Investigating the effect of codeswitching and L2 exclusive instructional strategies on learning and retention of L2 vocabulary

Author:Jon Nilsen

University of Birmingham

British Council ELT Master's Dissertation Awards: Commendation

INVESTIGATING THE EFFECT OF CODESWITCHING AND L2 EXCLUSIVE INSTRUCTIONAL STRATEGIES ON LEARNING AND RETENTION OF L2 VOCABULARY

ABSTRACT

This study sought to investigate the effect of teacher codeswitching, as compared to monolingual (L2) instruction, on L2 vocabulary learning and retention. Though a sizeable body of research on teacher beliefs and behavior regarding classroom L1 use has diminished the efficacy of monolingual pedagogy, the potential of strategic codeswitching for facilitating L2 learning has only recently come under empirical scrutiny. While past research has shown explicit vocabulary teaching and intentional learning to be advantageous, empirical evidence has been limited to reading comprehension studies giving little attention to the comparative value of providing lexical information in L1 or L2. However, evidence from psycholinguistic research on bilingual lexical development suggests similar advantages for codeswitching. Thus, 10 Japanese adult EFL learners were assigned to either codeswitching (CS) or English-only (NCS) conditions, and assessed regarding the quality and efficiency of learning and retention of 24 target words for which they had received condition-specific instruction and subsequent communicative exposure. Results echo similar studies that codeswitching is superior to L2-exclusive instruction for more accurate short-term learning, but participants' proficiency, among other limitations, potentially influenced this outcome. Results also show codeswitching to be more efficient despite long-term retention suffering more than L2-exclusive instruction. Pedagogical implications are discussed.

Contents

List of Tables and Figures

1 Introduction

In recent years, a fundamental question which has been the focus of considerable research and debate in the field of second language (L2) pedagogy is the role, if any, the learners' first language (L1) should have in the learning of other languages. The entrenched belief that languages are best learned without reference to learners' L1 has dominated theoretical research paradigms informing L2 pedagogy (e.g. Krashen, 1985; Long, 1981; Swain, 1985). However, these paradigms are often based on assumptions derived from observations of L1 acquisition (Turnbull & Dailey-O'Cain, 2009; Zhao & Macaro, 2014), and often ignore intrinsic EFL classroom realities or the fundamentally different nature of L2 learning, especially for learners who approach learning a new language with an already well established L1 system (Buzkamm & Caldwell, 2009). Following the path of previous studies conducted by Macaro and colleagues (e.g. Hennebry, Rogers, Macaro, & Murphy, 2013; Lee & Macaro, 2013, Tian & Macaro, 2012), the present study sought to investigate the potential value of utilizing learners' L1 (i.e. codeswitching) as a linguistic resource for L2 vocabulary learning and retention in comparison to monolingual strategies (i.e. definitions, explanations, contextual elaborations). In contrast to previous large-scale studies which used codeswitching to assist in more heavily teacher-fronted reading or listening comprehension activities, the present study takes advantage of a small sample size by introducing target vocabulary within a more communicative classroom environment.

The paper proceeds with a review of relevant literature followed by the present study in which 24 target vocabulary words were taught to two groups of Japanese adult EFL learners in a communicative classroom setting. A number of immediate and delayed post-instruction assessments are subsequently analyzed with regard to the research questions, the results of which are discussed and interpreted along with the study's limitations and pedagogical implications.

2 Literature Review

This section presents a brief historical overview of the monolingual principle, its influence on current L2 pedagogy, and an examination of the debates surrounding its efficacy in light of current empirical evidence and theoretical understandings of second/foreign language learning and development with a particular focus on vocabulary learning.

2.1 A Critical Reevaluation of Monolingual Instructional Practices

The exclusion of learners' own languages from the foreign language classroom in favor of monolingual teaching policies, the 'monolingual principle' (Howatt, 1984), represents an axiomatic thread running throughout modern language pedagogy, but one must wonder how such a principle came into such unquestioning acceptance and whether or not it was founded on pedagogically sound rationale.

2.1.1 A Brief Historical Overview of the Monolingual Principle

In response to increased migration of peoples around Europe and into the United States, classroom reference to learners' first languages (i.e. Grammar Translation) rapidly gave way to ‘naturalistic’ teaching methods, based on observations of child L1 acquisition, originating from the 'Reform Movement' (Howatt, 1984), such Gouin’s Series Method, which promoted the development of spoken communication (Brown, 2001; Richards & Rodgers, 2001). Proponents of these methods argued foreign languages could be taught without recourse to the learners' own language 'if meaning was conveyed directly though demonstration and action' (Richards & Rodgers, 2011: 11). As opposition to the use of learners' own languages was not uniform among leaders of the Reform Movement, such as Henry Sweet's pronounced advocacy for translation in vocabulary teaching (1899/1964:194 cited in Hall & Cook, 2012), staunch adherence to monolingual teaching arguably originated from the success of the Direct Method through Maximilian Berlitz's efforts (Hall & Cook, 2012: 275).

