanded the current understanding of how bilingual profile relates to L2 English speech and contributed against the neglect spoken production research has undergone. Moreover, we found effects of exposure and proficiency beyond AoA, which echoes the emphasis of these measures as important contributors to bilingual language processing (e.g., Abutalebi, 2005; Klein et al., 1995). Moving forward, researchers may employ less restricted narrative tasks knowing they will collect at least some significant spoken proficiency measures. Additionally, pedagogical practitioners should consider how learning and exposure to L2 English beyond in school, may be exploited to help bilingual pupils enhance their proficiency in speaking. There are, however, limitations to the current study. The most significant of these is the size and variation of the participant pool, an interesting issue to address in future research. Moreover, future research may be less time-constrained and find it feasible to investigate both the bilinguals' L1 and the L2 as in de Jong (2015).

In summary, the current study found that aspects of learning and exposure, such as learning by reading, gaming and exposure through friends and family did relate to varying degree to aspects of spoken proficiency. It would be compelling to revisit the methodologies employed with a larger participant pool and perhaps uncover more relationships between L2 learning and exposure and aspects of spoken proficiency.

13. References

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. Bilingualism: Language and Cognition, *19*(4), 689–698. https://doi.org/10.1017/S1366728916000225

Abutalebi, J., Cappa, S.F., & Perani, D. (2005). What Can Functional Neuroimaging Tell Us About The Bilingual Brain? In J.K., Kroll & de Groot, A.M.B. (Eds.). *Handbook of bilingualism: Psycholinguistic approaches*, 497-515.

Adi-Japha E, Berberich-Artzi J, Libnawi A. (2010). Cognitive flexibility in drawings of bilingual children. *Child Devoplemt*. *81*, 1356–66

Antoniou, M., Liang, E., Ettlinger, M & Wong, P. (2015). The bilingual advantage in phonetic learning. Bilingualism. 18. 683-695. 10.1017/S1366728914000777.

Antoniou, Mark. (2019). The Advantages of Bilingualism Debate. *Annual Review of Linguistic, 5*, doi:10.1146/annurev-linguistics-011718-011820.

Baker, C. (2011*). Foundations of Bilingual Education and Bilingualism*. *79, Multilingual Matters*.

Bialystok E. (2011). Reshaping the mind: the benefits of bilingualism. *Canadian journal of experimental psychology = Revue canadienne de psychologie experimentale*, *65*(4), 229–235. https://doi.org/10.1037/a0025406

Bialystok, E., & Craik, F. I. M. (2010). Cognitive and Linguistic Processing in the Bilingual Mind. *Current Directions in Psychological Science*, *19*(1), 19–23. doi: 10.1177/0963721409358571

Birdsong, D. (1999). Second language acquisition and the Critical Period Hypothesis**.** Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9781410601667

Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in Stimulus-Stimulus inhibition. *Bilingualism*, *17*(3), 610–629. https://doi.org/10.1017/S1366728913000564

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. *Language Learning & Language Teaching*, 21–46. https://doi.org/10.1075/LLLT.32.02BUL.

Burke, D., Mackay, D., Worthley, J., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and old adults? *Journal of Memory and Language - J MEM LANG*. *30*. 542-579.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*(1), 177–208. https://doi.org/10.1080/026432997381664

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: *A review. International Journal of Bilingualism, 7*(1), 71-87.

Costa, A. (2005). Lexical Access in Bilingual Production. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*, 289–307.

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.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. Journal of Experimental Psychology: Learning, Memory, and Cognition, *26*(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283.

Cummins J. (1978). Bilingualism and the development of metalinguistic awareness. J. *Cross-Cult. Psychol. 9*, 131– 49.

De Bot, K. (1992). A Bilingual Production Model: Levelt's 'Speaking' Model Adapted. *Applied Linguistics, 13*, 1-24.

De Groot, A. M. B., & Poot, R. (1997). Word translation at three levels of proficiency in a second language: The ubiquitous involvement of conceptual memory. *Language Learning*, *47*, 215-264.

de Jong, N. H. (2016). Fluency in second language assessment. In D. Tsagari, & J. Banerjee (Eds.), *Handbook of second language assessment*, 203-218.

de Jong, N.H., Groenhout, R., Schoonen, R., & Hulstijn, J. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*(2), 223-243. doi:10.1017/S0142716413000210.

de Jong, N.H., Steinel, M.P., Florijn, A.F., Schoonen, R., & Hulstijn, J.H. (2012). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics, 34*(5). doi:10.1017/S0142716412000069.

DeKeyser, R. (2000). The Robustness of Critical Period Effects in Second Language Acquisition. *Studies in Second Language Acquisition. 22*, 499-533.

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics, 31(*3), 413–438. https://doi.org/10.1017/S0142716410000056.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*(1-3), 287–314. https://doi.org/10.1016/0010-0277(92)90046-k

Derwing, T.M., Munro, M.J., Thomson, R.I., & Rossiter, M.J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition, 31*(4), 533-557. doi:10.1017/S0272263109990015.

Dijkstra, A., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. Bilingualism: *Language and Cognition, 5*, 175-197.

Dijkstra, A., Grainger, J., & Van Heuven, W. J. B. (1999). Recognizing cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language, 41*, 496-518.

Duran-Karaoz, Z., & Tavakoli, P. (2020). PREDICTING L2 FLUENCY FROM L1 FLUENCY BEHAVIOR: THE CASE OF L1 TURKISH AND L2 ENGLISH SPEAKERS. *Studies in Second Language Acquisition, 42*(4), 671-695. doi:10.1017/S0272263119000755.

FAB. (n.d.). *Fitness, Ageing, Bilingualism project*. OSF. Retrieved April 11, 2022, from https://osf.io/6fqg7.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*(1), 78–104. https://doi.org/10.1006/jmla.1999.2638.

Flege, J.E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis*, 101–131.

Flognfeldt, M,E., Tsagari, D., Šurkalović, D., Tishakov, T. (2020). The practice of assessing Norwegian and English language proficiency in multilingual elementary school classrooms in Norway. *Language Assessment Quarterly*, 17(5), 519-540.

Foster, P. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics - APPL LINGUIST*, *21*, 354-375. 10.1093/applin/21.3.354.

Francis, W. S., Tokowicz, N., & Kroll, J. F. (2003). Translation priming as a function of bilingual proficiency and item difficulty.

Gan, Z. (2012). Complexity Measures, Task Type, and Analytic Evaluations of Speaking Proficiency in a School-Based Assessment Context. *Language Assessment Quarterly*, *9*, 133-151. doi: 10.1080/15434303.2010.516041.

Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationship between temporal measures of fluency and oral English proficiency with implications for automated scoring. Language Testing, *27*, 379-399.

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: Language and Cognition*, *4*(1), 63–83. https://doi.org/10.1017/S136672890100013X

Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., & Jernigan, T. L. (2007). The bilingual effect on Boston Naming Test performance. *Journal of the International Neuropsychological Society : JINS*, *13*(2), 197–208. https://doi.org/10.1017/S1355617707070038.

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, *16*(4), 562–576. https://doi.org/10.1037/0894-4105.16.4.562

Gollan, T., Montoya, R., Bonanni, M. (2005a). Proper Names Get Stuck on Bilingual and Monolingual Speakers' Tip of the Tongue Equally Often.. *Neuropsychology*. *19*, 278-87.

Gollan, T., Montoya, R., Fennema-Notestine & C & Morris, S. (2005b). Bilingualism affects picture naming but not picture classification. *Memory & cognition. 33*, 1220-34. https://doi.org/10.3758/BF03193224.

Goodglass, H., Kaplan, E. and Barresi, B. (2001) The Assessment of Aphasia and Related Disorders.

Granena, G., & Long, M.H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research, 29*(3), 311–343. https://doi.org/10.1177/0267658312461497.

Green, D. (1998a). Author's response: Schemas, tags, and inhibition. *Bilingualism. Language and Cognition, 1*, 100-104.

Green, D. (1998b). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1* , 67-81.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377.

Hansen Edwards, J. (2017). Pronunciation and individual differences. In O. Kang. R. Thomson. & J. Murphy (Eds.) *The Routledge Handbook of Contemporary English pronunciation*, 385-398.

Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, *1*, 213-229.

Heyselaar, E., Segaert, K., & Wheeldon, L. (2021, July 20). Updated Cookie Theft Picture Set. https://doi.org/10.17605/OSF.IO/SHWCN

Huensch, A., & Tracy–Ventura, N. (2017). L2 Utterance Fluency Development Before, During, and After Residence Abroad: A Multidimensional Investigation. *The Modern Language Journal*, 101: 275-293. https://doi.org/10.1111/modl.12395

Hulstijn, J. (2012). The construct of language profciency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition*, *15*, 422–433. doi:10.1017/S1366728911000678.

Hulstijn, J.H. (2011). Language Proficiency in Native and Nonnative Speakers: An Agenda for Research and Suggestions for Second-Language Assessment, *Language Assessment Quarterly, 8*(3), 229-249. https://doi.org/10.1080/15434303.2011.565844.

Ishikawa., T. (2018). Bilingual speech production models.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production?. *Acta psychologica*, *127*(2), 277–288. https://doi.org/10.1016/j.actpsy.2007.06.003

Iwashita, N., Mcnamara, T & Elder, C. (2001). Can We Predict Task Difficulty in an Oral Proficiency Test? Exploring the Potential of an Information-Processing Approach to Task Design. *Language Learning, 51,* doi:10.1111/0023-8333.00160.

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language*, *44*, 2-31.

Jared, D., & Szucs, C. (2002). Phonological activation in bilinguals: Evidence from interlingual homograph recognition. Bilingualism, *Language and Cognition*, *5*, 225-239.

Jia, G., Aaronson, D., Wu, Y.H. (2002). Long-term language attainment of bilingual immigrants. Predictive factors and language group differences. *Applied Psycholinguistics, 23,* 599-621.

Johnson, J.S., & Newport, E.L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language**.** *Cognitive Psychology, 21*(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0.

Kałamała, P., Szewczyk, J., Chuderski, A., Senderecka, M., Wodniecka, Z. (2020). Patterns of bilingual language use and response inhibition: A test of the adaptive control hypothesis. *Cognition*. *204*. doi: 104373. 10.1016/j.cognition.2020.104373.

Klein, D., Milner, B., Zatorre, R., Meyer, E. & Evans, A. (1995). The neural substrates underlying word generation: a bilingual functional-imaging study. Proceedings of the National Academy of Sciences USA, *92*, 2899-2903.

Klein, D., Zatorre, R., Milner, B., Meyer, E. & Evans, A. (1994). Left putaminal activation when speaking a second language: evidence from PET. *Neuroreport*, (5), 2295-2297.

Köpke, B., Howells, R., Cortelazzo, F., Péran, P., De Boissezon, X., Lubrano, V. (2021). Functional and structural differences in brain networks involved in language processing and control in highly proficient early and late bilinguals. Journal of Neurolinguistics. *59*, 100988. 10.1016/j.jneuroling.2021.100988.

Kormos, J & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*. 145-164. https://doi.org/10.1016/j.system.2004.01.001.

Kormos, J. (2006). *Speech production and second language acquisition.* Lawrence Erlbaum Associates Publishers.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149-174.

Kroll, J. F., & Tokowicz, N. (2005). Models of Bilingual Representation and Processing: Looking Back and to the Future. In J. F. Kroll, & A. M. B. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches*, 531-553,

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: evidence for inhibitory processes. *Acta psychologica*, *128*(3), 416–430. https://doi.org/10.1016/j.actpsy.2008.02.001.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. Second Language Research, *18*, 137-171.

Krulatz, A., Dahl, A., & Flognfeldt, M. E. (2018). Enacting multilingualism: From research to teaching practice in the English classroom, *1*(1) Oslo: Cappelen Damm Akademisk.

La Heij, W., (2005). Selection processes in Monolingual and Bilingual lexical access. In J.F., Kroll, & A.M.B. de Groot (Eds.). *Handbook of bilingualism: Psycholinguistic approaches*, 289-307.

