The impact of Filipina domestic workers on Hong Kong primary school children’s L2 English spoken CAF and reading accuracy and fluency

Abstract
This study investigates the impact of Filipina domestic workers (FilDWs), a marginalised group in Hong Kong (HK), on HK children’s language development. It focuses on FilDWs’ influence on the second language (L2) English of bilingual HK primary school children attending an English Medium of Instruction school. The elements investigated are: L2 English spoken complexity, accuracy and fluency (CAF), and reading accuracy and fluency. Participants comprise thirty-four children (17 boys and 17 girls, mean age 8;11) from homes with FilDWs and 30 (15 boys and 15 girls, mean age 8;11) from homes with no FilDW. Participants completed an English reading and speaking task, and an English working memory capacity (WMC) test. Participants from households with FilDWs scored significantly higher on all aspects of both English language measures, while no significant differences for WMC test was observed. These suggest that FilDWs exert a positive impact on children’s L2 English proficiency, placing them in a different position to the low status they are usually ascribed. These findings have implications for decolonising and decentring language learning and teaching.

Keywords: L2 English acquisition, language proficiency, CAF, Filipina domestic workers, home language environment, working memory capacity

Background

The impact of caregivers on children’s language acquisition

Caregivers play a major role in children’s language acquisition. Previous research comparing the language use of L1 Mandarin, L1 Italian and L1 English children under the age of 3 found that their language patterns mirror that of their primary caregivers (Tardif et al., 1997). The pattern that children emulate the type of language
they hear from caregivers, at least during the first three years of life, also applies to L2 (Deuchar & Quay 1999).

Other studies showed that the amount parents and caregivers talk has a direct effect on children, resulting in children from households with more verbose parents/caregivers becoming more verbose themselves (De Houwer, 2011; Hart & Risley, 1995). This concurs with earlier research, showing that children from households where parents frequently use the target language outperformed those whose parents seldom or never use the target language to interact with their children (Krashen, 1976).

In cases where bilingual mothers are the primary caregivers, a study involving 58 bilingual Spanish-English speaking children between the ages of 5 and 7, shows that higher maternal SES and more diverse maternal vocabulary significantly contribute to children’s L2 English (but not L1 Spanish) vocabulary development (Buac et al., 2014). The researchers suggested that L1 vocabulary acquisition of young bilingual school aged children is not affected by the language input received from primary caregivers, but that their L2 vocabulary acquisition is. One recent research investigating the role of caregivers on normal bilingual children’s language proficiency in the Netherlands (Sczepurek et al. 2022) involved 72 bilingual children (acquiring Dutch and at least one other language) with a mean age of 35 months. Results showed that the amount of Dutch (the majority language) input by caregivers was a significant predictor of children’s L1 vocabulary knowledge. Although seemingly contradicting Buac et al. (2014), participants were of different ages. Perhaps primary caregivers have a significant impact on the L1 vocabulary of children under three, while the same does not hold true for children above five, once formal schooling begins.

It is clear that input received from primary caregivers affects the language output of children (see Zhang et al. 2023 for a meta-analysis). Though it should be noted that many previous studies have not considered the diversity of the target language and input, making the implicit assumption that the target of acquisition is
“monolithic”, representing only one variety. Less is known in contexts where the target language is represented by various varieties. One of the objectives of the current study is to establish to what extent previous findings hold true in the case of HK primary school children in the care of English-speaking Filipina domestic workers living with and taking care of them at home, while both parents are working. We explore whether the Filipino variety, which is not considered “mainstream”, has an impact on children’s development of L2 English. Below, we provide background about foreign domestic workers (FDWs) in HK. Next, we outline the local linguistic environment before discussing the constructs of complexity, accuracy and fluency. We then briefly review studies around the influence of FDWs on children’s L2 acquisition, followed by details about the present study. We conclude by calling for more research that investigate the influence of “non-traditional” input on SLA.

Foreign domestic workers in Hong Kong

HK families have been employing foreign domestic workers (FDWs) since the 1970s, as there is a lack of local domestic workers in the region (Tang & Chor, 2012). The vast majority of the 340,000 FDWs (Cheung, 2022) in HK originate from the Philippines and Indonesia, with Filipinas accounting for more than half of the total FDW population in HK (Cheung et al., 2019). FDWs are officially employed to do domestic duties. However, apart from the usual domestic duties, many also take care of their employers’ children (Ma et al., 2020). Through employing FDWs, parents in many households are freed to take up full time employment. FDWs therefore contribute to the quality of HK families’ lives, something which they rarely get credit for. In fact, they are a marginalised group.

As has been documented by Ladegaard (2020), FDWs in HK have been subjected to numerous forms of mistreatment, including verbal, physical and sexual abuse. Such mistreatment may spring from prejudiced views that FDWs are uneducated and unintelligent (Hsia & Smales, 2011), which are often baseless. Filipina
FDWs (henceforth FilDWs) in HK, on average, have 15.1 years of schooling, with many holding college degrees (Tang & Chor, 2012). Furthermore, FilDWs are at least bilingual (in Tagalog and English), with some being trilingual or multilingual. As an integral part of many HK children’s home language environment, FilDWs may therefore play a significant role in the English language acquisition of such children.