That confidence in monolingual teaching has largely endured a number of pivotal theoretical revolutions in understandings of language and language acquisition suggests its roots are more practical than theoretical (Hall and Cook, 2012: 274). Indeed, widespread acceptance of Berlitz's

monolingual model is often credited to practical factors, such as the inherent impracticality of bilingual instruction where monolingual native-speaker teachers teach groups of students of various linguistic backgrounds, and to the economic interests of materials publishers and language schools in English-speaking countries eager to 'promote monolingual products which could be implemented by native-speaker experts' (Hall & Cook, 2012: 275). Despite the Direct Method's inability to address criticism (e.g. poor theoretical grounding and non-transferability to public education settings (Brown, 2001; Richards & Rogers, 2001)), its emphasis on oral interaction and the dogmatic exclusion of learners' first languages would leave a lasting impression on language pedagogy throughout the twentieth century. Despite 'isolated voices of dissent' (Hall and Cook, 2012: 272), policy and practice still seem to operate as if the monolingual principle were widely accepted as 'common sense' (Cummins, 2007).

2.1.2 The Monolingual Principle's Influence on Current Pedagogy

As Hall and Cook (2012: 271-2) observe, the widely accepted belief that languages are 'best taught monolingually' has pervasively underpinned leading pedagogical literature throughout the twentieth century. Looking briefly at recent surveys of language teaching methodology (e.g. Larsen-Freeman, 2000; Richards & Rodgers, 2001) this belief underlies virtually 'every language teaching method that has found widespread official support', with the exception of a handful of 'alternative' methods (e.g. Community Language Learning, Suggestopedia) (Littlewood & Yu, 2011: 66). Indeed, even recent approaches, such as Communicative Language Teaching (CLT) and Task-based Language Teaching (TBLT), largely ignore learners' L1 unless strategies for its avoidance are mentioned (V. Cook, 2001: 404). Likewise, the theoretical framework supporting these methods, found within the significant body of second language acquisition (SLA) research pertaining to input and interaction carried out during in the 1980s and 1990s, either ignored or considered learners' first languages to be irrelevant to L2 acquisition (Macaro, 2009).

Of particular importance during this time were a number of hypotheses inspired by Chomsky's (1965) theory that language learning is derived from innate, subconscious functions in the brain. As it was claimed the L2 could be acquired through exposure to floods of comprehensible input (Krashen, 1982), interactive negotiation of meaning (Long, 1981; Pica, Young & Doughty, 1987), and being 'pushed' to provide comprehensible output (Ellis & He, 1999; Swain, 1985),

excluding learners' L1 in favor of maximizing L2 exposure and opportunities for its interactive use became the highest priority for educators. When L2 contact is limited to the classroom, such as in EFL contexts, any use of the learners' L1 was seen as interfering with their L2 development (Hummel, 2010; Littlewood and Yu, 2011; Turnbull and Dailey-O'Cain, 2009) and depriving them of valuable L2 exposure (Littlewood & Yu, 2011: 65). In this way, complete exclusion of learners' L1 in favor of maximizing L2 exposure has perpetuated the monolingual principle through what Macaro has termed the 'virtual position' (1997, 2001, 2009). Moreover, it is argued that the above theoretical and empirical support for the virtual position, in combination with the success of L2 immersion programs in Canada during the 1980s, played a significant role in influencing many governments, educators and publishers worldwide in both ESL and EFL contexts to 'accept the virtual position on target language use 'as '*best*' practice in second and foreign language learning and teaching' (Turnbull and Dailey-O'Cain, 2009: 3-4, original emphasis). Nevertheless, innovations in SLA literature often take a considerable amount of time to reach practitioners and policy makers in many parts of the world, and when they do, they are not necessarily received with equal enthusiasm (Hall and Cook, 2012: 272). In other words, although the virtual position has enjoyed significant support in mainstream literature for more than one hundred years, many educational contexts continue to employ learners' L1 in foreign language learning (Hall & Cook, 2012).

2.2 Acknowledging Classroom L1 Use

When teachers share their learners' L1, using it 'is so compelling that it emerges even when policies and assumptions mitigate against it' (Lucas & Katz, 1994: 558). Regardless, using learners' L1 has long been considered a 'skeleton in the cupboard ... a 'taboo' subject, a source of embarrassment', and especially for non-native teachers, 'a symptom of their failure to 'teach properly'' (Prodromou, 2002: 5). As Phillipson (1992: 191-2) points out, '[t]he overwhelming majority of teachers of English are non-native speakers', and yet they are expected emulate native-speaker English teachers. In EFL contexts where many teachers share their learners' L1, this expectation has had a 'devastating effect' on their ability to perform their jobs with confidence (Hall and Cook, 2012: 273). Despite virtual unanimity regarding the aim of maximizing teachers' L2 use (Turnbull & Arnett, 2002: 211), a majority of bilingual teachers still

see L1 use (i.e. codeswitching) in the classroom as 'unfortunate and regrettable but necessary' (Macaro, 2005: 68). This paradoxical perspective, labeled the 'maximal position' (Macaro, 1997, 2001), has been criticized for being more of a 'statement of inadequacy' than a sound theoretical framework on which representations and predictions of L2 learning and skill development can be based (Tian & Macaro, 2012: 370). Furthermore, for those holding this perspective, there is considerable evidence of a struggle with 'feelings of guilt' in deviating from the 'virtual position' ideal (e.g. Littlewood & Yu, 2011; Macaro, 2005; Mitchell, 1988). Indeed, guilt, as Macaro (2005: 69) rightly acknowledges, 'is not a healthy outcome of pedagogical debate'.