Leclercq, P., Edmonds, A., & Hilton, H. (2014). *Measuring L2 proficiency: Perspectives from Sla*. Multilingual Matters, https://doi.org/10.21832/9781783092291.

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological bulletin*, *144*(4), 394–425. https://doi.org/10.1037/bul0000142.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1995). The ability to speak: From intentions to spoken words. *European Review*, *3*, 13–23.

Levelt, W. J. M. (1999a). Language production: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *Neurocognition of language*, 83–122.

Levelt, W. J. M. (1999b). Models of word production. *Trends in Cognitive Sciences*, *3*, 223–232.

Levelt, W. J. M. , Roelofs, A. , & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Science, 22*, 1–38.

Lightbown, P., & Spada, N. M. (2013). *How languages are learned*.

Linck, J., Schwieter, J. W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. Bilingualism: *Language and Cognition, 15*(3), 651–662. doi:10.1017/S136672891100054X.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of cognitive psychology*, *25*(5), 605–621. https://doi.org/10.1080/20445911.2013.795574.

Marian, V., Blumenfeld, H.K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967.

Meuter, R.F.I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, (40), 25–40.

Michel, M., Kuiken, F. & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *Journal of Dental Research - J DENT RES, 45*, 241-259. doi:10.1515/iral.2007.011.

Moyer, A. (2004). Age, Accent and Experience in Second Language Acquisition. https://doi.org/10.21832/9781853597190.

Nadri, M., Baghaei, P., Zohoorian, Z. (2019). The contribution of cognitive abilities and general language proficiency to explaining listening comprehension in English as a foreign language. *Cogent Education*. *6*. 10.1080/2331186X.2019.1567010.

Norris, J.M., Ortega, K. (2009). Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity. *Applied Linguistics*, *30*(4), 555–578. https://doi.org/10.1093/applin/amp044

Norwegian Department of Education (2020). Kompetansemål og Vurdering Engelsk (ENG1-04). Retrieved February 2, 2022, from https://www.udir.no/lk20/eng01-04/kompetansemaal-og-vurdering/kv4.

Øystein A. Vangsnes, Göran B. W. Söderlund & Morten Blekesaune (2017) The effect of bidialectal literacy on school achievement, International Journal of Bilingual Education and Bilingualism, *20*(3), 346-361, DOI: 10.1080/13670050.2015.1051507.

Patkowski, M.S. (1980). THE SENSITIVE PERIOD FOR THE ACQUISITION OF SYNTAX IN A SECOND LANGUAGE . *Language* Learning, 30, 449-468. https://doi.org/10.1111/j.1467-1770.1980.tb00328.x

Peal, E., & Lambert, W. (1962). The Relation Of Bilingualism To Intelligence. *Psychological Monographs: General and Applied*, *76*(27), 1-23.

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, *15*(2), 202–206. https://doi.org/10.1016/j.conb.2005.03.007

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 539–557. https://doi.org/10.1037/0278-7393.24.3.539

Portocarrero, J. S., Burright, R. G., & Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, *22*(3), 415–422. https://doi.org/10.1016/j.acn.2007.01.015.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics, 15,* 36–57.

Poulisse, Nanda. (1999). Slips of the Tongue in First and Second Language Production. *Studia Linguistica*. *54.* 136 - 149.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes, 14*(4), 423–441. https://doi.org/10.1080/01638539109544795.

Roelofs, A. (2003). Shared phonological encoding processes and representations of languages in bilingual speakers. *Language and Cognitive Processes*, *18*(2), 175-204.

Runnqvist, E., Gollan, T.H., Costa, A., & Ferreira, V.S. (2013). A disadvantage in bilingual sentence production modulated by syntactic frequency and similarity across languages. *Cognition*, *129*, 256-263.

Schwartz, A., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language & Cognitive processes, 22(*1), 106-129, https://doi.org/10.1080/01690960500463920.

Schwieter, J. W. (2010). Cognition and bilingual speech: Psycholinguistic aspects of language production, processing, and inhibitory control. *Lambert Academic.*

Schwieter, J. W. (2013). Lexical inhibition in trilingual speakers. In J. Tirkkonen & E. Anttikoski (Eds.), Proceedings of The 24th Conference of Scandinavian Linguistics. *Reports and Studies in Education, Human*.

Schwieter, J. W., & Ferreira, A. (2013). Language selection, control, and conceptual-lexical development in bilinguals and multilinguals. In J. W. Schwieter (Ed.), *Innovative research and practices in second language acquisition and bilingualism*, 241–266.

Schwieter, J. W., & Sunderman, G. (2008). Language switching in bilingual speech production: In search of the language-specifc selection mechanism. *The Mental Lexicon, 3*(2), 214–238.

Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. Routledge.

Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching, 54*, 79-95.

Smith F. (1923). Bilingualism and mental development. *Br*. J. *Psychol*. *13,* 271–82.

Sunderman, G and Kroll, J F. (2006). First language activation during second language lexical processing. An Investigation of Lexical Form, Meaning and Grammatical Class. Studies in Second Language acquisition, *28*, 387-422.

Sunderman, G. (2002). Lexical development in a second language: Can the first language be suppressed? Unpublished doctoral dissertation.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition, 2*(1), 45–58. https://doi.org/10.1017/S1366728999000140.

Tarłowski, A., Wodniecka, Z., & Marzecová, A. (2013). Language switching in the production of phrases. *Journal of Psycholinguistic Research, 42,* 103–118 .

Tavakoli, P. and Foster, P. (2008), Task Design and Second Language Performance: The Effect of Narrative Type on Learner Output. *Language Learning*, *58,* 439-473. https://doi.org/10.1111/j.1467-9922.2008.00446.x.

Tavakoli, P., & Hunter, A.M. (2017). Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*, *22*. 10.1177/1362168817708462.

Tzelgov, J., & Eben-Ezra, S. (1992). Components of the between-language semantic priming effect. *European Journal of Cognitive Psychology*, *4*, 253-272.

Vaid, J. & Menon, R.. (2000). Correlates of bilinguals' preferred language for mental computations. *Spanish Applied Linguistics, 42*, 325-342.

van den Noort, M., Struys, E., Bosch, P., Jaswetz, L., Perriard, B., Yeo, S., Barisch, P., et al. (2019). Does the Bilingual Advantage in Cognitive Control Exist and If So, What Are Its Modulating Factors? A Systematic Review. *Behavioral Sciences, 9*(3), 27. MDPI AG. Retrieved from http://dx.doi.org/10.3390/bs9030027.

Van Hell, J. G ., De Groot, A. M. B., Dannenburg, L. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, *33*, 600-629.

Van Hell, J., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance: Evidence from trilinguals. *Psychonomic Bulletin and Review*, *9*, 780-789.

Yetkin, O., Yetkin, F.Z., Haughton, V. M., & Cox, R. W. (1996). Use of functional MR to map language in multilingual volunteers. *American Journal of Neuroradiology*, (17), 473–477.

Yoshioka JG. 1929. A study of bilingualism. J. Genet. *Psychol*, *36*, 473–79.

Indefrey, P., & Levelt, W. J. M. (2000). The speaking mind/brain: Where do spoken words come from? In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium*, 77–93.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning & Verbal Behavior*, *5*(4), 325–337. https://doi.org/10.1016/S0022-5371(66)80040-3

Harley, T.A. and Bown, H.E. (1998), What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. British Journal of Psychology, 89: 151-174. https://doi.org/10.1111/j.2044-8295.1998.tb02677.x

Wheeldon, L. R., & Monsell, S. (1992). The locus of repetition priming of spoken word production. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *44A*(4), 723–761. https://doi.org/10.1080/14640749208401307

Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 686–698. https://doi.org/10.1037/0278-7393.21.3.686

14. Appendices

Appendix A: LEAP-Q

Appendix LEAP-Q

Appendix (p. 1 of 2). Language Experience and Proficiency Questionnaire.

Last Name		First Name		Today's Date	
Age		Date of Birth		Male g	Female g

(1) Please list all the languages you know in order of dominance:

1 Language A	2 Language B	3 Language C	4 Language D	5 Language E

(2) Please list all the languages you know in order of acquisition (your native language first):

1 Language A	2 Language B	3 Language C	4 Language D	5 Language E

(3) Please list what percentage of the time you are CURRENTLY and *on average* exposed to each language.
(*YOUR percentages SHOULD add UP to 100%*):

List language here:	Language A	Language B	Language C	Language D	Language E
List percentage here:					

(4) When choosing to read a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you.
(*YOUR percentages SHOULD add UP to 100%*):

List language here	Language A	Language B	Language C	Language D	Language E
List percentage here:					

(5) When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of time would you choose to speak each language? Please report percent of total time.
(*YOUR percentages SHOULD add UP to 100%*):

List language here	Language A	Language B	Language C	Language D	Language E
List percentage here:					

(6) Please name the cultures with which you identify. On a scale from zero to ten, please rate the extent to which you identify with each culture. (Examples of possible cultures include US-American, Chinese, Jewish-Orthodox, etc.):

List cultures here	Culture A (click here for scale)	Culture B (click here for scale)	Culture C (click here for scale)	Culture D (click here for scale)	Culture E (click here for scale)

(7) How many years of formal education do you have?____________________
Please check your highest education level (or the approximate U.S. equivalent to a degree obtained in another country):

g Less than High School
g High School
g Professional Training
g Some College
g College
g Some Graduate
g Masters
g PhD/MD/JD
g Other:

(8) Date of immigration to the United States, if applicable: ____________________
If you have ever lived in another country, please provide name of country and dates of residence:

____

(9) Have you ever had a vision problem g, hearing impairment g, language disability g, or learning disability g? (Check all applicable).
If yes, please explain (including any corrections): ____________________

Appendix (p. 2 of 2). Language Experience and Proficiency Questionnaire.

Language: Language X

This is my (please select from scroll-down menu: First, Second, Third, etc.) language.
All questions below refer to your knowledge of Language X.

(1) Age when you ▮ :

began ACQUIRING Language X:	*became* FLUENT in Language X:	*began reading* in Language X:	*became* FLUENT *reading* in Language X:

(2) Please list the number of years and months you spent in each language environment:

	Years	Months
A country where Language X is spoken		
A family where Language X is spoken		
A school and/or working environment where Language X is spoken		

(3) On a scale from zero to ten, please select your *level of proficiency* in speaking, understanding, and reading Language X from the scroll-down menus:

Speaking	(click here for scale)	Understand spoken language	(click here for scale)	Reading	(click here for scale)

(4) On a scale from zero to ten, please select how much the following factors contributed to you learning Language X:

Interacting with friends	(click here for scale)	Language tapes/self instruction	(click here for scale)
Interacting with family	(click here for scale)	Watching TV	(click here for scale)
Reading	(click here for scale)	Listening to the radio	(click here for scale)

(5) Please rate to what extent you are currently exposed to Language X in the following contexts:

Interacting with friends	(click here for scale)	Listening to radio/music	(click here for scale)
Interacting with family	(click here for scale)	Reading	(click here for scale)
Watching TV	(click here for scale)	Language-lab/self-instruction	(click here for scale)

(6) In your perception, how much of a foreign accent do you have in Language X?
(click here for scale)

(7) Please rate how frequently others identify you as a non-native speaker based on your accent in Language X:
(click here for scale)

Appendix B: LEAP-Q Adapted

Q1 Welcome. In this questionnaire we will ask you questions about your language usage and proficiency. Some of these questions are a little tricky; just do your best and if you are unsure, please provide an estimate. The questionnaire should take around eight minutes to complete. Click on the arrow below to begin.

Q2 Please enter your participant ID (please do not insert any additional characters or symbols, such as a space or full-stop).

____

End of Block: Participant Identification

Start of Block: Block 2

Q8 Please list all the languages you speak in order of DOMINANCE (the more dominant language is the one you use more; up to 5).

○ 1 (1) ____

○ 2 (2) ____

○ 3 (3) ____

○ 4 (4) ____

○ 5 (5) ____

Q9 Please list all the languages you speak in order of ACQUISITION (acquisition refers to when you have learned the language; up to 5).