**Hong Kong’s Language Environment**

Cantonese is the most widely spoken language in HK, with 88.2% of the population using it as their usual spoken language (Li, 2022). The other two major languages in the region are Mandarin and English, with 42% and 40% of the population respectively being able to speak each (Poon, 2010). Despite HK no longer being a British colony after 1997 and less than 5% of the population above the age of five using English as their usual spoken language, English continues to enjoy a very high level of prestige in post-colonial HK and it remains one of the major languages of instruction (Hansen-Edwards 2018; Li, 1999; Wakefield, 2021). HK primary and secondary education mainly divide into Chinese Medium of Instruction (CMI) and English Medium of Instruction (EMI) schools. There are also international schools, but they are only attended by 6% of students. CMI schools conduct all classes (apart from English language classes) in Cantonese or Mandarin.

Notwithstanding the predominant use of English in EMI schools, research suggests that the standard of English in HK declined drastically after the turn of the century, which lead to the introduction of the "Fine-tuning Medium of Instruction Policy" by the HK government in 2010 (Poon, 2016). Not only does the local government want to see high standards of English, but the majority of HK parents desire for their children to be proficient in English, partly because all universities in HK almost exclusively use English as the medium of instruction (Wang & Kirkpatrick, 2020). However, perhaps mainly due to the lack of an English learning environment outside of the classroom, English in HK can be better described as a foreign language (FL),...
rather than an L2 (Li, 2009). Local Hong Kongers being trained as English teachers also reported facing numerous challenges when it comes to effective communication in English, with one of the main reasons cited being lack of opportunity to converse in English outside classrooms (Gan, 2012).

The above suggests that, despite the prestige that English enjoys in HK, there is a general lack of proficiency and performance of the language among HK locals. The picture is, however, not the same for all Hong Kongers, with a lot being highly proficient in English (Wakefield, 2021). The presence of a FilDW in many households means that English will be used at home on a regular basis, which could possibly make English for children growing up in such homes an L2, rather than an FL.

The complexity, accuracy and fluency triad

The combination of complexity, accuracy, and fluency (CAF) has frequently been used in previous empirical studies to assess, among others, language learning progress and language proficiency (Housen & Kuiken, 2009; Tavakoli & Wright, 2020). Furthermore, factor analysis provides evidence that all three of the CAF triad must be considered to draw meaningful conclusions on learners’ overall L2 proficiency (Housen et al., 2012, p. 3).

Although there is agreement among most scholars that CAF forms an essential part of L2 research, no consensus exists when it comes to the definition of complexity, accuracy, and fluency. First, complexity is seen as the most challenging construct to define (Housen & Kuiken, 2009; Housen et al., 2012; Juola, 2008; Pallotti, 2009, 2015), with various definitions. For the purposes of this study, (syntactic) complexity refers to the number of verbs, nouns, adjectives and adverbs per independent clause, which relates to Yuan and Ellis’ (2003) definition, the extent to which the language produced in performing a task is elaborate and varied. The primary basis for evaluating syntactic complexity in this way is rooted in a method proposed by Cook and MacDonald (2012), where they use the number of verbs, nouns, and adjectives per independent clause to index
syntactic complexity (see below).

Second, accuracy, often refers to the degree of deviation from a norm (Pallotti, 2009). Although some scholars (notably, Housen et al., 2012, p. 4) argue that such a definition is problematic in that it does not specify the nature and extent of the deviation, and that it would be preferable to define accuracy in terms of appropriateness and acceptability. Our view aligns with that of Pallotti (2009), arguing that adequacy should be assessed in addition to all three constructs constituting the CAF triad. Furthermore, although being important additional elements of language proficiency, pragmatic competence and appropriateness cannot easily be measured effectively (Sell et al., 2019).

In view of the above, this study will adopt a more traditional view to define accuracy by measuring it against a predetermined norm, i.e., descriptive grammar. This is chosen in favour of prescriptive grammar, as certain forms of prescriptive grammar, like whom (as in the sentence, I do not know with whom I will go), now appear less than the “incorrect form”, who, in major corpora, like the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC). Moreover, the authenticity of descriptive grammar has been recognised for more than a quarter century already (see, e.g., McCarthy & Carter, 1995).

Third, there is consensus among scholars that fluency is an essential element of linguistic performance (e.g., Housen & Kuiken, 2009; Skehan, 2009; Yuan & Ellis, 2003), but there is no explicit agreement about the exact definition of fluency (Ellis, 2009). Given the diverse definitions of fluency, clarification of what constitutes L2 spoken fluency is needed.

Utterance fluency measures the number of words or syllables uttered per second or minute. This is a useful and widely used indicator of fluency, but there is more to (L2) fluency than just the number of words/syllables
uttered in a given time frame. One may not always be more comprehensible when producing more syllables in a certain time frame. Slowing down or pausing are commonly used as markers of dysfluency. However, some of these “dysfluency markers” are part of normal L1 speech and should therefore not by default be seen as signs of dysfluencies in L2 speakers. Rather than categorically classifying these as markers of poor fluency, one should consider whether the way something was expressed by an L2 speaker would be considered appropriate and normal if it was uttered in the same way by a native speaker (Tavakoli & Wright, 2020). In this regard, Davies (2003) showed that mid clause pauses are plentiful in L2 speech, whereas normal L1 speech is not typically characterised by mid clause pauses. Pauses at the end of clauses are, however, not uncommon in normal L1 speech (Skehan, 2009). It therefore makes more sense to classify mid clause pauses as markers of dysfluencies, but to exclude pauses at the end of clauses from such classification (unless the pauses at end of clauses are unnaturally long and hinders communication flow). It should also be noted that, while useful to refer to L1 speech, it is not a gold standard, as there is significant variation among L1 speakers.