That teachers' classroom behavior seemingly subverts established institutional policies has led to significant accumulation of research investigating the functions and amount of bilingual teachers' classroom codeswitching. Polio & Duff (1994) identified some key pedagogic functions of codeswitching: grammar instruction, classroom management and administrative tasks, showing empathy/solidarity, translating unknown vocabulary, compensation for lack of comprehension, and in responding to learners' L1 use. Similar findings can be found in studies from Atkinson (1987), Macaro (1997, 2001), Rolin-Ianziti & Brownlie (2002), and Liu *et al.* (2004). Rolin-Ianziti & Varshney (2008), using Ellis's (1994: 577-8) classifications of classroom interaction, divide teachers' L1 use into 'medium-oriented goals', such as explaining vocabulary or teaching grammar (e.g. Polio & Duff, 1994; Kim & Elder, 2005, 2008), and 'framework goals', such as giving procedural instructions, or assigning homework (e.g. De La Campa & Nassaji, 2009; Macaro, 2001; Polio and Duff, 1994). Kim & Elder (2008) have drawn similar distinctions, but also recognize codeswitching for social goals, such as expressing concern and sympathy. Additionally, Edstrom (2006) suggests the debates revolving around L1 use transcend pedagogical matters, instead pointing to teachers' 'moral obligation' to recognize learners as individuals and to create supportive, affective environments by communicating interest, respect, and concern through 'judicious' use of the L1.

Affective support is particularly important for L2 learners' lack of proficiency, which is often cited as the main reason teachers codeswitch (see, for example, Macaro, 1997; Mitchell, 1988). In classroom environments where L1 use is strictly proscribed, learners' lack of proficiency can quickly breed frustration (Butzkamm, 2003; Butzkamm & Caldwell, 2009) and stress stemming

from feelings of inequity created by the disproportionate power teachers hold in relation to them (Brooks-Lewis, 2009). Thus, Littlewood & Yu (2011: 70) suggest that immersing 'disoriented and powerless' learners into environments deprived of any L1 support is potentially demotivating. Another often cited reason for codeswitching is teachers' own limited proficiency (e.g. Carless, 2004; Liu, *et al.*, 2004). Indeed, the 'purposeless', 'lazy' (Edstrom, 2006: 289), and 'time-consuming' use of the L1 by less proficient, 'disaffected teachers' has often resulted in situations where they 'succumb' to teaching lessons in the L1 (Butzkamm & Caldwell, 2009: 86). Even a limited cross section of studies concerning the amount of teacher L1 use show wide variation ranging from as narrow as 0% to 15.2% (Macaro, 2001), 0% to 18.15% (Rolin-Ianziti & Brownlie, 2002), 4.6% to 25.1% (De La Campa & Nassaji, 2009), to as wide as 12% to 77% (Kim & Elder, 2005), 7% to 70% (Edstrom, 2006), 10% to 90% (Liu, *et al*., 2004), and 0% to 90% (Duff & Polio, 1990). Although these studies did not examine EFL contexts (with the exception of Liu, *et al*. (2004)), similar patterns can be expected, provided teachers share their learners' L1. An inherent weakness of these studies is their tendency to ignore the pedagogical approaches and intentions of teachers, and so are claimed to be of little value for informing the debate over the efficacy of codeswitching to enhance L2 learning (Macaro, 2009; Tian & Macaro, 2012).

2.3 'Optimal use': Exploring the Pedagogical Potential of Codeswitching

Contrary to mainstream beliefs, some have viewed codeswitching favorably, recognizing its pedagogical value in facilitating L2 learning beyond exclusive L2 use (Hall and Cook, 2012), a perspective identified as the 'optimal position' (Macaro, 1997, 2001). Central to this perspective is the concept of *optimal use*, defined as 'codeswitching in broadly communicative classrooms [which] can enhance second language acquisition and/or proficiency better than second language exclusivity' (Macaro, 2009: 38). Recognition of this potential has led some researchers to compile collections of principled instructional techniques which exploit learners' L1 in classroom discourse (e.g. Atkinson, 1993; Butzkamm & Caldwell, 2009; Swain, Kirkpatrick, & Cummins, 2011). The most comprehensive of these is Butzkamm & Caldwell's (2009) volume, *The Bilingual Reform*, which contains numerous ideas and techniques for systematically integrating learners' L1s into classroom instruction and activities. Butzkamm & Caldwell (2009: 13) argue

Table 2.1. Butzkamm & Caldwell's (2009) maxims for bilingual instruction

Maxim 1	**The direct principle is a delusion. The L2 learner must build upon existing skills and knowledge acquired in and through the L1.**
Maxim 2	**Limited, incomplete understanding and blank incomprehension are a frequent source of frustration in L2 classes because monolingual ersatz-techniques of meaning-conveyance function less well than the L1.**
Maxim 3	**Poring over the meaning is likely to be less effective for meaning retention than putting the new expression to use right away**
Maxim 4	**L1 translations and explanations are more accurate than most monolingual ersatz-techniques that can be understood by the learners .**
Maxim 5	**L1 aids make it easier to conduct whole lessons in the foreign language and can promote more authentic, message-oriented communication than might be found in lessons where they are avoided. Pupils gain confidence and, seemingly paradoxically, become less dependent on their L1.**
Maxim 6	**L1 techniques allow teachers to use richer, more authentic texts sooner and to transmit larger vocabularies . The thin language soup served up to modern learners is the price paid for the L1 taboo.**
Maxim 7	**Bilingual techniques allow teachers partially to bypass the grammatical progression of textbooks. No postponement of, let's say, do-negation or the past tenses.**
Maxim 8	**It is not possible to avoid interference, but it can be greatly reduced.**
Maxim 9	**The counter-productive, haphazard use of the mother tongue, which may end up in a total breakdown of teaching, could be an unwanted side-effect of the doctrine of monolingualism.**
Maxim 10	**All newly-acquired L2 items have to sink roots in our minds which are eventually deep enough for the items to function independently of the L1.**
Maxim 11	**Direct method lessons can be fun. Monolingual explanations and paraphrases are not outlawed but will become ever more important.**