○ 1 (1) ____

○ 2 (2) ____

○ 3 (3) ____

○ 4 (4) ____

○ 5 (5) ____

Q10 Please list what percentage of the time you are on average exposed to each language (e.g. exposure in terms of talking, listening, and reading, including TV, films and music). All your answers should add up to 100%.

○ ____

○ ____

○ ____

○ ____

○ ____

Q11 Please list what percentage of the time you spend speaking each language. All your answers should add up to 100%.

○ ____

○ ____

○ ____

Q12 Please list what percentage of the time you typically spend reading in each language. All your answers should add up to 100%.

○ ____

○ ____

○ ____

○ ____

○ ____

Q13 When choosing a language to speak, with a person who is equally fluent in all your languages, what percentage of time would you choose to speak each language? Please report percentage of total time. All your answers should add up to 100%.

○ ____

○ ____

○ ____

○ ____

○ ____

End of Block: Block 2

Start of Block: Block 3

Q14 Do you feel that you were once better in one of your languages and that you have become less fluent?

◯ Yes (please explain which one) (____

◯ No

Q15 In which language do you usually do the following tasks?

◯ Simple maths (count, add) ____

◯ Dream ____

◯ Express anger or affection ____

◯ Talk to yourself ____

End of Block: Block 3

Start of Block: NO and ENG Proficiency

Q16 Please answer the following questions about your experience with Norwegian **and** English.

Q17 Please list the number of years and months you have spent in each language environment. E.g., if your response is 6 months, please enter 0y 6m. If your response is 2 years and 3 months, please enter 2y 3m.

	Norwegian	English
A country where this language is spoken		
A family where this language is spoken		

Q18 Please rate how much the following factors contributed to your learning of each language on a scale of 0-10 whereby 0 = not a contributor, 5 = moderate contributor and 10 = most important contributor.

	Norwegian	English
Interacting with friends / colleagues		
Interacting with family		
Reading (e.g., books, magazines, online material)		
School and education		
Self-instruction (e.g., language learning videos or apps)		
Watching TV / streaming / listening to music / media		
Interactive gaming		

Q19 Please rate to what extent you are currently (e.g. in the last month or so) exposed to each language on a scale of 0-10 whereby 0 = never, 5 = half of the time and 10 = almost always.

	Norwegian	English
Interacting with friends / colleagues		
Interacting with family		
Reading (e.g., books, magazines, online material)		
School and education		
Self-instruction (e.g., language learning videos or apps)		
Watching TV / streaming / listening to music / media		
Interactive gaming		

Q20 Please rate your level of proficiency in the following aspects of each language on a scale of 0-10 whereby: 0 = none; 1 = very low; 2 = low; 3 = fair; 4 = slightly less than adequate; 5 = adequate; 6 = slightly more than adequate; 7 = good; 8 = very good; 9 = excellent; 10 = perfect.

	Norwegian	English
Speaking (general fluency)		
Pronunciation (accent)		
Listening (understanding spoken language)		
Reading		
Writing		
Vocabulary		

Q21 Please list the AGE (in years) you were when the following occurred for each language.

	Norwegian	English
Started hearing this language		
Started speaking this language		
Started reading this language		

Q22 How often do you have to switch between your languages in order to communicate effectively with others?

- ◯ Roughly daily
- ◯ Roughly weekly
- ◯ Roughly monthly
- ◯ Very rarely

Q23 Please rate your level of proficiency in switching between your languages when you need to, on a scale of 0-10 whereby: 0 = none; 1 = very low; 2 = low; 3 = fair; 4 = slightly less than adequate; 5 = adequate; 6 = slightly more than adequate; 7 = good; 8 = very good; 9 = excellent; 10 = perfect.

____

Q24 When you are speaking do you ever find yourself **accidentally** mixing words or sentences from Norwegian and English?

○ Yes (4)

○ No (5)

Skip To: Q27 If When you are speaking do you ever find yourself accidentally mixing words or sentences from Norwegian... = No

Q25 (a) If yes, how often does English accidentally intrude in your Norwegian on a scale of 0-10 (whereby 0 = never, 5 = half of the time, 10 = all of the time)?

____

Q26 (b) And how often does Norwegian accidentally intrude into your English on a scale of 0-10 (whereby 0 = never, 5 = half of the time, 10 = all of the time)?

____

Q27 When you are speaking with a person who also knows both Norwegian and English do you ever find yourself intentionally mixing words or sentences from Norwegian and English?

○ Yes (4)

○ No (5)

Skip To: End of Block If When you are speaking with a person who also knows both Norwegian and English do you ever find yoouself intentionally mixing words or sentences from Norwegian and English? = No

Q28 (a) If yes, how often do you intentionally use English words or sentences when speaking Norwegian on a scale of 0-10 (whereby 0 = never, 5 = half of the time, 10 = all of the time)?

____

Q29 (b) And how often do you intentionally use Norwegian words or sentences when speaking English on a scale of 0-10 (whereby 0 = never, 5 = half of the time, 10 = all of the time)?

____

Appendix C: Vocabulary depth Task Stimuli

Table 18: ANTONYMS TASK LIST

TEST	Item	Word	Length	WordFreq	DomPos_UK	DomPos_dic	Correct	FoilA	FoilB	FoilC	CorrectFreq	FoilAFreq	FoilBFreq	FoilCFreq	Source
anto	1	concerned	6	4,93	verb	a	uncaring	scarce	misleading	understanding	2,48	3,51	3,64	4,47	SubtlexUK
anto	2	timorous	8	1,81	adjective	a	fearless	forestry	funny	emotive	3,51	3,56	5,06	3,22	SubtlexUK
anto	3	disdain	9	2,73	noun	n	admire	unload	misfortune	huge	3,97	3,12	3,18	5,32	SubtlexUK
anto	4	acerbic	7	2,17	adjective	a	sweet	itchy	loud	beautiful	4,98	3,45	4,48	5,42	SubtlexUK
anto	5	nonplus	7			v	enlighten	subtract	gain	disadvantage	3,12	2,62	4,34	3,6	SubtlexUK
anto	6	surfeit	7	1,84	noun	n	lack	southern	excess	fake	4,62	4,5	4	4,51	SubtlexUK
anto	7	vicious	7	3,87	adjective	a	gentle	slippery	fierce	disobedient	4,23	3,84	3,97	2,34	SubtlexUK
anto	8	saunter	7	2,25	verb	v	rush	fry	punish	daydream	4,44	4,19	3,64	2,9	SubtlexUK
anto	9	slipshod	8	1,6	adjective	a	careful	difficult	clumsy	footwear	4,84	5,4	3,63	3,24	SubtlexUK
anto	10	umbrage	7	2,14	noun	a/n	delight	dungeon	demanding	appeal	4,04	3,14	4,09	4,59	SubtlexUK
anto	11	strenuous	9	2,92	adjective	a	effortless	arduous	smooth	tricky	3,13	3,06	4,36	4,51	SubtlexUK
anto	12	divulge	7	2,67	verb	v	conceal	purchase	disclose	smuggle	3,24	4,28	3,22	3,12	SubtlexUK
anto	13	loathe	6	3,16	verb	v	cherish	rejoice	kindle	undress	3,37	3,22	2,73	2,68	SubtlexUK
anto	14	querulous	9	1,3	adjective	a	agreeable	feathered	blatant	squeaky	3,03	3,14	3,19	3,42	SubtlexUK
anto	15	forgo	5	2,55	verb	v	acquire	precede	journey	disappear	3,46	2,53	4,91	4,21	SubtlexUK
anto	16	conquer	7	3,64	verb	v	surrender	demand	retain	release	3,76	4,56	3,9	4,6	SubtlexUK
anto	17	hovel	5	2,57	noun	n	palace	float	cloudy	stairwell	4,55	3,96	4,17	2,93	SubtlexUK
anto	18	adversity	9	3,24	noun	n	advantage	delay	grudge	persevere	4,67	4,11	3,19	2,85	SubtlexUK
anto	19	alacrity	8	2,08	noun	n	slowness	annoyance	fog	ingenuity	2,33	3,03	4,08	3,31	SubtlexUK
anto	20	penury	6	2,02	noun	n	wealth	dispatch	cunning	famine	4,33	3,34	3,82	3,62	SubtlexUK

Table 19: SYNONYMS TASK LIST

TEST	Item	Word	Length	WordFreq	DomPos_UK	DomPos_dic	Correct	FoilA	FoilB	FoilC	CorrectFreq	FoilAFreq	FoilBFreq	FoilCFreq	Source
syno	1	caprice	7	2,73	noun	n	whim	cattle	brute	lounge	3,16	4,25	3,33	4,19	SubtlexUK
syno	2	baffle	7	2,6	verb	v	confuse	hide	warp	bully	3,5	4,66	3,19	3,72	SubtlexUK
syno	3	ponderous	9	2,39	adjective	a	unwieldy	useless	supportive	thoughtful	2,56	4,1	3,87	3,52	SubtlexUK
syno	4	banter	8	3,66	noun	n/v	chatting	whispering	denial	beating	4,03	3,43	3,6	4,31	SubtlexUK
syno	5	garish	6	2,85	adjective	a	tasteless	spiky	green	bland	2,96	3,51	5,22	3,5	SubtlexUK
syno	6	sequin	6	2,36	noun	n	bead	stamp	sledge	order	2,93	4,2	3,3	5,15	SubtlexUK
syno	7	loquacious	10	1,87	adjective	a	talkative	broad	roomy	marshy	2,69	4,22	2,95	2,63	SubtlexUK
syno	8	covet	5	2,39	verb	v	desire	pad	cradle	cave	4,24	3,97	3,46	4,19	SubtlexUK
syno	9	acumen	6	2,65	noun	n	cleverness	blame	spicy	wealth	2,44	4,66	3,98	4,33	SubtlexUK
syno	10	drench	6	2,06	noun	n/v	soak	raise	erase	flatten	3,78	4,85	3,08	3,24	SubtlexUK
syno	11	abide	5	3,45	verb	v	endure	inhabit	crave	depart	3,5	2,99	3,13	3,27	SubtlexUK
syno	12	vocation	8	3,14	noun	n	occupation	holiday	pronunciation	vocabulary	3,86	4,94	3,09	3,43	SubtlexUK
syno	13	gulch	5	1,81	name	n	crevasse	swallow	shed	dislike	2,42	3,89	4,38	3,52	SubtlexUK
syno	14	cogitate	8	1,9	verb	v	ponder	achieve	succeed	enquire	3,06	4,66	4,15	2,75	SubtlexUK
syno	15	vexatious	9	2,02	adjective	a	effortful	engaging	horrifying	priceless	1,3	3,75	3,19	3,7	SubtlexUK
syno	16	peril	5	3,39	noun	n	danger	shiny	delight	shelter	4,75	4,15	4,04	4,13	SubtlexUK
syno	17	feral	5	3,13	adjective	a	savage	hungry	impartial	ugly	3,83	4,55	3,24	4,23	SubtlexUK
syno	18	ludicrous	9	4,41	adjective	a	ridiculous	developed	nasty	certain	4,57	4,51	4,45	4,99	SubtlexUK
syno	19	brisk	5	3,35	adjective	a	energetic	disposable	section	stern	3,58	3,31	4,53	3,57	SubtlexUK
syno	20	truculent	9	1,74	adjective	a	defiant	delicious	juicy	tardy	3,22	4,66	3,86	2,3	SubtlexUK

Appendix D: Vocabulary and Comprehension E-Mail Example

Dear __________________

Thank you so much for having completed the two online tasks sent before.

Now, we would like you to complete **two similar digital tasks, but this time in English.** As the ones before, these will take approx. 15 minutes. Unfortunately, it is not possible to carry out the tests on a mobile phone, so here you have to find your PC or laptop. It is important that you complete the entire test, and do not press "escape" or close the tab until the green message saying "Thank you for your patience" appears.