Spending several seconds to come up with an answer or idea is not necessarily indicative of inferior fluency either. Depending on the nature of the answers or ideas required, even the best speakers may spend some time thinking before producing speech. Therefore, despite some studies (e.g., Tavakoli & Skehan, 2005) using total silence time as a measure of dysfluency, it does not seem like the most reliable measure. In deciding which fluency measure/s to implement, one would need to define what fluency entails.

Lennon (1990), in his pioneering work on establishing universal spoken fluency measures, distinguished between fluency in a broad and narrow sense. He pointed out that the broad sense of fluency, includes judgements on other aspects of performance, like pronunciation, that should be measured separately. The narrow sense seems more appropriate, as it focuses on speed of delivery, with native-like rapidity cited as a popular definition of fluency that was widely applied in the late 1980s. Since then, other definitions have been
suggested, but there still does not appear to be a universal definition of spoken fluency or a standardised way of measuring it.

Considering the above-mentioned issues relating to fluency, as well as the fact that this study is not dealing with absolute, but rather with relative scores (those of children from homes with FilDWs vs. those from homes without FilDWs) the definition of L2 fluency for the purposes of this study will be as follows: *the ability to produce L2 English relatively rapidly and relatively free from dysfluency markers, while applying pauses relatively effectively*. An advantage of this definition is that it will not by default regard pauses as dysfluency markers, but do so only after considering whether they hinder, enhance, or do not affect normal flow of speech. A further benefit of this definition for a study such as this is that it allows for direct comparisons between two groups, rather than comparing each group to a norm and indirectly comparing the groups to each other by calculating how far each group deviated from the idealised norm.

In investigating CAF, it should be noted that planning time allotted to tasks can also play a role (e.g. Ahmadian, et al., 2015). However, since planning time is not a factor that we investigate in the present study, we will not go into any further detail here.

*Reading fluency*

Reading accuracy refers to the extent that participants produced target like reading (i.e., did they actually read what's written in the passage that they were asked to read). The target for reading accuracy was therefore the actual text provided and those deviating (e.g., omitting, adding to or contracting some of the text) the least from it received the highest reading accuracy scores.
One of the first projects in this area, focusing on phonological development, was that of Leung (2010), investigating the impact of FilDWs on HK kindergartners’ English accent. The study explored whether children’s daily exposure to Philippine English (PE) would result in some of PE’s phonological features transferring to them. Results showed no such transfer to the children’s speech and it was concluded that FilDWs’ distinctive English accent has no bearing on the L2 English of children in the HK households they are employed in. Leung (2011) demonstrated that although speech production of children under FilDWs care was again unaffected (no traces of PE), speech perception was positively influenced. Recordings of minimal pairs produced in PE were played to 10 participants. Results showed that children with FilDWs were easily able to distinguish between similar sounding words (e.g., fan vs. pan vs. van) produced in PE, while those who never had a FilDW were unable to make such distinctions.

As a follow up, Leung (2012), launched a larger scale investigation with 94 participants comprising children aged four and a half to six from four kindergartens, and 11- to 14-year-old pupils from two secondary schools. 60 participants with FilDWs (experimental group) and 34 without FilDWs (control group) were assessed on their ability to comprehend single words read in a carrier sentence in American English (AE), British English (BE), Hong Kong English (HKE), and PE. Results showed that participants in the experimental group understood words in BE, AE and HKE as well as those in the control group, Leung concluded that children in HK with FilDWs are not inferior to those without FilDWs as far as word level comprehension in the English varieties assessed is concerned. The findings also provide evidence that participants in the experimental group are superior with regards to identifying sounds of PE, yet their production did not contain traces of PE (Leung & Young-Scholten, 2013).

Another study, in which 50 third year HK kindergartners participated, showed that the presence of a FilDW helps children develop a significantly larger English vocabulary, compared to those without FilDWs (Chan &
McBride-Chang, 2005). Authors suggested that the superior English ability displayed by those from homes with FilDWs is probably because they needed English for daily communication with their helpers, as opposed to simply using it at school, where English was taught as a formal subject.

To date, the largest study investigating the impact of FilDWs on HK children’s English was conducted by Tse et al. (2009). They examined FilDW’s influence on HK children’s reading proficiency by assessing 4352 grade four students. Participants were required to read two of eight 400- to 700-word English passages from the Progress in International Reading Literacy Study (PIRLS) 2001, standardised for their age group, and answer questions about the content. All eight passages were distributed equally among research participants and results showed a strong positive correlation between participants’ reading proficiency and the presence of a FilDW at home. The correlation was significantly stronger than that between children’s reading proficiency and whether or not their parents spoke some English.

Wolfaardt (2015) was the first to investigate the impact of FilDWs on both a receptive- (listening) and a productive skill (speaking). This study assessed the influence that FilDWs have on the English spoken fluency and listening comprehension of primary school children enrolled at CMI schools in HK. Participants comprised two groups of eight- to 12-year-olds (six boys and four girls). Participants from Group 1 came from homes with FilDWs, while those from Group 2 were from homes without FilDWs. Results showed Group 1 consistently outperformed Group 2 by substantial margins in both receptive and productive skills. These serve as evidence that children from households with FilDWs may have superior L2 English listening comprehension and spoken fluency skills, compared to similar children from homes without FilDWs.