Adapted from Butzkamm & Caldwell (2009: 73-86)

that learners' L1 essentially 'lays the foundations' for learning any other language. They base their recommendations on a series of maxims, listed in **Table 2.1**, focusing on issues related to building on learners' prior L1 knowledge, focusing on meaning in authentic, message-oriented communication, enriching the learning environment with more authentic texts, etc. However, in acknowledging the importance of establishing and maintaining a 'foreign language atmosphere' by maximizing L2 use for both communicative tasks and conducting the 'normal business' of the classroom (i.e. administrative purposes and spontaneous, authentic teacher-learner interaction) (Butzkamm & Caldwell, 2009.: 31), heed is also paid to those who caution against overly enthusiastic L1 use (e.g. Cummins, 2007; Macaro, 2005; Turnbull, 2001). Indeed, as teachers' unrestrained L1 use significantly reduces L2 exposure, it is arguably 'the single biggest danger in any monolingual class' (Atkinson, 1993: 13). Furthermore, if teachers carelessly abandon the L2 whenever difficulties arise, convincing learners to accept the L2 as the predominant mode of classroom communication becomes more difficult (Littlewood, 1981 cited in Littlewood & Yu, 2011). Nevertheless, as classroom time constraints unavoidably prevent sufficient L2 exposure 'for learners to sort out the many bewildering complexities of a language by themselves'

(Butzkamm & Caldwell, 2009: 30), a balanced approach which compensates for these limitations by maximizing comprehensible L2 exposure through judicious L1 use is necessary. Disregarding the insufficient class time available for recreating naturalistic learning conditions in many EFL contexts is regarded as the 'fundamental error' made by proponents who extrapolate monolingual teaching principles from theories of acquisition (Butzkamm & Caldwell, 2009). While exposure to L2 input is unquestionably necessary, it alone does not provide the requisite conditions for learning since input must become intake (Ellis, 1994, but see also V. Cook, 2001). In other words, even if learners are exposed to massive amounts of input, they will learn very little if the input is incomprehensible. As Zhao & Macaro (2014: 4) observe, while monolingual instruction increases L2 exposure, the input learners receive may overburden their cognitive capacity to effectively grasp meaning by exceeding the limits of their prior experiential knowledge, potentially resulting in inaccurate or erroneous understandings. In light of this, it is argued that 'judicious and theoretically principled' L1 use can aid in comprehension by easing learners' cognitive burden resulting in better learning outcomes (Turnbull & Dailey-O'Cain, 2009: 5).

2.3.1 Theoretical Support for Classroom Codeswitching

According to Macaro (2009: 37-8), theories in support of the 'optimal position' come from three sources: cognitive processing theory (e.g. N. Ellis, 2005), sociocultural theory (e.g. Antón & DiCamilla, 1998; Brooks & Donato, 1994; Swain & Lapkin, 2000), and codeswitching in naturalistic environments, which he argues are essentially no different from the 'broadly communicative' classroom. These sources will be examined in turn.

As Macaro (2009) observes, cognitive processing theory (N. Ellis, 2005) posits that working memory and long term memory coordinate to allow language to be perceived, processed and stored in similar ways to other information, but more importantly that the L1 and L2 are not conceptually compartmentalized, contrary to previous claims (e.g. Weinreich's (1953 cited in Macaro, 2009) *co-ordinate bilingual model*), but rather that they share an increasingly interconnected, dynamic, conceptual network, all of which is brought to bear in the processing either language. Meanwhile, it has been proposed that, in addition to the L1 and L2 sharing a common conceptual network, learning an L2 entails 'a complete metamorphosis' of the learners' psycholinguistic system rather than 'an overlap between the two subsystems' (Jessner, 2006 cited

in Cummins, 2007: 234). Furthermore, there is empirical evidence suggesting that, given the effort involved in processing two languages, working memory in bilinguals is superior to monolinguals, with capacity increasing with L2 proficiency (Bartolotti & Marian, 2013). Along these lines, it is suggested that

> [l]earning a L2 is not just the adding of rooms to your house by building an extension in the back: it is the rebuilding of all the internal walls. Trying to put languages in a separate compartment in the mind is doomed to failure since the compartments are connected in many ways' (V. Cook, 2001: 407).

These theories effectively underpin the concept of Common Underlying Proficiency (CUP), which recognizes that interdependency across languages allows for 'the transfer of cognitive/academic or literacy-related proficiency from one language to another' (Cummins, 2007: 232). Of further relevance is the role of engaging learners' prior knowledge, which constitutes an amalgamation of all previous learning, skills and experiences 'that have shaped the learner's identity and cognitive functioning' (Cummins, 2007: 232). As Brooks-Lewis (2009: 228) points out, prior knowledge 'makes learning significant', and in the context of language learning, if the prior knowledge learners bring with them into the classroom is encoded in their L1, their engagement with it is unavoidably facilitated through their L1 (Cummins, op. cit.). By prohibiting L1 use, then, teachers are essentially depriving learners of their lifetime accumulation of knowledge and experience (Brooks-Lewis, op. cit.), 'an important cognitive tool' (Antòn & DiCamilla, 1998; Swain & Lapkin, 2000), and 'the greatest asset [they] bring to the task of foreign language learning' (Butzkamm, 2003; Butzkamm & Caldwell, 2009). Thus, what remains is a learning environment where learners' access to prior knowledge is potentially limited to what can be expressed in their L2 (Cummins, op. cit.), which may decrease opportunities for L2 cognitive and metacognitive development (Macaro, 2009: 49).