Task1: Requires you to choose words that are either similar or opposite in meaning to the word presented.

https://pavlovia.org/run/foyzulr/fab_vocabtask

Task2. Here you will read a series of sentences and answer presented follow-up questions. Some of the sentences can be difficult to understand - this is part of the task. In this task, it is importanthat you press the space key as soon as you have finished reading the sentence. This is because we want to know how fast you can process the information.

https://pavlovia.org/run/foyzulr/textcomp_eng_l4a

Remember to use the participant ID _________, which you have already received. This is entered under "participant", while you can leave "session" untouched.

After completion of these tasks we can book the sessions in our lab!

Appendix E: Vocabulary Task Information on Monitor

Introduction information first page	Introduction information second page
Welcome. In this task, you are required to choose a word that is either similar or opposite in meaning to the centrally presented word. The central word will be presented in big bold letters. You will also see another four word options at the bottom – your job is to simply pick one of these words which you believe is either similar opposite in meaning. Press SPACE to continue.	Trials where you are picking similar words will appear in one block and trials where you are picking opposite words will appear in a separate block. At the beginning of each block, you will be instructed with whether you should select Similar or Opposite words. Use the mouse to select (left click) your chosen response option. You should complete this task in one sitting and please do not attempt to get help from any external sources (phone, internet, friend, etc.). Press SPACE to begin experiment.
Experiment information page (Similar or opposite depending on whether it is synonym or antonym block)	**Experiment testing page (Similar or opposite depending on whether it is synonym or antonym block)**
Choose a word with similar meaning	SIMILAR Caprice whim cattle brute lounge

Appendix F: Text Comprehension Stimuli

item	syntax level	AnTrans Verb	freq	sylls	letters	InanTrans Verb	freq	sylls	letters	Adjective	freq	sylls	letters	Inan Noun	freq	sylls	letters	Lead Noun	An Noun	Gender	Sentence	Question1	Answer1	Question2	Answer2
0	1	teasing	3,44	2	7	colouring	3,61	3	9	silly	4,61	2	5	picture	5,13	2	7	girl	boy	she	The girl is teasing the boy and she is colouring a silly picture.	Who is colouring the silly picture?	girl	Who is being teased?	boy
0	2	teasing	3,44	2	7	colouring	3,61	3	9	silly	4,61	2	5	picture	5,13	2	7	girl	boy	she	The girl who is teasing the boy is colouring a silly picture.	Who is colouring the silly picture?	girl	Who is being teased?	boy
0	3	teased	3,44	1	6	colouring	3,61	3	9	silly	4,61	2	5	picture	5,13	2	7	girl	boy	she	The boy is being teased by the girl who is colouring a silly picture.	Who is colouring the silly picture?	girl	Who is being teased?	boy
0	4	teasing	3,44	2	7	colouring	3,61	3	9	silly	4,61	2	5	picture	5,13	2	7	girl	boy	she	The picture which the girl who is teasing the boy is colouring is silly.	Who is colouring the silly picture?	girl	Who is being teased?	boy
1	1	blessing	3,81	2	8	hugging	3,42	2	7	fuzzy	3,28	2	5	cushion	4,25	2	7	boy	girl	he	The boy is blessing the girl and he is hugging a fuzzy cushion.	Who is hugging the fuzzy cushion?	boy	Who is being blessed?	girl
1	2	blessing	3,81	2	8	hugging	3,42	2	7	fuzzy	3,28	2	5	cushion	4,25	2	7	boy	girl	he	The boy who is blessing the girl is hugging a fuzzy cushion.	Who is hugging the fuzzy cushion?	boy	Who is being blessed?	girl
1	3	blessing	3,81	2	8	hugging	3,42	2	7	fuzzy	3,28	2	5	cushion	4,25	2	7	boy	girl	he	The girl is being blessed by the boy who is hugging a fuzzy cushion.	Who is hugging the fuzzy cushion?	boy	Who is being blessed?	girl
1	4	blessing	3,81	2	8	hugging	3,42	2	7	fuzzy	3,28	2	5	cushion	4,25	2	7	boy	girl	he	The cushion which the boy who is blessing the girl is hugging is fuzzy.	Who is hugging the fuzzy cushion?	boy	Who is being blessed?	girl
2	1	annoying	4,26	3	8	treating	4,04	2	8	weeping	3,36	2	7	wound	4,1	1	5	girl	boy	she	The girl is annoying the boy and she is treating a weeping wound.	Who is treating the weeping wound?	girl	Who is being annoyed?	boy
2	2	annoying	4,26	3	8	treating	4,04	2	8	weeping	3,36	2	7	wound	4,1	1	5	girl	boy	she	The girl who is annoying the boy is treating a weeping wound.	Who is treating the weeping wound?	girl	Who is being annoyed?	boy
2	3	annoying	4,26	3	8	treating	4,04	2	8	weeping	3,36	2	7	wound	4,1	1	5	girl	boy	she	The boy is being annoyed by the girl who is treating a weeping wound.	Who is treating the weeping wound?	girl	Who is being annoyed?	boy
2	4	annoying	4,26	3	8	treating	4,04	2	8	weeping	3,36	2	7	wound	4,1	1	5	girl	boy	she	The wound which the girl who is annoying the boy is treating is weeping.	Who is treating the weeping wound?	girl	Who is being annoyed?	boy
3	1	watching	5,12	2	8	swallowing	3,11	3	10	sticky	4,27	2	6	sweet	4,98	1	5	boy	girl	he	The boy is watching the girl and he is swallowing a sticky sweet.	Who is swallowing the sticky sweet?	boy	Who is being watched?	girl
3	2	watching	5,12	2	8	swallowing	3,11	3	10	sticky	4,27	2	6	sweet	4,98	1	5	boy	girl	he	The boy who is watching the girl is swallowing a sticky sweet.	Who is swallowing the sticky sweet?	boy	Who is being watched?	girl
3	3	watching	5,12	2	8	swallowing	3,11	3	10	sticky	4,27	2	6	sweet	4,98	1	5	boy	girl	he	The girl is being watched by the boy who is swallowing a sticky sweet.	Who is swallowing the sticky sweet?	boy	Who is being watched?	girl
3	4	watching	5,12	2	8	swallowing	3,11	3	10	sticky	4,27	2	6	sweet	4,98	1	5	boy	girl	he	The sweet which the boy who is watching the girl is swallowing is sticky.	Who is swallowing the sticky sweet?	boy	Who is being watched?	girl
4	1	frightening	4,06	2	11	concealing	2,87	3	10	mouldy	3,23	2	6	glove	3,81	1	5	girl	boy	she	The girl is frightening the boy and she is concealing a mouldy glove.	Who is concealing the mouldy glove?	girl	Who is being frightened?	boy

item	syntax level	AnTrans Verb	freq	sylls	letters	InanTrans Verb	freq	sylls	letters	Adjective	freq	sylls	letters	Inan Noun	freq	sylls	letters	Lead Noun	An Noun	Gender	Sentence	Question1	Answer1	Question2	Answer2
4	2	frightening	4,06	2	11	concealing	2,87	3	10	mouldy	3,23	2	6	glove	3,81	1	5	girl	boy	she	The girl who is frightening the boy is concealing a mouldy glove.	Who is concealing the mouldy glove?	girl	Who is being frightened?	boy
4	3	frightening	4,06	2	11	concealing	2,87	3	10	mouldy	3,23	2	6	glove	3,81	1	5	girl	boy	she	The boy is being frightened by the girl who is concealing a mouldy glove.	Who is concealing the mouldy glove?	girl	Who is being frightened?	boy
4	4	frightening	4,06	2	11	concealing	2,87	3	10	mouldy	3,23	2	6	glove	3,81	1	5	girl	boy	she	The glove which the girl who is frightening the boy is concealing is mouldy.	Who is concealing the mouldy glove?	girl	Who is being frightened?	boy
5	1	offending	3,45	3	9	knitting	3,59	2	8	wooly	2	2	5	mitten	2,43	2	6	boy	girl	he	The boy is offending the girl and he is knitting a wooly mitten.	Who is knitting the wooly mitten?	boy	Who is being offended?	girl
5	2	offending	3,45	3	9	knitting	3,59	2	8	wooly	2	2	5	mitten	2,43	2	6	boy	girl	he	The boy who is offending the girl is knitting a wooly mitten.	Who is knitting the wooly mitten?	boy	Who is being offended?	girl
5	3	offending	3,45	3	9	knitting	3,59	2	8	wooly	2	2	5	mitten	2,43	2	6	boy	girl	he	The girl is being offended by the boy who is knitting a wooly mitten.	Who is knitting the wooly mitten?	boy	Who is being offended?	girl
5	4	offending	3,45	3	9	knitting	3,59	2	8	wooly	2	2	5	mitten	2,43	2	6	boy	girl	he	The mitten which the boy who is offending the girl is knitting is wooly.	Who is knitting the wooly mitten?	boy	Who is being offended?	girl
6	1	avoiding	3,82	3	8	eating	4,82	2	6	juicy	3,86	2	5	peach	3,62	1	5	girl	boy	she	The girl is avoiding the boy and she is eating a juicy peach.	Who is eating the juicy peach?	girl	Who is being avoided?	boy
6	2	avoiding	3,82	3	8	eating	4,82	2	6	juicy	3,86	2	5	peach	3,62	1	5	girl	boy	she	The girl who is avoiding the boy is eating a juicy peach.	Who is eating the juicy peach?	girl	Who is being avoided?	boy
6	3	avoiding	3,82	3	8	eating	4,82	2	6	juicy	3,86	2	5	peach	3,62	1	5	girl	boy	she	The boy is being avoided by the girl who is eating a juicy peach.	Who is eating the juicy peach?	girl	Who is being avoided?	boy
6	4	avoiding	3,82	3	8	eating	4,82	2	6	juicy	3,86	2	5	peach	3,62	1	5	girl	boy	she	The peach which the girl who is avoiding the boy is eating is juicy.	Who is eating the juicy peach?	girl	Who is being avoided?	boy
7	1	embracing	3,4	3	9	admiring	3,35	3	8	bright	4,82	1	6	sunset	3,79	2	6	boy	girl	he	The boy is embracing the girl and he is admiring a bright sunset.	Who is admiring the bright sunset?	boy	Who is being embraced?	girl
7	2	embracing	3,4	3	9	admiring	3,35	3	8	bright	4,82	1	6	sunset	3,79	2	6	boy	girl	he	The boy who is embracing the girl is admiring a bright sunset.	Who is admiring the bright sunset?	boy	Who is being embraced?	girl
7	3	embracing	3,4	3	9	admiring	3,35	3	8	bright	4,82	1	6	sunset	3,79	2	6	boy	girl	he	The girl is being embraced by the boy who is admiring a bright sunset.	Who is admiring the bright sunset?	boy	Who is being embraced?	girl
7	4	embracing	3,4	3	9	admiring	3,35	3	8	bright	4,82	1	6	sunset	3,79	2	6	boy	girl	he	The sunset which the boy who is embracing the girl is admiring is bright.	Who is admiring the bright sunset?	boy	Who is being embraced?	girl
8	1	chasing	4,25	2	7	throwing	4,4	2	8	purple	4,29	2	6	toy	4,39	1	3	girl	boy	she	The girl is chasing the boy and she is throwing a purple toy.	Who is throwing the purple toy?	girl	Who is being chased?	boy
8	2	chasing	4,25	2	7	throwing	4,4	2	8	purple	4,29	2	6	toy	4,39	1	3	girl	boy	she	The girl who is chasing the boy is throwing a purple toy.	Who is throwing the purple toy?	girl	Who is being chased?	boy
8	3	chasing	4,25	2	7	throwing	4,4	2	8	purple	4,29	2	6	toy	4,39	1	3	girl	boy	she	The boy is being chased by the girl who is throwing a purple toy.	Who is throwing the purple toy?	girl	Who is being chased?	boy
8	4	chasing	4,25	2	7	throwing	4,4	2	8	purple	4,29	2	6	toy	4,39	1	3	girl	boy	she	The toy which the girl who is chasing the boy is throwing is purple.	Who is throwing the purple toy?	girl	Who is being chased?	boy