A longitudinal study, with 142 participating HK children (73% of the initial sample) remaining on board till the fifth and final year of the study showed that children from homes with English speaking FDWs acquired a
significantly larger initial English lexicon, compared to those from homes with non-English speaking- or no helpers (Dulay et al., 2017). Furthermore, there was no negative correlation between the presence of an English speaking FDW and Chinese vocabulary development, although those in the group with English speaking FDWs did have poorer performances when it came to initial Chinese character recognition.

Wolfaardt (2022) is the most recent study on the topic. This study compared two groups of eight- to nine-year-olds attending the same EMI primary school in HK, with one group (n=34) coming from homes with FilDWs and the other group (n=30) from households without a FilDW. Results showed that participants from homes with FilDWs had a significantly larger L2 English receptive vocabulary, while there were no significant differences between the groups on the other tests (WMC, L1 Cantonese-receptive vocabulary and word reading). It was concluded that FilDWs have a positive impact on HK primary school children’s L2 English receptive vocabulary, while there is no trade-off effect on their Cantonese receptive vocabulary or word reading.

Overall, FDWs/FilDWs seem to have a positive impact on children’s L2 English. There is, however, one study from which some evidence emerged that the L2 development of those cared for by FilDWs may suffer. Cheuk and Wong (2005), investigating the relationship of FilDWs’ presence and the L2 English development of children under the age of three with a specific language impairment (SLI), found a negative correlation. Yet, there is no evidence that these results could be generalised to those without an SLI.

This study

Research Questions

This study focuses on the potential impact that FilDWs have on specific facets of HK primary school children’s L2 English. Data was collected with the aim of answering the following research questions (RQs):
RQ1: Do FilDWs influence the level of L2 English CAF of the children in their care?

RQ2: Do FilDWs influence the level of L2 English reading accuracy and fluency of children in their care?

Unlike previous research, this study makes the critical distinction between types of schools in HK and drew the entire sample from an EMI population in order to eliminate confounding effects from the differences between types of schools. This study is also the first investigation looking into the impact of FilDWs on HK children’s language proficiency to employ a WMC test. Prior to this study, CAF has not been employed in a study examining the role of FilDWs on HK children’s language acquisition.

Target population and participants of this study

Participants were drawn from EMI schools. The rationale behind this decision was as follows: (i) the native-like level of English proficiency of pupils attending international schools makes them unsuitable for a study like the present one, (ii) learners attending CMI schools may not be proficient enough in English to complete a meaningful set of English (Evans, 2013). To avoid the potential confound of age, we focused on eight- and nine-year-olds.

Parents of participating children filled out a questionnaire that sought other demographic information. One of the primary purposes of this questionnaire was to ensure that all participants come from homes with similar SES. Information around the following were also collected: main and secondary languages as well as the quantity (in hours per week) of interaction between parents and participants, and between helpers and participants; the number of siblings participants have, including birth order; the amount of English TV watched per week; the types of TV programmes watched; the number of English books at home; the amount of time per week participants spent reading these books; the amount of time spent using English on electronic devices, such as computers, phones and iPads; and whether participants received extracurricular English classes.
A total of 64 children participated. Prior to commencement of testing, all participating children returned signed consent forms from their parents. Parents were informed about the nature and purpose of the study before providing consent. Participant groupings were as follows: 34 with FilDWs (17 males, 17 females; mean age: 8;11, SD = 6 months, age range: 8;1 to 9;10) and 30 without FilDWs (15 females, 15 males; mean age: 8;11, SD = 6 months, age range: 8;2 to 9;10).1

Materials and procedures

Participants took part individually. The materials used are as follows: (1) a picture description task, (2) a reading passage, and (3) a WMC test. (1) was designed specifically for this study; (2) was taken from a children’s textbook (Smith & Boyle, 2008), and (3) is part of a standardised test. A brief description of each measure used, and procedures followed is given below.

(1) Picture description task

Participants described 15 colour photographs (see Appendix for a sample), depicting images familiar to them. Participants were presented with one photograph at a time and had to describe each. Before they commence, participants were told to provide sufficient detail so that people who had not seen the photo can understand. They were instructed to not simply say something like, I see a bird, but to be more specific, referring to details, such as colours, the background, season and time of day. All participants were allowed planning time before every picture and they were recorded as they rendered their descriptions. Planning time was neither compulsory nor time constrained, in line with previous research (Ellis, 2009). All participants took less than 10 seconds to plan.

1 We thank one of the reviewers for raising questions about the level of education of the helpers. The average educational level of FilDWs in this study is a post-secondary certificate, with thirteen (nearly 40%) holding degrees. A recent study about Filipino domestic helpers’ English also concluded that their variety should be regarded as more advance than a basilect (Borlongan 2022).
Transcription and coding of the picture description task

The first three photographs in the speaking test were treated as practice items and transcriptions are the 90 seconds from the fourth picture onwards. Using fixed, predetermined segments of recordings to transcribe for analyses is normal practice (e.g. Sato, 2014). Transcriptions were checked for fluency, grammatical accuracy, and syntactical complexity. For illustrative purposes, two extracts from actual transcripts are shown below.