Sociocultural theory is another source in support of 'optimal use'. Under this research paradigm, learning and language development are facilitated through social interaction (Swain & Lapkin, 2000; Vygotsky, 1978). Through this perspective, learners' L1 use is viewed as 'a cognitive tool... through which learning is scaffolded' (Hall & Cook, 2012: 291) at both interpsychological (i.e. collaborative talk for explaining tasks, problem solving) and intrapsychological levels (i.e. private speech) (Antòn & DiCamilla, 1998). For example, Swain & Lapkin (2000) found

English-speaking French immersion students used their L1 in collaborative dictogloss and jigsaw tasks, to move the task along, to focus attention on form, and to initiate and maintain interpersonal interaction. Another example of this is Carless (2002, 2004) who, studying elementary school EFL learners in Hong Kong, found L1 use increased among learners in correspondence with task complexity. It is additionally suggested that L1 use may grant less proficient learners access to the 'higher-level knowledge' of more proficient learners by aiding in the maintenance of sustained interaction (Thoms, Liao & Szustak, 2005 cited in Hall & Cook, 2012: 292). Similarly, as it is now widely acknowledged that virtually all but 'the most advanced L2 learner/user' think in their L1, learners will actively use it to lighten the cognitive load during demanding learning tasks (Macaro, 2005: 68). For instance, Kern (1994), studying English-speaking learners in a college French class, noted his subjects used their L1 to reduce constraints on working memory, to avoid losing track of meaning, to consolidate meaning in long term memory, to convert input into more familiar terms, and to clarify the syntactic roles of certain lexical items (see also Macaro, 2005; Swain & Lapkin, 2000). Given the inevitability of learners utilizing their L1, with or without permission, it would make sense for teachers to tap into this behavior to maximize its potential to facilitate L2 learning.

The third source of support for the 'optimal position' comes from codeswitching in naturalistic environments (i.e. informal, non-instructional contexts), which is now considered a normal occurrence in bilingual communities worldwide (Macaro, 2005). By viewing the L2 learner as a developing bilingual, it is suggested that codeswitching in a classroom context 'can be seen as a reflection of what bi- and multilingual speakers do in everyday life' (Turnbull & Dailey-O'Cain, 2009:1). While it is questionable whether the predominantly message-oriented, naturalistic codeswitching found in bilingual communities resembles pedagogical, medium-oriented teacher codeswitching in 'broadly communicative' classrooms (Macaro, 2009: 38), in reference to his 1997 study, Macaro (2005) points out that teachers' codeswitching was, in fact, mainly for message-oriented functions (e.g. complex procedural instructions). He suggests the reason for this stems from having inadequate time to provide the amount of input modification necessary to successfully communicate complex message-oriented utterances, but avoiding input modification is precisely the reason codeswitching occurs in naturalistic settings (Macaro, 2005). If message-oriented codeswitching can take place in the classroom, there is no reason this context should be

considered any different than naturalistic codeswitching (Macaro, 2009). Excessive input modification for the sake of upholding the monolingual principle in communicative classrooms 'often results in the teacher hogging the discourse space' (Macaro, 2005: 73), which can potentially lead to their lessons becoming ironically less 'communicative' (Macaro, 2000: 184). As Macaro and Mutton (2002) found, teachers' codeswitching effectively provided learners with more 'discourse space' simply because it took less time than the alternative. This, by no means, devalues input modification, but optimal use of codeswitching represents a judgment teachers must make regarding

> the possible detrimental effects of *not* drawing the learners' attention to aspects of their first language, or not making comparisons between the first and second languages. It involves a principled decision regarding the effects of not conveying important information simply because this might be too difficult for the learners to understand in the second language – a teacher avoidance strategy. It involves decisions about the relative merits of second language input modification as opposed to activating first language connections (Macaro, 2009: 38-9).

Despite the need for these judgments to be informed, there is a considerable dearth of empirical research into the pedagogical value of codeswitching as opposed to L2 exclusivity (Macaro, 2009). Most of the observational studies investigating the functions and amounts of codeswitching described earlier (see **Section 2.2**) do very little to inform its pedagogical potential as they ignore the 'learning environment that the teacher was trying to create' (Macaro 2009: 48). To fill this gap, Macaro and his colleagues have conducted a number of studies into the effect of codeswitching on L2 learning (see, for example Hennebry, *et al*., 2013; Tian & Macaro, 2012). Among these studies, the potential for codeswitching in explicit vocabulary teaching has been a particular focal point. A selection of these studies will be examined in more detail along with reference to psycholinguistic research below.