item	syntax level	AnTrans Verb	freq	sylls	letters	InanTrans Verb	freq	sylls	letters	Adjective	freq	sylls	letters	Inan Noun	freq	sylls	letters	Lead Noun	An Noun	Gender	Sentence	Question1	Answer1	Question2	Answer2
9	1	examining	3,59	4	9	toasting	2,99	2	8	stale	3,28	1	5	sandwich	4,25	2	8	boy	girl	he	The boy is examining the girl and he is toasting a stale sandwich.	Who is toasting the stale sandwich?	boy	Who is being examined?	girl
9	2	examining	3,59	4	9	toasting	2,99	2	8	stale	3,28	1	5	sandwich	4,25	2	8	boy	girl	he	The boy who is examining the girl is toasting a stale sandwich.	Who is toasting the stale sandwich?	boy	Who is being discussed?	girl
9	3	examining	3,59	4	9	toasting	2,99	2	8	stale	3,28	1	5	sandwich	4,25	2	8	boy	girl	he	The girl is being examined by the boy who is toasting a stale sandwich.	Who is toasting the stale sandwich?	boy	Who is being discussed?	girl
9	4	examining	3,59	4	9	toasting	2,99	2	8	stale	3,28	1	5	sandwich	4,25	2	8	boy	girl	he	The sandwich which the boy who is examining the girl is toasting is stale.	Who is toasting the stale sandwich?	boy	Who is being discussed?	girl
10	1	soothing	3,16	2	8	slicing	3,29	2	7	yellow	4,84	2	6	lemon	4,46	2	5	girl	boy	she	The girl is soothing the boy and she is slicing a yellow lemon.	Who is slicing the yellow lemon?	girl	Who is being soothed?	boy
10	2	soothing	3,16	2	8	slicing	3,29	2	7	yellow	4,84	2	6	lemon	4,46	2	5	girl	boy	she	The girl who is soothing the boy is slicing a yellow lemon.	Who is slicing the yellow lemon?	girl	Who is being soothed?	boy
10	3	soothing	3,16	2	8	slicing	3,29	2	7	yellow	4,84	2	6	lemon	4,46	2	5	girl	boy	she	The boy is being soothed by the girl who is slicing a yellow lemon.	Who is slicing the yellow lemon?	girl	Who is being soothed?	boy
10	4	soothing	3,16	2	8	slicing	3,29	2	7	yellow	4,84	2	6	lemon	4,46	2	5	girl	boy	she	The lemon which the girl who is soothing the boy is slicing is yellow.	Who is slicing the yellow lemon?	girl	Who is being soothed?	boy
11	1	kicking	4,35	2	7	translating	3,02	3	11	dull	4,04	1	4	novel	4,43	2	5	boy	girl	he	The boy is kicking the girl and he is translating a dull novel.	Who is translating the dull novel?	boy	Who is being kicked?	girl
11	2	kicking	4,35	2	7	translating	3,02	3	11	dull	4,04	1	4	novel	4,43	2	5	boy	girl	he	The boy who is kicking the girl is translating a dull novel.	Who is translating the dull novel?	boy	Who is being kicked?	girl
11	3	kicking	4,35	2	7	translating	3,02	3	11	dull	4,04	1	4	novel	4,43	2	5	boy	girl	he	The girl is being kicked by the boy who is translating a dull novel.	Who is translating the dull novel?	boy	Who is being kicked?	girl
11	4	kicking	4,35	2	7	translating	3,02	3	11	dull	4,04	1	4	novel	4,43	2	5	boy	girl	he	The novel which the boy who is kicking the girl is translating is dull.	Who is translating the dull novel?	boy	Who is being kicked?	girl
12	1	punishing	3,45	3	9	squashing	2,8	2	9	small	5,43	1	5	grape	3,54	1	5	girl	boy	she	The girl is punishing the boy and she is squashing a small grape.	Who is squashing the small grape?	girl	Who is being punished?	boy
12	2	punishing	3,45	3	9	squashing	2,8	2	9	small	5,43	1	5	grape	3,54	1	5	girl	boy	she	The girl who is punishing the boy is squashing a small grape.	Who is squashing the small grape?	girl	Who is being punished?	boy
12	3	punishing	3,45	3	9	squashing	2,8	2	9	small	5,43	1	5	grape	3,54	1	5	girl	boy	she	The boy is being punished by the girl who is squashing a small grape.	Who is squashing the small grape?	girl	Who is being punished?	boy
12	4	punishing	3,45	3	9	squashing	2,8	2	9	small	5,43	1	5	grape	3,54	1	5	girl	boy	she	The grape which the girl who is punishing the boy is squashing is small.	Who is squashing the small grape?	girl	Who is being punished?	boy
13	1	tickling	3,36	2	8	drawing	4,47	2	7	pretty	5,5	2	6	flower	4,5	2	6	boy	girl	he	The boy is tickling the girl and he is drawing a pretty flower.	Who is drawing the pretty flower?	boy	Who is being tickled?	girl
13	2	tickling	3,36	2	8	drawing	4,47	2	7	pretty	5,5	2	6	flower	4,5	2	6	boy	girl	he	The boy who is tickling the girl is drawing a pretty flower.	Who is drawing the pretty flower?	boy	Who is being tickled?	girl
13	3	tickling	3,36	2	8	drawing	4,47	2	7	pretty	5,5	2	6	flower	4,5	2	6	boy	girl	he	The girl is being tickled by the boy who is drawing a pretty flower.	Who is drawing the pretty flower?	boy	Who is being tickled?	girl

item	syntax level	AnTrans Verb	freq	sylls	letters	InanTrans Verb	freq	sylls	letters	Adjective	freq	sylls	letters	Inan Noun	freq	sylls	letters	Lead Noun	An Noun	Gender	Sentence	Question1	Answer1	Question2	Answer2
13	4	tickling	3,36	2	8	drawing	4,47	2	7	pretty	5,5	2	6	flower	4,5	2	6	boy	girl	he	The flower which the boy who is tickling the girl is drawing is pretty.	Who is drawing the pretty flower?	boy	Who is being tickled?	girl
14	1	shaking	4,21	2	7	snatching	2,93	2	9	pricy	2,23	2	5	necklace	3,8	2	8	girl	boy	she	The girl is shaking the boy and she is snatching a pricy necklace.	Who is snatching the pricy necklace?	girl	Who is being shaken?	boy
14	2	shaking	4,21	2	7	snatching	2,93	2	9	pricy	2,23	2	5	necklace	3,8	2	8	girl	boy	she	The girl who is shaking the boy is snatching a pricy necklace.	Who is snatching the pricy necklace?	girl	Who is being kissed?	boy
14	3	shaking	4,21	2	7	snatching	2,93	2	9	pricy	2,23	2	5	necklace	3,8	2	8	girl	boy	she	The boy is being shaken by the girl who is snatching a pricy necklace.	Who is snatching the pricy necklace?	girl	Who is being kissed?	boy
14	4	shaking	4,21	2	7	snatching	2,93	2	9	pricy	2,23	2	5	necklace	3,8	2	8	girl	boy	she	The necklace which the girl who is shaking the boy is snatching is pricy.	Who is snatching the pricy necklace?	girl	Who is being kissed?	boy
15	1	protecting	4,21	3	10	removing	3,9	3	8	heavy	4,9	2	5	armchair	3,31	2	8	boy	girl	he	The boy is protecting the girl and he is removing a heavy armchair.	Who is removing the heavy armchair?	boy	Who is being protected?	girl
15	2	protecting	4,21	3	10	removing	3,9	3	8	heavy	4,9	2	5	armchair	3,31	2	8	boy	girl	he	The boy who is protecting the girl is removing a heavy armchair.	Who is removing the heavy armchair?	boy	Who is being protected?	girl
15	3	protecting	4,21	3	10	removing	3,9	3	8	heavy	4,9	2	5	armchair	3,31	2	8	boy	girl	he	The girl is being protected by the boy who is removing a heavy armchair.	Who is removing the heavy armchair?	boy	Who is being protected?	girl
15	4	protecting	4,21	3	10	removing	3,9	3	8	heavy	4,9	2	5	armchair	3,31	2	8	boy	girl	he	The armchair which the boy who is protecting the girl is removing is heavy.	Who is removing the heavy armchair?	boy	Who is being protected?	girl
16	1	visiting	4,3	3	8	closing	4,35	2	7	creaky	2,69	2	6	gate	4,34	1	4	girl	boy	she	The girl is visiting the boy and she is closing a creaky gate.	Who is closing the creaky gate?	girl	Who is being visited?	boy
16	2	visiting	4,3	3	8	closing	4,35	2	7	creaky	2,69	2	6	gate	4,34	1	4	girl	boy	she	The girl who is visiting the boy is closing a creaky gate.	Who is closing the creaky gate?	girl	Who is being visited?	boy
16	3	visiting	4,3	3	8	closing	4,35	2	7	creaky	2,69	2	6	gate	4,34	1	4	girl	boy	she	The boy is being visited by the girl who is closing a creaky gate.	Who is closing the creaky gate?	girl	Who is being visited?	boy
16	4	visiting	4,3	3	8	closing	4,35	2	7	creaky	2,69	2	6	gate	4,34	1	4	girl	boy	she	The gate which the girl who is visiting the boy is closing is creaky.	Who is closing the creaky gate?	girl	Who is being visited?	boy
17	1	praising	3,13	2	8	washing	4,41	2	7	muddy	3,87	2	5	coat	4,41	1	4	boy	girl	he	The boy is praising the girl and he is washing a muddy coat.	Who is washing the muddy coat?	boy	Who is being praised?	girl
17	2	praising	3,13	2	8	washing	4,41	2	7	muddy	3,87	2	5	coat	4,41	1	4	boy	girl	he	The boy who is praising the girl is washing a muddy coat.	Who is washing the muddy coat?	boy	Who is being praised?	girl
17	3	praising	3,13	2	8	washing	4,41	2	7	muddy	3,87	2	5	coat	4,41	1	4	boy	girl	he	The girl is being praised by the boy who is washing a muddy coat.	Who is washing the muddy coat?	boy	Who is being praised?	girl
17	4	praising	3,13	2	8	washing	4,41	2	7	muddy	3,87	2	5	coat	4,41	1	4	boy	girl	he	The coat which the boy who is praising the girl is washing is muddy.	Who is washing the muddy coat?	boy	Who is being praised?	girl
18	1	confusing	3,98	3	9	hiding	4,44	2	6	big	5,94	1	3	bottle	4,65	2	6	girl	boy	she	The girl is confusing the boy and she is hiding a big bottle.	Who is hiding the big bottle?	girl	Who is being confused?	boy