The words/parts in red are indicative of grammatical errors, while colons mark elongated sounds. Underlined parts indicate dysfluency markers (false starts, inappropriate silences/pauses, repetitions, elongated words, and hesitations). The abbreviations following the participant’s turn indicate the following: DFM = dysfluency markers, GE = grammatical error/s, S = number of sentences (every independent clause is counted as a sentence for calculation purposes, as recommended in Cook and MacDonald (2012)), while VNAA = verbs, nouns, adjectives and adverbs. False starts and repetitions are not brought into the equation when calculating GE, S and VNAA.²

Extract 1

01:43-01:56 uhm it is (pause) a student sitting (pause) sitting on a (pause) sitting (pause) bel (pause) sitting under the tree: doing its homework and writing (pause) things. [13 words, 18 syllables, 13s, 14DFM, 2GE, 1S, VNAA: 8]

Extract 2

01:30-01:44 with (pause) (NA) blue: face (pause) climbing (pause) on: (pause) a tree (pause) and it has four legs. [12 words, 13 syllables, 14s, 6DFM, 1GE, 2S, VNAA: 7]

² We thank one of the reviewers for pointing out the conundrum of having to assess fluency using dysfluency measures. This is indeed an important and interesting epistemological question to consider. However, at the moment of writing, we have not come across and nor able to devise an alternative way to measure fluency, so in this paper we have followed the tradition of existing literature. But it will indeed be useful to explore ways to break free from the "deficient" driven model in future.
In Extract 1 all pauses are underlined as dysfluency markers, since every one of them appears in the middle of a clause. The first grammatical error in this extract relates to the incorrect use of articles. No tree has previously been identified by this participant, so the indefinite article *a* should be used, instead of the definite article *the*. The second grammatical error is the incorrect choice of pronouns. The homework referred to by the participant is done by a female student, requiring the possessive determiner *her*, instead of the possessive pronoun *its*. The calculation of VNAA was as follows: 4 verbs (*is, sitting, doing, writing*) + 4 nouns (*student, tree, homework, things*) = 8.

In Extract 2 the red letters NA indicate that there is no article where the indefinite article *a* should have been used. The second and final pauses in this extract are not underlined as dysfluency markers, as they appear at the end of independent clauses. Such pauses are considered normal and natural (Skehan, 2009). This extract contains two adjectives (*blue and four*), two verbs (*has and climbing*) and three nouns (*face, tree, legs*), making the total of VNAA 7.

(2) Reading passage

Each participant was required to read the same passage aloud, while they were recorded. They were presented with the text printed on an A4 page and could commence reading when they were ready. Most opted to start reading straight away.

*Scoring of tasks 1 and 2*

Syntactic complexity scores were calculated in two ways. The first measure is an adapted version of that used by Cook and MacDonald (2012). They calculate complexity by dividing the total number of verbs, nouns and adjectives (VNA) by the total number of sentences produced. Useful as this method is, it does have a shortcoming, as will be illustrated with two sample sentences below:
Example 1: *The girl sings beautifully.* [VNA: 2, VNAA: 3]

Example 2: *The beautiful girl sings.* [VNA: 3, VNAA: 3]

Example 1 contains four words in total, including the noun *girl* and the verb *sings.* This makes the syntactic complexity score of this sentence two, in Cook and MacDonald (2012)’s method. Example 2, which is very similar to the first, additionally contains the adjective *beautiful,* giving it a complexity score of three. However, it is easy to see that *beautiful* is already present in the first example. The only differences are that it appears after, rather than before, the verb and contains the suffix *ly.* If anything, a prefix, infix or suffix makes a word more, instead of less complex. Although we are dealing with syntactic, rather than lexical complexity in the analysis, it does not seem reasonable that one four-word sentence should be classified as one and a half times more complex than another, where the only differences are that the less complex sentence has a different word order and a modified, longer form of a word contained in the more complex sentence. It would be more accurate to classify these sentences as equally complex. A modified application, where we also consider adverbs, will do exactly this. In such an adapted version of Cook and Macdonald’s method, instead of ignoring the adverb *beautifully* that appears in Example 1, it will be included in the calculations, resulting in a complexity score of three for the sentence in Example 1. Note also, that the score does not change for the sentence in Example 2. Both are now classified as equally complex.

The adapted version is therefore more accurate than the original, but it still does not tell the full story. In some cases, to get a clearer, more accurate picture of syntactic complexity, additional information is required. Two further examples below illustrate this:

Example 3: *Two fat men make hot tea quickly.* [VNAA: 7]

Example 4: *Your money in the envelope lies neatly next to mine under the crockery in the cupboard above the refrigerator.* [VNAA: 7]
Example 3 contains seven words in total and is comprised of three adjectives (*two, fat, hot*), two nouns (*men and tea*), one adverb (*quickly*) and one verb (*make*). This gives it a complexity score of seven. Example 4 contains 19 words, including five nouns (*money, envelope, crockery, cupboard, refrigerator*), one adverb (*neatly*), and one verb (*lies*). According to the modified version of Cook and Macdonald’s method, proposed above, these two sentences are of equal complexity. This is a more accurate classification than the original version, that would not consider the adverb neatly, thus making the sentence in Example 4 less complex than that in Example 3. However, even the modified version’s classification that assigns equal complexity to both examples does not capture the fact that Example 4 contains nearly three times as many words as the first. It would therefore be more accurate to include the additional classification of words per sentence to be used in conjunction with the first, that solely considers VNAA. Furthermore, it should be noted that none of the additional words contained in the second example are false starts or repetitions.