2.3.2 Codeswitching to Assist L2 Vocabulary Learning

A number of studies have compared the effect of codeswitching against L2-only explanations on vocabulary learning, all of which found a comparative advantage in providing L1 vocabulary equivalents over L2-only strategies, especially for short-term learning (Hennebry *et al*. 2013; Lee & Macaro, 2013; Tian & Macaro, 2012; Zhao & Macaro, 2014). Tian & Macaro (2012) found

that, following listening comprehension activities, Chinese university English majors benefitted significantly from being given L1 equivalents compared to L2 exclusive explanations, regardless of proficiency level, although the effect was not durable in a delayed posttest. A similar result was found when Hennebry *et al*. (2013) measured the effect of listening followed by form-focused instruction vs. listening-only conditions with Year 9 French students in the UK. Lee & Macaro (2013) found that both young learners (12-year-olds) and adults' receptive vocabulary recall benefitted from receiving L1 (Korean) equivalents during reading comprehension tasks, though the young learners performed better than adults on delayed recognition/retention tests, suggesting an age effect on retention. Finally, Zhao & Macaro (2014) found teaching L1 equivalents after reading comprehension activities to be far superior to L2-only explanations, regardless of word type (i.e. concrete or abstract). The results of these studies are comparable to those in support the argument that L1 use for intentional vocabulary learning (i.e. through explicit form-focused instruction, memorization of word pairs, provision of marginal glosses, bilingual dictionary use) is more effective than incidental learning in which learners are expected to infer the meaning of target words embedded in context (e.g. Laufer, 2005; Mondria, 2003) or from pictures (Lotto & de Groot, 1998). Indeed, as Macaro (2009: 40) suggests, making unfamiliar vocabulary more salient in written texts (e.g. providing glosses or textual elaboration) is 'not that dissimilar to codeswitching in oral interaction.'

Incidental vocabulary learning is conceptualized as a 'byproduct' of activities without a focus on vocabulary learning while intentional vocabulary learning entails activities explicitly aimed at memorizing lexical information (Hulstijn, 2001). In studies investigating vocabulary acquisition, learners who read texts with the assistance of a bilingual dictionary (e.g. Luppescu & Day, 1993; Knight, 1994; Laufer & Girsai, 2008) or marginal L1 glosses (e.g. Hulstijn, Hollander & Greidanus, 1996; Laufer & Shmueli, 1997; Watanabe, 1997) significantly outperformed those who read without assistance, the latter (i.e. marginal glosses) being superior overall. For instance, Hulstijn *et al.* (1996) found that Dutch learners of French who were given marginal L1 glosses outperformed the other two conditions (reading-only and reading with bilingual dictionaries) in immediate tests, presumably because glosses made target words more salient to learners than in, for example, the dictionary condition where learners rarely consulted their dictionaries (Hulstijn *et al.*, 1996: 335). Hulstijn *et al.* (1996) conclude that when learners read for global

comprehension, they are less likely to invest time and effort into inferring meanings of unknown words. Nevertheless, it has been argued that while glossing yields more accurate comprehension than contextual inferencing, unless glossed lexical items are encountered multiple times, their retention is more tenuous than words which have been inferred with greater effort leading to deeper processing (Hulstijn, 2008). Codeswitching is potentially similar, though it may grant opportunities for deeper semantic processing given the potential for learners' more intense strategic reaction in discovering cross-linguistic similarities (Macaro, 2009: 47).

Despite advantages in terms of retention, inferencing from context is still dependent on learners' capacity to notice unknown words, but for learners engaged in listening for comprehension, the inherently ephemeral nature of spoken communication makes noticing and inferring the meaning of unknown words even less likely, given the unique demands placed on learners' working memory. Since less proficient learners often have more limited vocabularies and strategic repertoires for coping with these demands, the prospect of successfully identifying and inferring the meanings of novel words in the speech stream is diminished considerably (Macaro, 2014). In this way, codeswitching is seen as freeing up learners' cognitive resources to process larger chunks of input during spoken interaction (Macaro, 2005). Furthermore, as it may be beneficial to reduce or even eliminate context (e.g. memorizing L2/L1 word pairs) to give new vocabulary more direct attention (Prince, 1996; Laufer, 2003, 2005; Laufer & Shmueli, 1997), it is argued that decontextualized, form-focused treatment of new vocabulary (e.g. writing sentences with target words (Webb, 2005)) leads to better learning outcomes than activities which embed vocabulary in context (e.g. reading comprehension tasks) (e.g. Laufer, 2003; Min, 2008; Mondria, 2003). Laufer (2005: 226-7) consolidates the reasons for these arguments into four main points: (1) learners do not focus on individual word meanings provided global meaning is understood; (2) correctly inferring word meanings becomes unreliable when less than 98% of the words in the given context are known; (3) if word meanings are easily inferred from context, lack of engagement will result in less durable learning; and (4) learners, on average, do not read nearly enough material to reach optimal frequencies of exposure for incidental learning to take place (i.e. Nation & Wang's (1999) recommended 10 exposures). Many of these studies provide support for theoretical models of bilingual lexical development discussed below.

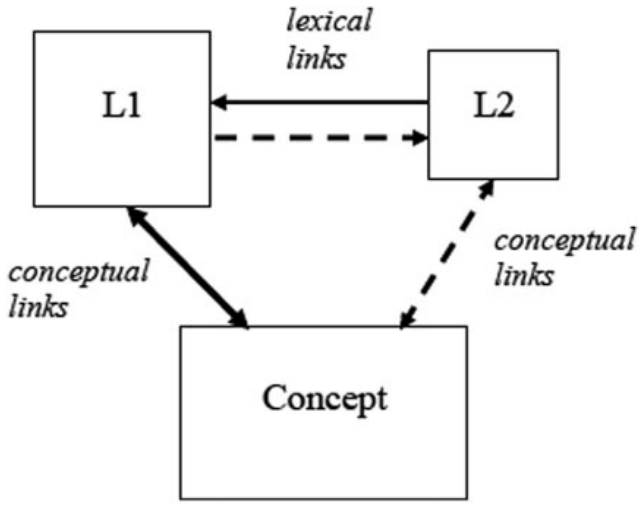

Figure 2.1. The revised hierarchical model (Kroll and Stewart 1994)