item	syntax level	AnTrans Verb	freq	sylls	letters	InanTrans Verb	freq	sylls	letters	Adjective	freq	sylls	letters	Inan Noun	freq	sylls	letters	Lead Noun	An Noun	Gender	Sentence	Question1	Answer1	Question2	Answer2
18	2	confusing	3,98	3	9	hiding	4,44	2	6	big	5,94	1	3	bottle	4,65	2	6	girl	boy	she	The girl who is confusing the boy is hiding a big bottle.	Who is hiding the big bottle?	girl	Who is being confused?	boy
18	3	confusing	3,98	3	9	hiding	4,44	2	6	big	5,94	1	3	bottle	4,65	2	6	girl	boy	she	The boy is being confused by the girl who is hiding a big bottle.	Who is hiding the big bottle?	girl	Who is being confused?	boy
18	4	confusing	3,98	3	9	hiding	4,44	2	6	big	5,94	1	3	bottle	4,65	2	6	girl	boy	she	The bottle which the girl who is confusing the boy is hiding is big.	Who is hiding the big bottle?	girl	Who is being confused?	boy
19	1	pushing	4,58	2	7	repeating	3,58	3	9	nasty	4,45	2	5	rumour	3,75	2	6	boy	girl	he	The boy is pushing the girl and he is repeating a nasty rumour.	Who is repeating the nasty rumour?	boy	Who is being pushed?	girl
19	2	pushing	4,58	2	7	repeating	3,58	3	9	nasty	4,45	2	5	rumour	3,75	2	6	boy	girl	he	The boy who is pushing the girl is repeating a nasty rumour.	Who is repeating the nasty rumour?	boy	Who is being pushed?	girl
19	3	pushing	4,58	2	7	repeating	3,58	3	9	nasty	4,45	2	5	rumour	3,75	2	6	boy	girl	he	The girl is being pushed by the boy who is repeating a nasty rumour.	Who is repeating the nasty rumour?	boy	Who is being pushed?	girl
19	4	pushing	4,58	2	7	repeating	3,58	3	9	nasty	4,45	2	5	rumour	3,75	2	6	boy	girl	he	The rumour which the boy who is pushing the girl is repeating is nasty.	Who is repeating the nasty rumour?	boy	Who is being pushed?	girl
20	1	blaming	3,79	2	7	chewing	3,7	2	7	homemade	3,41	2	8	biscuit	4,1	2	7	girl	boy	she	The girl is blaming the boy and she is chewing a homemade biscuit.	Who is chewing the homemade biscuit?	girl	Who is being blamed?	boy
20	2	blaming	3,79	2	7	chewing	3,7	2	7	homemade	3,41	2	8	biscuit	4,1	2	7	girl	boy	she	The girl who is blaming the boy is chewing a homemade biscuit.	Who is chewing the homemade biscuit?	girl	Who is being blamed?	boy
20	3	blaming	3,79	2	7	chewing	3,7	2	7	homemade	3,41	2	8	biscuit	4,1	2	7	girl	boy	she	The boy is being blamed by the girl who is chewing a homemade biscuit.	Who is chewing the homemade biscuit?	girl	Who is being blamed?	boy
20	4	blaming	3,79	2	7	chewing	3,7	2	7	homemade	3,41	2	8	biscuit	4,1	2	7	girl	boy	she	The biscuit which the girl who is blaming the boy is chewing is homemade.	Who is chewing the homemade biscuit?	girl	Who is being blamed?	boy
21	1	pinching	3,22	2	8	greasing	2,29	2	8	large	5,1	1	5	tin	4,35	1	3	boy	girl	he	The boy is pinching the girl and he is greasing a large tin.	Who is greasing the large tin?	boy	Who is being pinched?	girl
21	2	pinching	3,22	2	8	greasing	2,29	2	8	large	5,1	1	5	tin	4,35	1	3	boy	girl	he	The boy who is pinching the girl is greasing a large tin.	Who is greasing the large tin?	boy	Who is being pinched?	girl
21	3	pinching	3,22	2	8	greasing	2,29	2	8	large	5,1	1	5	tin	4,35	1	3	boy	girl	he	The girl is being pinched by the boy who is greasing a large tin.	Who is greasing the large tin?	boy	Who is being pinched?	girl
21	4	pinching	3,22	2	8	greasing	2,29	2	8	large	5,1	1	5	tin	4,35	1	3	boy	girl	he	The tin which the boy who is pinching the girl is greasing is large.	Who is greasing the large tin?	boy	Who is being pinched?	girl
22	1	persuading	3,45	3	10	fitting	4,1	2	7	fragile	3,99	2	7	lightbulb	2,68	2	9	girl	boy	she	The girl is persuading the boy and she is fitting a fragile lightbulb.	Who is fitting the fragile lightbulb?	girl	Who is being persuaded?	boy
22	2	persuading	3,45	3	10	fitting	4,1	2	7	fragile	3,99	2	7	lightbulb	2,68	2	9	girl	boy	she	The girl who is persuading the boy is fitting a fragile lightbulb.	Who is fitting the fragile lightbulb?	girl	Who is being persuaded?	boy
22	3	persuading	3,45	3	10	fitting	4,1	2	7	fragile	3,99	2	7	lightbulb	2,68	2	9	girl	boy	she	The boy is being persuaded by the girl who is fitting a fragile lightbulb.	Who is fitting the fragile lightbulb?	girl	Who is being persuaded?	boy

item	syntax level	AnTrans Verb	freq	sylls	letters	InanTrans Verb	freq	sylls	letters	Adjective	freq	sylls	letters	Inan Noun	freq	sylls	letters	Lead Noun	An Noun	Gender	Sentence	Question1	Answer1	Question2	Answer2
22	4	persuading	3,45	3	10	fitting	4,1	2	7	fragile	3,99	2	7	light bulb	2,68	2	9	girl	boy	she	The lightbulb which the girl who is persuading the boy is fitting is fragile.	Who is fitting the fragile lightbulb?	girl	Who is being persuaded?	boy
23	1	vexing	2,04	2	6	watering	3,84	3	8	fragrant	3,35	2	8	herb	3,76	1	4	boy	girl	he	The boy is vexing the girl and he is watering a fragrant herb.	Who is watering the fragrant herb?	boy	Who is being vexed?	girl
23	2	vexing	2,04	2	6	watering	3,84	3	8	fragrant	3,35	2	8	herb	3,76	1	4	boy	girl	he	The boy who is vexing the girl is watering a fragrant herb.	Who is watering the fragrant herb?	boy	Who is being vexed?	girl
23	3	vexing	2,04	2	6	watering	3,84	3	8	fragrant	3,35	2	8	herb	3,76	1	4	boy	girl	he	The girl is being vexed by the boy who is watering a fragrant herb.	Who is watering the fragrant herb?	boy	Who is being vexed?	girl
23	4	vexing	2,04	2	6	watering	3,84	3	8	fragrant	3,35	2	8	herb	3,76	1	4	boy	girl	he	The herb which the boy who is vexing the girl is watering is fragrant.	Who is watering the fragrant herb?	boy	Who is being vexed?	girl
24	1	worrying	4,41	3	8	investigating	4,24	2	13	strange	4,83	1	7	device	4,27	2	6	girl	boy	she	The girl is worrying the boy and she is investigating a strange device.	Who is investigating the strange device?	girl	Who is being worried?	boy
24	2	worrying	4,41	3	8	investigating	4,24	2	13	strange	4,83	1	7	device	4,27	2	6	girl	boy	she	The girl who is worrying the boy is investigating a strange device.	Who is investigating the strange device?	girl	Who is being worried?	boy
24	3	worrying	4,41	3	8	investigating	4,24	2	13	strange	4,83	1	7	device	4,27	2	6	girl	boy	she	The boy is being worried by the girl who is investigating a strange device.	Who is investigating the strange device?	girl	Who is being worried?	boy
24	4	worrying	4,41	3	8	investigating	4,24	2	13	strange	4,83	1	7	device	4,27	2	6	girl	boy	she	The device which the girl who is worrying the boy is investigating is strange.	Who is investigating the strange device?	girl	Who is being worried?	boy

Appendix G: Text Comprehension Task Information on Monitor

Introduction information first page	Introduction information second page
Welcome. First, you will see a sentence appear on the screen. Please read the sentence carefully. Once you have finished reading, press the SPACE key. Then you will be asked a follow up question. Please respond to this question based on the sentence you have just read. It is important to be as fast and as accurate as you possibly can. Press SPACE to continue.	For every question, the answer will be either GIRL or BOY; using the keyboard, press the Z key if your answer is GIRL, or M key if your answer is BOY. Remember: Z= GIRL M = BOY When you have finished reading, press SPACE to proceed- You will now do a few practice trials. Press SPACE to begin.
Trial and actual task rounds sentence screen	**Trial and actual task round question screen**
The girl is teasing the boy and she is colouring a silly picture.	Who is colouring the silly picture? Z= GIRL M= BOY

Appendix H: Free Picture Description Task

Figure 15: Free picture description pictures

Table 20: Information on monitor during free picture description task

First information page	Experiment countdown page
Welcome to the experiment. You are about to see a pictured scene. Please describe what you see in the picture in as much detail as you can. You can describe the surroundings, people, animals, and objects, as well as any actions or events that you observe. Please speak clearly and loudly. This should not take you more than 2 minutes. Press SPACE to begin.	The picture will appear in: 5.0

Appendix I: Free Picture Description Images Matching List

"Cookie theft"	**animates**	**verbs**	**objects**		**"Party"**	**animates**	**verbs**	**objects**		**"Garden"**	**animates**	**verbs**	**objects**		**"Dinner"**	**animates**	**verbs**	**objects**
	man	washing	stool			dancing woman 1	vomiting	speaker			boy	jumping	book/kindle			woman	falling	fish bowl
	woman	mowing	dishes			dancing woman 2	dancing	sofa			girl	falling	trampoline			boy	watching	chair
	girl	falling	lawnmower			bored woman	filming	bottles			young boy	defecating	nest			girl	eating	table
	boy	laughing	tin biscuits			filming man	watching	cups			dog	crawling	poop			cats	tripping	baby seat
	dog	stealing	kitchen			pointing man	pointing	plant pot			cat	sleeping	fence			fish	laughing	glass
	cat	flying	cupboards			vomiting woman	sitting	plant			bird	climbing	tree			dog	charging	knife
	birds	eating	lawnmower			cat	drinking	phone			man	sitting	branch			cat	sitting	fork
		chatting	phone				smiling	window				reaching	plant pot				walking	ipad
		chasing	tap				sleeping	curtains				hunting	garden				smiling	charging cord
			sink					table					bush					dining room
			drawers					living room					frisbee					curtains
			garden					glass					slippers					plate
			water					picture frame					rattle					window
			cookies					clock					ball					games console/tablet
			fence					cat basket					cup/drink					spoon

Appendix J: Leap-QA Data

Table 21: Exposure in English

sub	Overall exposure	Overall speaking	Overall Reading	Exposure friends	Exposure family	Exposure reading	Exposure school	Exposure selfinstruct	Exposure media	Exposure gaming
1	40%	15%	85%	2	0	1	9	0	8	10
2	10%	1%	30%	2	0	4	2	0	5	0
3	40%	20%	80%	5	0	10	8	6	7	0
4	40%	10%	70%	5	1	7	7	10	9	8
5	30%	15%	30%	1	0	2	1	0	3	0
6	55%	45%	43%	8	1	8	9	0	9	4
7	30%	20%	50%	2	0	5	0	0	5	0
8	40%	10%	10%	5	1	7	3	1	6	1
9	70%	20%	90%	3	0	9	5	0	10	10
10	40%	30%	70%	8	0	10	10	0	10	10
11	50%	10%	50%	2	0	0	10	0	10	10
12	50%	20%	80%	5	0	10	10	0	10	0
13	40%	20%	30%	2	0	2	3	7	8	6
14	50%	20%	40%	2	0	5	5	0	5	0
15	50%	20%	40%	7	0	4	3	0	10	8
16	20%	20%	10%	0	0	2	5	0	6	0
17	50%	30%	50%	5	0	5	5	0	7	10
18	5%	5%	20%	0	0	0	5	0	8	0

Table 22: Learning in English

sub	Friends	Family	Reading	School	Selfinstruction	Media	Gaming
1	8	1	8	7	1	10	5
2	2	8	9	9	7	10	0
3	8	0	8	10	6	8	0
4	7	1	8	7	6	9	8
5	5	1	9	10	1	8	5
6	4	2	8	9	1	8	3
7	2	0	8	10	4	10	5
8	3	3	10	10	1	7	1
9	8	3	10	7	0	10	8
10	2	0	5	10	0	10	10
11	4	0	0	5	0	10	10
12	1	0	10	10	0	10	0
13	8	5	7	10	3	10	8
14	5	0	10	10	0	8	0
15	8	0	10	10	6	7	5
16	1	0	5	10	0	10	0
17	5	2	0	7	0	10	10
18	0	0	0	10	0	7	0