Applying the conclusions drawn from the above discussion, calculation of the two measures employed to capture syntactic complexity scores were done as follows: (1) VNAA/S and (2) Words/S. For example, a participant that produced 100 VNAA and 200 words in 30 sentences during the 90 second transcribed extract will have a Syntactic Complexity 1 score of 3.33 and a Syntactic Complexity 2 score of 6.66 (100/30 and 200/30, rounded off to the second decimal) according to these measures.

Grammatical accuracy scores as well as those on the third fluency measure were arrived at by calculating the number of grammatical errors/dysfluency markers per 100 words and deducting that from 100. Should a participant’s speech include, for example, 10 grammatical errors and 35 dysfluency markers per 100 words, their accuracy and Fluency 3 scores would be 90 and 65, respectively. Dysfluency markers were not included in the calculation of grammatical accuracy scores. Fluency 3 scores were calculated similarly as in Lennon (1990).
The calculation of the first and second fluency measures is a simple division of the number of words (first measure) or syllables (third measure) by 1.5 (90 seconds equals one and a half minutes), which gives the number of words/syllables per minute. If, for example, a participant produced 120 words and 180 syllables, their Fluency 1 and Fluency 2 scores would be 80 (120/1.5) and 120 (180/1.5) respectively. As in the case of the third fluency measure, as well as grammatical accuracy scores, parts identified as dysfluency markers were not included in the calculation of Fluency 1 and 2, as recommended by Lennon (1990). Expressing fluency in wpm. has previously been done by Oh and Lee (2012), while syllables per minute was employed by De Jong and Perfetti (2011). Reading fluency scores were derived from number of correct words read per minute, as in Pretorius and Spaull (2016).

To recapitulate, our study measured two types of accuracies: (1) Spoken accuracy (measured via the picture description task) focuses on grammatical accuracy and refers to the ability to produce target like language, which is in line with the definition used by Suzuki and Kormos (2020). The target/norm in our study is descriptive grammar and participants deviating the least from this obtained the highest spoken accuracy scores. (2) Reading accuracy which refers to the extent that participants produced target like reading (i.e., did they actually read what's written in the passage that they were asked to read). The target for reading accuracy was therefore the actual text provided and those deviating (e.g., omitting, adding to or contracting some of the text) the least from it received the highest reading accuracy scores. For the purposes of our study, we operationalised accuracy in terms of deviation from a norm, but we acknowledge that this “deficient view” can be problematic.³

³ We thank one of the reviewers for drawing our attention to the matter of appropriateness and pragmatic competence. Appropriateness has been evaluated in recent L2 studies investigating argumentative speech (e.g., Suzuki & Kormos, 2020) and speech acts (Li et al., 2023), but is less of a concern when it comes to picture description tasks. Participants of this study were briefed on what would be appropriate descriptions of the pictures presented to them and given 3 practice items to help ensure that they all described the pictures appropriately (as noted above). However, our study was not designed to investigate appropriateness directly. The research questions were not aimed at obtaining insights into pragmatic competence either, as pragmatic competence relates to conversational contexts (Kecskes, 2019), while the speech elicitation task in the current study is not of a conversational nature.
Cross-checking of transcriptions

All transcriptions were done by author 1, while a trained linguist (a native English speaker and final year undergraduate student in linguistics at the University of Cambridge, with three years of experience in transcribing audio recordings for linguistic analyses) listened to 26 (20.31%) randomly selected recordings and checked the transcriptions thereof. The levels of agreement were as follows: (1) Reading: WPM 99.59%; Accuracy: 99.60% (2) Speaking: Fluency 1: 98.58%, Fluency 2: 98.77%, Fluency 3: 90.54%, Accuracy: 99.89%, Syntactic Complexity 1: 100%, Syntactic Complexity 2: 99.40%. The overall level of agreement on all of these measures is 98.03%.

(3) WMC test

This is a subtest of the standardised test battery, Clinical Evaluation of Language Fundamentals, fourth edition (CELF-4) (Semel, et al., 2003), which consists of 15 items, with two task types, namely a forward digit span task (assessing the storage component) and a backward digit span task (assessing the processing component). Each item had a pair of a fixed number of digits (e.g., 1 (a) “3-5” and 1 (b) “7-2”) that participants should repeat after hearing them (i.e., forward digit span task; e.g. the correct answers are “3-5” and “7-2”) or reverse the order and then repeat (i.e., backward digit span task; e.g., the correct answers are “5-3” and “2-7”). The first eight items were a forward digit span task, and the last seven a backward digit span task. In both tasks, the question consisted of a pair of two digits, with a digit added to each subsequent question until the number of the digits was nine for the forward digit span task and eight for the backward digit span task. Participants continued with the test until they incorrectly answered both pairs of a question set. When this happened during the forward span task, the participant moved to the backward span task, continuing until they got both parts of the same question wrong, at which point the test was terminated.