2.3.3 Lexical Representation in the Bilingual Mind

There is now a large body of psycholinguistic empirical research regarding bilingual lexical representation, which demonstrates the role of learners' L1 in facilitating L2 lexical processing (for example, Jiang, 2002, 2004a, 2004b; Kroll & Stewart, 1994; Lotto & de Groot, 1998). For adult learners who already have an established L1, the majority of conceptual knowledge they possess is derived from experiences in their L1 (Hennebry, *et al.*, 2013). A model of the bilingual lexicon that takes this (i.e. successive bilingualism) into consideration is Kroll & Stewart's (1994) Revised Hierarchical Model (RHM) (see **Figure 1**), which shows how, at initial stages of L2 learning, learners form associations between L2 and L1 words through which access to the L1 conceptual store is gained, with more direct links between L2 words and concepts being established with increasing proficiency. Given that forward translation (L1 to L2) involves accessing conceptual knowledge through a longer route than backward translation (L2 to L1), this model accounts for the apparent asymmetrical latency between the two directions (Dufour & Kroll, 1995; Kroll & Stewart, 1994). The RHM's implications for L2 pedagogy suggest that beginning learners may benefit more from L1 information while more proficient learners may not. However, a number of studies have shown that all learners appear to benefit from L1 information regardless of proficiency (e.g. Hennebry, *et al.*, 2013, Tian & Macaro, 2012). Jiang (2000) proposed a similar model detailing three stages (L1-L2 word association, L1 lemma mediation, full integration) learners go through in the process of establishing more direct

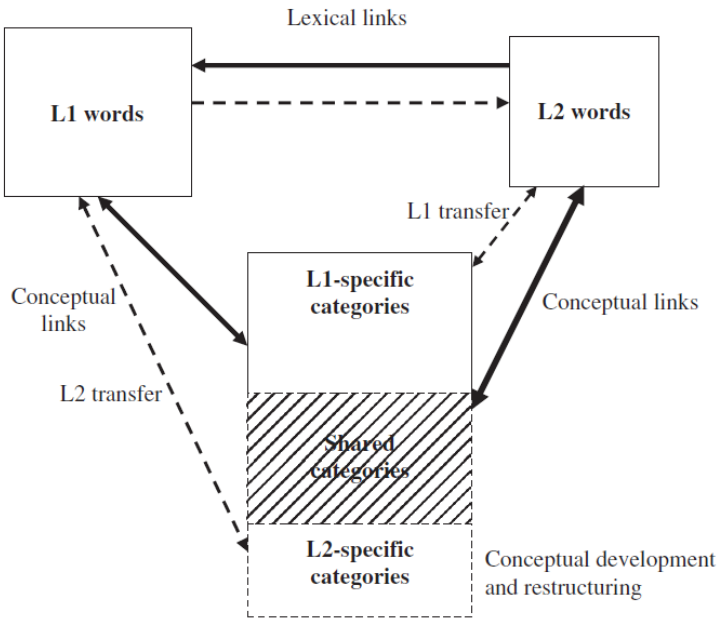

Figure 2.2. The Modified Hierarchical Model (MHM) (Pavlenko, 2009)

links to L2 words. Based on this three-stage model, Jiang (2002, 2004a, 2004b), conducting a series of lexical priming experiments with Chinese and Korean learners of English, found compelling evidence that even highly proficient learners remain at the 'L1 lemma mediation' stage in which L2 lexical representations are still dependent on shared L1 lemma (i.e. semantic and syntactic) information to access the conceptual store. In one study, Jiang (2002) found, that participants responded more quickly when given semantically similar L2 (English) word pairs sharing the same L1 (Chinese) translation (e.g. *problem* and *question* translate to *wenti*) than word pairs that did not (e.g. *interrupt* and *interfere* translate respectively to *daduan* and *ganrao*). These results were successfully replicated in the studies of Korean-English bilinguals that followed (Jiang, 2004a, 2004b). As no significant difference in response time was found for native English speakers, Jiang concluded that the learners' L1 still played a role in processing L2 lexical information. Similarly, in a recent review of event-related potential (ERP) studies, van Hell & Kroll (2013: 153) observe that co-activation of lexical level-links not only persists in highly proficient bilinguals, but that proficiency level does not necessarily determine the access route for L2 word-form-to-concept mappings. In other words, even less proficient learners can potentially access conceptual knowledge through their L2. They go on to conclude that

> ... a hallmark of proficient language use is the ability to fully exploit lexical and conceptual links depending on linguistic and contextual task demands, the difficulty of the language materials the bilingual perceives or produces, and the situational context of language learning and language use. The high end of bilingual processing may be a bilingual whose L1–L2 and L2–L1 processing is fully symmetrical' while 'L2 learners' and less proficient bilinguals' processing may be characterized by asymmetries in L1–L2 and L2–L1processing...' (van Hell & Kroll, 2013: 154).