Table 23: English Proficiency Self-Ratings

sub	Friends	Family	Reading	School	Selfinstruction	Media	Gaming	Speaking	Accent	Listening	Reading	Writing	Vocab
1	2	0	1	9	0	8	10	7	8	9	8	7	7
2	2	0	4	2	0	5	0	7	7	8	8	8	8
3	5	0	10	8	6	7	0	7	6	8	8	5	5
4	5	1	7	7	10	9	8	9	7	9	8	9	9
5	1	0	2	1	0	3	0	7	7	7	7	7	6
6	8	1	8	9	0	9	4	8	7	9	9	8	8
7	2	0	5	0	0	5	0	6	6	6	5	5	4
8	5	1	7	3	1	6	1	5	6	8	8	9	7
9	3	0	9	5	0	10	10	9	8	9	9	9	8
10	8	0	10	10	0	10	10	7	7	7	7	8	8
11	2	0	0	10	0	10	10	7	7	7	8	8	7
12	5	0	10	10	0	10	0	8	8	10	10	10	10
13	2	0	2	3	7	8	6	8	2	9	8	7	7
14	2	0	5	5	0	5	0	5	5	5	5	5	5
15	7	0	4	3	0	10	8	8	6	9	8	8	8
16	0	0	2	5	0	6	0	7	7	7	7	7	7
17	5	0	5	5	0	7	10	6	6	8	7	7	7
18	0	0	0	5	0	8	0	7	7	7	7	7	6

Table 24: Self-reported Language Switching

sub	Switching freq	Prof switching	ENG accidentally in NOR	NOR accidentally in ENG	ENG intentionally in NOR	NOR intentionally in ENG
1	3	9	2	2	6	3
2	1	7	0	0	2	2
3	2	8	4	1	0	0
4	2	8	0	0	2	1
5	3	6	0	0	2	0
6	2	8	0	0	0	0
7	1	5	2	8	3	3
8	2	3	3	2	4	2
9	2	7	7	2	4	1
10	3	9	2	6	2	6
11	2	5	2	2	6	1
12	3	8	6	2	6	1
13	0	8	4	0	0	0
14	3	5	3	6	3	3
15	2	8	6	5	7	4
16	0	5	0	0	0	0
17	3	8	2	2	5	1
18	0	5	0	0	0	0

Appendix K: Objective Language Proficiency Results

Subject	Vocab percent correct	Comp percent correct	Comp correct reading time
1	60,7	100,0	8177
2	50,0	87,5	6658
3	50,0	100,0	7575
4	75,0	87,5	3224
5	60,0	95,8	6165
6	78,6	100,0	3561
7	35,7	79,2	6082
8	42,9	95,8	5010
9	57,1	87,5	4425
10	39,3	83,3	2981
11	35,7	91,7	3933
12	32,1	100,0	3970
13	25,0	87,5	5392
14	63,0	100,0	9082
15	67,9	95,8	4056
16	28,6	91,7	5950
17	46,4	100,0	4767
18	29,6	95,8	4667

Appendix L: Correlations of Objective Proficiency, Self-ratings, and Mean Proficiency

Correlation	z value	z value
Correct vocabulary percent and		
vocab proficiency	0.86401	0.38
mean proficiency	0.83909	0.40
Text Comprehension mean duration and reading proficiency	**-1.4944**	**0.13**
Correct comprehension percent and		
reading proficiency	0.70885	0.47
mean proficiency	-0.51228	0.60

Appendix M: Linear Regression of Objective measures and Learning and Exposure Variables Full Outputs

All significance across all output data is marked.

Vocabulary percent correct by learning

		Estimate	Std. error	t value	Pr(>|t|)
Vocabulary percent correct by learning	Friends	3.449e-02	1.664e+00	0.021	0.9839
	Family	-6.286e-01	1.600e+00	-0.393	0.7026
	Reading	2.837e+00	1.230e+00	2.307	0.0437
	School	-6.283e+00	2.805e+00	-2.240	0.0490
	Self-instruction	7.804e-01	1.423e+00	0.549	0.5953
	Media	-7.083e+00	3.177e+00	-2.230	0.0499
	Gaming	6.424e-01	1.184e+00	0.543	0.5993

Vocabulary percent correct by exposure

		Estimate	Std. error	t value	Pr(>|t|)
Vocabulary percent correct by exposure	Friends	2.268e+00	3.113e+00	0.729	0.483
	Family	1.545e+01	1.324e+01	1.167	0.270
	Reading	-2.507e-03	1.932e+00	-0.001	0.999
	School	2.851e-01	1.757e+00	0.162	0.874
	Self-instruction	-1.791e-01	1.504e+00	-0.119	0.908
	Media	-3.491e+00	3.269e+00	-1.068	0.311
	Gaming	1.377e+00	1.315e+00	1.048	0.319

Text comprehension percent correct by learning

		Estimate	Std. error	t value	Pr(>|t|)
Text comprehension percent correct by learning	Friends	1.439e+00	6.517e-01	2.208	0.0517
	Family	-1.366e-01	6.263e-01	-0.218	0.8317
	Reading	-4.667e-01	4.816e-01	-0.969	0.3554
	School	-9.519e-01	1.099e+00	-0.866	0.4065
	Self-instruction	-1.161e+00	5.571e-01	-2.084	0.0637
	Media	-1.373e+00	1.244e+00	-1.103	0.2957
	Gaming	-1.343e+00	4.636e-01	-2.897	0.0159

Text comprehension percent correct by exposure

		Estimate	Std. error	t value	Pr(>|t|)
Text comprehension percent correct by exposure	Friends	1.305e+00	1.289e+00	1.013	0.335
	Family	-6.728e-02	5.480e+00	-0.012	0.990
	Reading	-8.668e-01	7.997e-01	-1.084	0.304
	School	1.310e+00	7.274e-01	1.801	0.102
	Self-instruction	-2.282e-01	6.226e-01	-0.367	0.722

	Media	-1.055e+00	1.354e+00	-0.780	0.454
	Gaming	-4.914e-01	5.443e-01	-0.903	0.388

Text comprehension mean reading duration by learning

		Estimate	Std. error	t value	Pr(>|t|)
Text comprehension mean reading duration by learning	Friends	4.642e+02	2.008e+02	2.312	0.0434
	Family	3.731e+00	1.930e+02	0.019	0.9850
	Reading	-1.183e+02	1.484e+02	-0.797	0.4438
	School	1.706e+02	3.385e+02	0.504	0.6251
	Self-instruction	-9.593e+01	1.716e+02	-0.559	0.5885
	Media	5.834e+02	3.832e+02	1.522	0.1589
	Gaming	-4.207e+02	1.428e+02	-2.946	0.0146

Text comprehension mean reading duration by exposure

		Estimate	Std. error	t value	Pr(>|t|)
Text comp mean reading duration by learning	Friends	-9.456e+01	2.569e+02	-0.368	0.7204
	Family	-1.454e+03	1.092e+03	-1.331	0.2127
	Reading	2.387e+01	1.594e+02	0.150	0.8839
	School	1.975e+02	1.450e+02	1.363	0.2029
	Self-instruction	6.721e+01	1.241e+02	0.542	0.5999
	Media	-6.903e+02	2.698e+02	-2.559	0.0284
	Gaming	1.565e+01	1.085e+02	0.144	0.8881

Appendix N: Free Speech Description data

subject	utterance_no	utterance_tot	utterance_mean	pause_out_no	pause_out_tot	pause_out_mean	pause_in_no	pause_in_tot	pause_in_mean
1	13	73433	5649	6	8164	1361	14	10478	748
2	6	42318	7053	5	8164	1633	15	9848	657
3	24	87633	3651	7	7051	1007	2	1804	902
4	16	83188	5199	13	15618	1201	20	12216	611
5	15	79632	5309	14	13263	947	8	5733	717
6	12	95889	7991	9	12540	1393	20	17422	871
7	12	83386	6949	9	16597	1844	6	3663	610
8	12	50877	4240	11	22617	2056	10	7434	743
9	14	88136	6295	7	4493	642	5	5582	1116
10	25	108188	4328	18	28358	1575	19	11060	582
11	13	56125	4317	12	26934	2244	6	6792	1132
12	17	69980	4116	9	8008	890	18	9816	545
13	17	52722	3101	15	23787	1586	13	7624	586
14	11	64979	5907	10	13401	1340	25	18181	727
15	10	43718	4372	6	9123	1521	9	5575	619
16	8	37295	4662	6	9629	1605	10	6808	681
17	14	59203	4229	7	7181	1026	18	11785	655
18	9	46059	5118	8	11608	1451	6	4335	722

subject	sylls_tot	sylls_mean	contwords_tot	contwords_mean	morphs_tot	morphs_mean	cognates_tot	cognates_mean	noncogs_tot	noncogs_mean
1	202	15,5	75	5,8	185	14,2	31	2,4	9	0,7
2	93	15,5	27	4,5	74	12,3	5	0,8	6	1,0
3	254	10,6	84	3,5	249	10,4	37	1,5	47	2,0
4	304	19,0	90	5,6	286	17,9	49	3,1	40	2,5
5	211	14,1	67	4,5	190	12,7	33	2,2	34	2,3
6	207	17,3	68	5,7	188	15,7	15	1,3	53	4,4
7	205	17,1	71	5,9	193	16,1	26	2,2	45	3,8
8	148	12,3	50	4,2	140	11,7	23	1,9	27	2,3
9	277	19,8	86	6,1	250	17,9	40	2,9	46	3,3
10	370	14,8	98	3,9	359	14,4	48	1,9	50	2,0
11	134	10,3	51	3,9	124	9,5	25	1,9	26	2,0
12	237	13,9	75	4,4	230	13,5	36	2,1	40	2,4
13	200	11,8	62	3,6	193	11,4	35	2,1	75	4,4
14	166	15,1	49	4,5	149	13,5	36	3,3	13	1,2
15	113	11,3	45	4,5	105	10,5	31	3,1	14	1,4
16	94	11,8	38	4,8	91	11,4	25	3,1	13	1,6
17	148	10,6	47	3,4	130	9,3	27	1,9	20	1,4
18	142	15,8	47	5,2	130	14,4	27	3,0	20	2,2

Appendix O: Free Speech Description Key measures data

sub	totaltime	Mean_Duration	MLU_morphs_mean	speechrate	articulationrate	phonationtime	mean_sylls_run	pause_out_mean	pause_in_mean
1	92075	7083	14,23	0,0022	0,0028	0,80	15,54	1360,61	748,42
2	60330	10055	12,33	0,0015	0,0022	0,70	15,50	1632,76	656,56
3	96488	4020	10,38	0,0026	0,0029	0,91	10,58	1007,25	901,75
4	111022	6939	17,88	0,0027	0,0037	0,75	19,00	1201,39	610,80
5	98628	6575	12,67	0,0021	0,0026	0,81	14,07	947,35	716,57
6	125851	10488	15,67	0,0016	0,0022	0,76	17,25	1393,32	871,10
7	103646	8637	16,08	0,0020	0,0025	0,80	17,08	1844,10	610,46
8	80928	6744	11,67	0,0018	0,0029	0,63	12,33	2056,08	743,44
9	98211	7015	17,86	0,0028	0,0031	0,90	19,79	641,86	1116,41
10	147606	5904	14,36	0,0025	0,0034	0,73	14,80	1575,45	582,08
11	89851	6912	9,54	0,0015	0,0024	0,62	10,31	2244,50	1131,98
12	87804	5165	13,53	0,0027	0,0034	0,80	13,94	889,78	545,35
13	84133	4949	11,35	0,0024	0,0038	0,63	11,76	1585,77	586,43
14	96561	8778	13,55	0,0017	0,0026	0,67	15,09	1340,14	727,24
15	58416	5842	10,50	0,0019	0,0026	0,75	11,30	1520,56	619,48
16	53732	6717	11,38	0,0017	0,0025	0,69	11,75	1604,84	680,82
17	78169	5584	9,29	0,0019	0,0025	0,76	10,57	1025,79	654,72
18	62002	6889	14,44	0,0023	0,0031	0,74	15,78	1451,01	722,43

Appendix P: Linear Regression of Key Speech Data Measures and Learning and Exposure Variables Full Outputs

All significance across all output data is marked.