Scoring of the WMC test
One mark was awarded for the correct answer of each part of every pair on all items. This made the maximum possible scores 16 (8 questions × 2 marks) and 14 (7 questions × 2 marks) for the forward- and the backward span tasks respectively, giving a total possible score of 30. Participants had to repeat every digit in the correct order (i.e., the same order in which it was read to them for the forward task and reverse order for the backward task) to receive any score. For example, if “5-0-9-1” was read to them in the forward span task, one mark was awarded for answering “5-0-9-1”. When a participant incorrectly repeated (e.g., “4-0-9-1”), missed a digit (e.g., “5-9- 1”), or mixed up the order in any way (e.g., “5-0-1-9”), they received no score.

Establishing validity of the language tests.

The validity of tasks 1 and 2 was tested via the Content Validity Index (CVI) (Ozer et al. 2014). The CVI aims to establish the extent to which the content of a test adequately represents the domain that the test is supposed to measure through the judgement by at least two experts on the relevancy of each item on the scale against the test’s objectives. Two judges were selected, based on their linguistic repertoire and level of expertise. Judge 1 was a postdoctoral fellow in linguistics, Judge 2 is an experienced language teacher at a local HK school, both are Cantonese-English bilinguals. They independently applied the CVI to both language measures. The CVI of both tests was established at 1.0 (in case of the picture description task, the score of 1.0 was given for both the Scale-CVI and Item-CVI), therefore can be accepted as valid measures (an S-CVI of .80 or higher indicates acceptable validity (Polit & Beck, 2006)). Before implementing the tests, a pilot was run with eight children from the target population. As everything ran smoothly, the tests remained unchanged for the main study. Since the WMC test is a subtest from a standardised test we did not further establish its content validity.

Results and discussion

Independent t-tests were conducted to compare the groups. As seen from Table 1 below, participants from homes with a FilDW obtained significantly higher scores on the English language tests. This applies to all
measures, which include reading accuracy and fluency, all three spoken fluency measures, spoken accuracy, and both spoken complexity measures. Frequent self-corrections were made during the picture description task (often involving grammar points, like gender marking). These are perhaps indicative that children in HK start to monitor their own L2 speech much earlier than at the age of 12, as previous research suggested (Krashen, 1981, p. 35).

**Table 1**

*Results English language tests*

<table>
<thead>
<tr>
<th>Variables</th>
<th>+FilDW (N = 34)</th>
<th>-FilDW (N = 30)</th>
<th>Effect Size&lt;sup&gt;6&lt;/sup&gt;</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading accuracy</td>
<td>97.99 1.89</td>
<td>95.99 4.36</td>
<td>.60</td>
<td>&lt;.05**</td>
</tr>
<tr>
<td>Reading fluency</td>
<td>140.09 18.31</td>
<td>116.30 19.59</td>
<td>1.25</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>Spoken fluency&lt;sup&gt;1&lt;/sup&gt;</td>
<td>95.36 15.72</td>
<td>69.93 14.55</td>
<td>1.67</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>Spoken fluency&lt;sup&gt;2&lt;/sup&gt;</td>
<td>116.14 20.96</td>
<td>88.75 19.71</td>
<td>1.35</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>Spoken fluency&lt;sup&gt;3&lt;/sup&gt;</td>
<td>68.65 10.88</td>
<td>41.02 17.38</td>
<td>1.91</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>Spoken accuracy</td>
<td>95.79 2.75</td>
<td>92.69 3.99</td>
<td>.90</td>
<td>&lt;.01**</td>
</tr>
<tr>
<td>Spoken complexity&lt;sup&gt;1&lt;/sup&gt;</td>
<td>5.48 1.36</td>
<td>3.96 0.97</td>
<td>1.29</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>Spoken complexity&lt;sup&gt;2&lt;/sup&gt;</td>
<td>9.26 2.37</td>
<td>6.68 1.54</td>
<td>1.29</td>
<td>&lt;.001***</td>
</tr>
</tbody>
</table>

<sup>1</sup>Words per minute; <sup>2</sup>Syllables per minute; <sup>3</sup>Percentage of dysfluency markers subtracted from total words; <sup>4</sup>Number of verbs, nouns, adjectives and adverbs per sentence; <sup>5</sup>Number of words per sentence; <sup>6</sup>Cohen’s D was used to calculate effect size in Table 1 and 2. Significant at *95%/***99% confidence interval.

Table 2, displaying the scores of the WMC test, shows no significant difference between the two groups. The processing component (requiring manipulation of information by participants) is perhaps the most important here, as the storage component (not involving any information manipulation) is essentially the same as short-term memory.

**Table 2**

*Mean scores of the WMC Test*
The results of the English language tests clearly indicate that children from homes with FilDWs have better English reading fluency and accuracy as well as superior CAF in spoken English. No previous study investigating the impact of FilDWs on HK children’s L2 English considered the effect that FilDWs potentially have on spoken CAF. The results of this study provide evidence that L2 oral English of children growing up in households with FilDWs is characterised by significantly better CAF. The superior fluency scores are in line with previous research that found a positive correlation between the presence of FilDWs and L2 spoken fluency (Wolfaardt, 2015). Considering that language practice involving extended periods of meaningful participation is required for the automatization of L2/FL skills, it is perhaps no surprise that children from homes with FilDWs outperformed those from homes without FilDWs. Children from homes without FilDWs, simply do not have the same amount of meaningful participation in English language activities, like conversations. Thus, for school aged children, FilDWs provide significant, additional (to school) opportunities for children to actively participate in authentic English activities. In the case of preschoolers, FilDWs may be the primary providers of such opportunities.

The results of the language tests add to the existing evidence (e.g., Buac et al., 2014; De Houwer, 2011; Hart & Risley, 1995; Krashen, 1976; Tardif et al., 1997) that home language environment, particularly that of primary caregivers, plays an important role in children’s language proficiency and performance. More specifically, this study shows that a FilDW at home could be a critical factor in HK primary school children’s L2 English acquisition. As FilDWs have been employed in large numbers in Taiwan (Lan, 2006), where English is also spoken as an L2, the same may apply to Taiwanese children’s L2 English.

<table>
<thead>
<tr>
<th>Variables</th>
<th>+FilDW (N = 34)</th>
<th>-FilDW (N = 30)</th>
<th>Effect Size</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Storage component</td>
<td>10.47</td>
<td>2.26</td>
<td>9.53</td>
<td>1.93</td>
</tr>
<tr>
<td>Processing component</td>
<td>5.88</td>
<td>2.29</td>
<td>5.53</td>
<td>2.06</td>
</tr>
<tr>
<td>WMC (total)</td>
<td>16.24</td>
<td>4.07</td>
<td>15.07</td>
<td>3.39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variables</th>
<th>+FilDW (N = 34)</th>
<th>-FilDW (N = 30)</th>
<th>Effect Size</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Storage component</td>
<td>10.47</td>
<td>2.26</td>
<td>9.53</td>
<td>1.93</td>
</tr>
<tr>
<td>Processing component</td>
<td>5.88</td>
<td>2.29</td>
<td>5.53</td>
<td>2.06</td>
</tr>
<tr>
<td>WMC (total)</td>
<td>16.24</td>
<td>4.07</td>
<td>15.07</td>
<td>3.39</td>
</tr>
</tbody>
</table>
The positive effects of FilDWs on HK primary school children’s English are clearly demonstrated through the results of the English language tests administered during this investigation. Furthermore, English is increasingly being used as a lingua franca (Suzuki, 2019) and learning to do so from one’s childhood years is a very useful skill to acquire. Unfortunately, as mentioned in the background, FDWs in HK receive little or no recognition, do not enjoy the same rights as most other workers in HK, and many have been subjected to numerous forms of mistreatment. Although not justified, some of the abuse stems from HK employers’ general low regard for FDWs. If this view is changed, it may result in better treatment of FilDWs and FDWs overall. While it is welcoming that the HK government recently increased the penalty against agencies exploiting FDWs sevenfold (Chung & Mak, 2020), it is disconcerting that HK employers have been treating FDWs worse during the ongoing Covid-19 pandemic (Lui et al., 2021).

Given the significant positive impact that FilDWs have on local, bilingual HK children’s L2 English acquisition, it is hoped that the present study’s findings would contribute to changing the general negative view towards them, giving them more recognition and improving their status as well as their living and working conditions in HK. Recent research argued that FilDWs’ English proficiency could be used as a strength in “bargaining” to get better treatment (Tong & Jiang, 2020). There are some HK parents, educators and academics who appreciate FilDWs’ contribution to local children’s English proficiency and they could perhaps, through word-of-mouth, contribute to a shift towards a more positive view, coupled with better treatment of FilDWs.

**Conclusion**

This study set out to investigate the potential influence of FilDWs on HK primary school children’s L2 English proficiency. Overall, the results indicate that FilDWs have a positive impact on the L2 English proficiency of participants attending an EMI school. More specifically, participants with FilDWs obtained superior scores on tests measuring English spoken CAF as well as reading accuracy and fluency. The
evidence presented in this investigation, provides a strong argument that FilDWs, as an extra source of 
language input, may be responsible for the superior English displayed in the children they interact with in 
English for several thousand hours over many years. Although the English variety of FilDWs might not be 
considered “mainstream” and indeed stigmatised, our findings demonstrate a clear benefit of such exposure 
on the L2 English development of children under their care. This, in our view, is welcome and reassuring, 
especially given the current discussions around the need to “decentre”, “decolonise” and “democratise” 
language learning and teaching (e.g. Canagarajah 2022; Saraceni & Jacob 2021).

Acknowledgements

We would like to thank our participants; without their participation our study would not have been possible. 
We would also like to thank the editors and the reviewers for their kind comments which have helped us 
enhance the clarity of our paper. Thanks goes to the support from the editorial team of Applied Linguistics as 
well.

References

Ahmadian, M., Tavakoli, M., & Vahid Dastjerdi, H. (2015). The combined effects of online planning and 
task structure on complexity, accuracy and fluency of L2 speech. The language learning journal, 43(1), 41- 
56.

67-79.

Buac, M., Gross, M., & Kaushanskaya, M. (2014). The role of primary caregiver vocabulary knowledge in 
the development of bilingual children's vocabulary skills. Journal of Speech, Language, and Hearing 
Research, 57(5), 1804-1816.

Canagarajah, S. (2022). Challenges in decolonizing linguistics: the politics of enregisterment and the 

Societies, 6(2), 179-193.

Cheung, T. (2022). Coronavirus: domestic helper shortage continues to plague Hong Kong as number of such workers hits new low. South China Morning Post: Coronavirus: domestic helper shortage continues to plague Hong Kong as number of such workers hits new low | South China Morning Post (scmp.com)


Appendix: Sample item for picture description task.