Mean length of utterance by learning

		Estimate	Std. error	t value	**Pr(>|t|)**
Mean length of utterance by learning	Friends	-4.362e-01	3.721e-01	-1.172	0.2683
	Family	-1.739e-01	3.577e-01	-0.486	0.6373
	Reading	5.645e-01	2.750e-01	2.053	0.0672 .
	School	-5.437e-01	6.273e-01	-0.867	0.4064
	Self-instruction	-1.510e-02	3.181e-01	-0.047	0.9631
	Media	-1.273e-01	7.103e-01	-0.179	0.8613
	Gaming	2.184e-01	2.648e-01	0.825	0.4287

Mean length of utterance by exposure

		Estimate	Std. error	t value	**Pr(>|t|)**
Mean length of utterance by exposure	Friends	-1.033e+00	4.671e-01	-2.211	0.0515
	Family	3.913e+00	1.986e+00	1.971	0.0771
	Reading	6.551e-01	2.898e-01	2.260	0.0473
	School	-1.903e-01	2.636e-01 -	-0.722	0.4869
	Self-instruction	-1.481e-01	-2.256e-01	-0.657	0.5262
	Media	4.513e-01	4.905e-01	0.920	0.3793
	Gaming	1.574e-01	1.972e-01	0.798	0.4433

Articulation rate by learning

		Estimate	Std. error	t value	**Pr(>|t|)**
Articulation rate by learning	Friends	4.421e-05	7.295e-05	0.606	0.558
	Family	-8.719e-06	7.012e-05	-0.124	0.904
	Reading	1.942e-06	5.391e-05	0.036	0.972
	School	1.480e-04	1.230e-04	1.204	0.256
	Self-instruction	-1.639e-05	6.236e-05	-0.263	0.798
	Media	6.681e-05	1.392e-04	0.480	0.642
	Gaming	4.467e-05	5.190e-05	0.861	0.410

Articulation rate by exposure

		Estimate	Std. error	t value	**Pr(>|t|)**
Articulation rate by exposure	Friends	-7.893e-05	7.408e-05	-1.065	0.3118
	Family	-1.403e-04	3.150e-04	-0.446	0.6654
	Reading	5.619e-05	4.597e-05	1.222	0.2496
	School	-2.321e-05	4.181e-05	-0.555	0.5910

	Self-instruction	9.328e-05	3.579e-05	2.606	0.0262
	Media	1.203e-04	7.781e-05	1.545	0.1533
	Gaming	1.873e-06	3.129e-05	0.060	0.9534

Mean utterance duration by learning

		Estimate	Std. error	t value	**Pr(>|t|)**
Mean utterance duration by learning	Friends	-2.670e+02	1.789e+02	1.492	0.166
	Family	8.070e+01	1.720e+02	0.469	0.649
	Reading	2.143e+02	1.322e+02	1.620	0.136
	School	-4.765e+02	3.016e+02	-1.580	0.145
	Self-instruction	2.934e+01	1.529e+02	0.192	0.852
	Media	-1.805e+02	3.415e+02	-0.528	0.609
	Gaming	-5.973e+00	1.273e+02	-0.047	0.963

Mean utterance duration by exposure

		Estimate	Std. error	t value	Pr(>|t|)
Mean utterance duration by exposure	Friends	-2.139e+02	2.477e+02	-0.864	0.4080
	Family	1.837e+03	1.053e+03	1.744	0.1117
	Reading	1.265e+02	1.537e+02	0.823	0.4295
	School	-4.962e+01	1.398e+02	-0.355	0.7299
	Self-instruction	-2.389e+02	1.196e+02	-1.997	0.0737
	Media	-6.870e+01	2.601e+02	-0.264	0.7971
	Gaming	4.619e+01	1.046e+02	0.442	0.6682

Total time by learning

		Estimate	Std. error	t value	Pr(>|t|)
Total time by learning	Friends	-2.334e+03	3.081e+03	-0.758	0.4662
	Family	-2.849e+03	2.961e+03	-0.962	0.3586
	Reading	3.725e+03	2.277e+03	1.636	0.1328
	School	5.906e+02	5.194e+03	0.114	0.9117
	Self-instruction	-9.720e+02	2.634e+03	-0.369	0.7198
	Media	-6.315e+02	5.881e+03	-0.107	0.9166
	Gaming	4.532e+03	2.192e+03	2.067	0.0656

Total time by exposure

		Estimate	Std. error	t value	Pr(>|t|)
Total time by exposure	Friends	4.834e+02	3.946e+03	0.123	0.905
	Family	7.518e+03	1.678e+04	0.448	0.664
	Reading	3.409e+03	2.448e+03	1.393	0.194

	School	3.252e+03	2.227e+03	1.460	0.175
	Self-instruction	1.989e+02	1.906e+03	0.104	0.919
	Media	-6.100e+03	4.144e+03	-1.472	0.172
	Gaming	2.450e+03	1.666e+03	1.470	0.172

Speech rate by learning

		Estimate	Std. error	t value	Pr(>|t|)
Speech rate by learning	Friends	5.334e-05	6.319e-05	0.844	0.418
	Family	-5.558e-05	6.073e-05	-0.915	0.382
	Reading	1.712e-05	4.669e-05	0.367	0.722
	School	8.307e-05	1.065e-04	0.780	0.454
	Self-instruction	-6.389e-07	5.401e-05	-0.012	0.991
	Media	9.772e-05	1.206e-04	0.810	0.437
	Gaming	1.568e-05	4.496e-05	0.349	0.735

Speech rate by exposure

		Estimate	Std. error	t value	Pr(>|t|)
Speech rate by exposure	Friends	-9.698e-05	5.886e-05	-1.648	0.1304
	Family	-2.610e-04	2.503e-04	-1.043	0.3215
	Reading	1.009e-04	3.652e-05	2.764	0.0200
	School	-1.484e-05	3.322e-05	-0.447	0.6647
	Self-instruction	6.330e-05	2.843e-05	2.226	0.0502
	Media	8.441e-05	6.182e-05	1.365	0.2020
	Gaming	1.038e-05	2.486e-05	0.418	0.6850

Phonation time by learning

		Estimate	Std. error	t value	Pr(>|t|)
Phonation time by learning	Friends	9.429e-03	1.197e-02	0.788	0.449
	Family	-1.536e-02	1.151e-02	-1.335	0.211
	Reading	4.098e-03	8.847e-03	0.463	0.653
	School	-1.452e-03	2.018e-02	-0.072	0.944
	Self-instruction	3.806e-03	1.023e-02	0.372	0.718
	Media	1.770e-02	2.285e-02	0.775	0.457
	Gaming	-4.871e-03	8.517e-03	-0.572	-0.580

Phonation time by exposure

		Estimate	Std. error	t value	Pr(>|t|)
Phonation time by exposure	Friends	-1.015e-02	1.592e-02	-0.638	0.5379
	Family	-6.122e-02	6.768e-02	-0.905	0.3870
	Reading	1.884e-02	9.876e-03	1.907	0.0856
	School	-5.140e-04	8.983e-03	-0.057	0.9555

	Self-instruction	1.883e-05	7.689e-03	0.002	0.9981
	Media	-2.220e-03	1.672e-02	-0.133	0.8970
	Gaming	2.845e-03	6.722e-03	0.423	0.6811

Mean pause outside by learning

		Estimate	Std. error	t value	Pr(>|t|)
Mean pause out by learning	Friends	-8.343e+01	6.033e+01	-1.383	0.197
	Family	1.690e+01	5.798e+01	0.291	0.777
	Reading	-1.250e+01	4.458e+01	-0.280	0.785
	School	-3.393e+01	1.017e+02	-0.334	0.746
	Self-instruction	4.415e+01	5.156e+01	0.856	0.412
	Media	-8.092e+01	1.151e+02	-0.703	0.498
	Gaming	2.831e+01	4.292e+01	0.660	0.524

Mean Pause outside by exposure

		Estimate	Std. error	t value	Pr(>|t|)
Mean pause out by exposure	Friends	6.460e+01	8.070e+01	0.800	0.4420
	Family	2.774e+02	3.432e+02	0.808	0.4377
	Reading	-9.453e+01	5.008e+01	-1.888	0.0884
	School	-4.208e+00	4.555e+01	-0.092	0.9282
	Self-instruction	-1.559e+01	3.899e+01	-0.400	0.6976
	Media	2.312e+01	8.476e+01	0.273	0.7906
	Gaming	-2.107e+01	3.408e+01	-0.618	0.5502

Mean pause inside by learning

		Estimate	Std. error	t value	Pr(>|t|)
Mean pause in by learning	Friends	1.490e+01	2.018e+01	0.738	0.4774
	Family	3.712e+00	1.940e+01	0.191	0.8521
	Reading	-1.484e-02	1.492e+01	-0.001	0.9992
	School	-7.860e+01	3.402e+01	-2.310	0.0435
	Self-instruction	-2.027e+01	1.725e+01	-1.175	0.2672
	Media	-3.113e+01	3.852e+01	-0.808	0.4378
	Gaming	-1.385e+01	1.436e+01	-0.964	0.3576

Mean pause inside by exposure

		Estimate	Std. error	t value	Pr(>|t|)
Mean pause in by learning	Friends	-5.471e+01	3.408e+01	-1.606	0.139
	Family	1.461e+02	1.449e+02	1.008	0.337
	Reading	2.008e+01	2.114e+01	0.950	0.365
	School	1.199e+01	1.923e+01	0.624	0.547

	Self-instruction	-1.724e+01	1.646e+01	-1.047	0.320
	Media	7.409e+00	3.579e+01	0.207	0.840
	Gaming	1.583e+01	1.439e+01	1.100	0.297

Mean syllables per run by learning

		Estimate	Std. error	t value	Pr(>|t|)
Mean syllables per run by learning	Friends	-5.263e-01	4.109e-01	-1.281	0.229
	Family	1.990e-02	3.949e-01	0.050	0.961
	Reading	6.158e-01	3.036e-01	2.028	0.070
	School	-8.251e-01	6.926e-01	-1.191	0.261
	Self-instruction	-4.816e-03	3.512e-01	-0.014	0.989
	Media	-2.692e-01	7.842e-01	-0.343	0.739
	Gaming	1.908e-01	2.923e-01	0.653	0.529

Mean syllables per run by exposure

		Estimate	Std. error	t value	Pr(>|t|)
Mean syllables per run by exposure	Friends	-1.156e+00	5.349e-01	-2.162	0.0559
	Family	4.481e+00	2.274e+00	1.970	0.0771
	Reading	7.071e-01	3.319e-01	2.130	0.0590
	School	-2.412e-01	3.019e-01	-0.799	0.4429
	Self-instruction	-2.432e-01	2.584e-01	-0.941	0.3687
	Media	3.882e-01	5.618e-01	0.691	0.5053
	Gaming	2.140e-01	2.259e-01	0.947	0.3659

Mean cognate by learning

		Estimate	Std. error	t value	Pr(>|t|)
Mean cognate by learning	Friends	6.235e-02	9.700e-02	0.643	0.535
	Family	-1.543e-01	9.323e-02	-1.655	0.129
	Reading	1.654e-02	7.168e-02	0.231	0.822
	School	6.134e-03	1.635e-01	0.038	0.971
	Self-instruction	-5.095e-02	8.292e-02	-0.614	0.553
	Media	-4.128e-02	1.851e-01	-0.223	0.828
	Gaming	-7.011e-03	6.901e-02	-0.102	0.921

Mean cognate by exposure

		Estimate	Std. error	t value	Pr(>|t|)
Mean cognate by exposure	Friends	-1.620e-01	1.489e-01	-1.087	0.302
	Family	1.823e-01	6.333e-01	0.288	0.779
	Reading	1.943e-02	9.242e-02	0.210	0.838
	School	-6.672e-02	8.406e-02	-0.794	0.446

	Self-instruction	5.435e-03	7.195e-02	0.076	0.941
	Media	1.737e-01	1.564e-01	1.110	0.293
	Gaming	1.157e-02	6.290e-02	0.184	0.858