Another critique comes from Pavlenko (2009:130) who points out how tasks using decontextualized words and/or pictures of prototypical objects coupled with serial avoidance of cross-linguistic differences in the stimulus materials has resulted in a 'somewhat skewed picture of the bilingual lexicon' which assumes a unified conceptual store. She put forward the Modified Hierarchical Model (MHM) (see **Figure 2.2**) which attempts to reconcile the RHM's inability to account for partial and/or complete lexical (non)equivalence by dividing the conceptual store into both shared and language-specific categories (Pavlenko, 2009: 147-151). Thus, it is argued that L2 instruction should take advantage of conceptual equivalence by strengthening links between L1/L2 words, draw attention to L1/L2 similarities and differences in order to facilitate 'conceptual restructuring' in the case of partial (non)equivalence, and promote the development of new L2 conceptual knowledge in the case of conceptual non-equivalence (Pavlenko, 2009: 154-5). In this way, while codeswitching is seen as essential for accurately establishing the initial form-meaning mapping of novel vocabulary, developing deeper paradigmatic and syntagmatic vocabulary knowledge is best achieved through a combination of frequent incidental exposure (Laufer, 2005; Schmitt, 2008) and more direct attention through rich, form-focused instruction (Laufer, 2005; Nation, 2001).

The importance of incidental exposure cannot be overstated given that such opportunities facilitate 'conceptual restructuring' in cases of partial (non)equivalence (Pavlenko, 2009). For example, while the Japanese language consolidates the act of visual perception into a single verb (*miru*), the concept in English is divided further to account for increasing levels of attention (i.e. see, look, watch). This case of partial (non)equivalence often leads to errors such as 'I *watch* the internet' or 'Did you *look* the baseball game yesterday?'. While Japanese learners can be made aware of these distinctions through explicit instruction, they will struggle to apply this knowledge when engaged in activities where the primary goal is communication. They will need to encounter these words frequently in various contexts over time for conceptual restructuring to

take place. This fits within the wider context provided by Nation's (2001) four-stranded approach to vocabulary teaching and learning which emphasizes the importance of meaning-focused input, meaning-focused output, language-focused learning (i.e. explicit attention to lexical items), and fluency development.

3 The Study

In light of the preceding research regarding codeswitching, and bilingual lexical development, the present study sought answers to the following research questions:

1. Is learners' receptive vocabulary learning better facilitated by giving L1 equivalents through the use of codeswitching or by maintaining monolingual instructional strategies (e.g. definitions, explanations, contextualized examples)?

2. Is codeswitching a more efficient way of teaching L2 vocabulary than exclusively monolingual strategies?

3. Is learners' retention of vocabulary meanings better facilitated by the use of codeswitching or by maintaining monolingual instructional strategies?

3.1 Design

The present study was conducted in a small, private conversation school located in a rural town in Western Japan between March and June 2016. The study adopted a quasi-experimental design with experimental and comparison groups (5 participants in each group). The experimental group received target vocabulary instruction through the use of L1 codeswitching (CS), while the comparison group received monolingual (i.e. English-only) instruction (NCS). Instruction for both groups was provided prior to a series of communicative activities designed to familiarize participants with and deepen their understanding of the target vocabulary. Following each instructional session participants were tested to measure learning outcomes, and delayed posttests were administered at 2 weeks and 8 weeks after instruction to measure retention.

As scheduling difficulties prevented organizing time outside of regularly scheduled lessons, it was necessary to conduct the intervention in place of participants' regular lessons, meaning both study groups had to be established as intact classes. Thus, although 5 students were allocated to

Week		Week 1	Week 2		Week 3		Week 4		Week 5			Week 7		Week 13
Class		VST	Session 1		Session 2		Session 3		Session 4			2-week delayed test		8-week delayed test
CS			N = 4*	Posttest 1	N = 4	Posttest 2	N = 4*	Posttest 3	N = 5*	Posttest4				
NCS	NCS1		N = 2*		N = 1				N = 1					
	NCS2		N = 1		N = 2		N = 2		N = 1					
	NCS3		N = 2		N = 1		N = 2		N = 2					

Note: * indicates an excluded participant attended the session

Figure 3.1. Intervention structure and progression

each condition, those in the CS condition represented a single class, while those in the NCS condition had to be divided across 3 separate classes. An attempt was made to consolidate the NCS condition into a single class before the intervention, but was met with resistance as most participants could not commit to the change for the entire duration of the intervention. Nevertheless, to avoid problems resulting from excessive delinquency, participants in the NCS condition were permitted to join any one of the NCS sessions if they were unable attend their regular session. Consequently, there were several occasions where only a single participant received instruction in the NCS condition because their classmate was either absent or attended a different session. In contrast, participants in the CS condition regularly received instruction as a group during the intervention (see **Figure 3.1** for a diagram of the invention structure, progression, and session attendance). It is acknowledged that this arrangement was less than ideal and represents a considerable limitation due to the disparate nature of the instruction NCS participants received compared to the CS condition. However, given the small sample size for this study and the potential for delinquency, the arrangement was necessary to ensure an adequate amount of data could be gathered without causing participants undue stress. As this issue could not be avoided, its implications will be discussed later.

3.2 Participants

The participants for this study were a small convenience sample of 10 adults (5 male and 5 female) who agreed to participate in the study. All participants were of various backgrounds

Table 3.1 Demographic and baseline test information

Demographic/baseline		L1 Codeswitching Group				L2-only Group			
		N	**%**	**Mean**	**SD**	**N**	**%**	**Mean**	**SD**
Gender	Male	3	75			2	50		
	Female	1	25			2	50		
VST scores				58.04	7.18			65.54	5.45

Notes: VST: Vocabulary size test

between the ages of 40 and 60, and regularly attended weekly lessons with the researcher over an average period of 8 years prior to this study. Based on the researcher's regular contact with the participants, their overall English proficiency level was determined to be within the intermediate to upper-intermediate range. All participants were accustomed to monolingual CLT instruction, and their capacity to handle classroom study materials at the B2 level on the CEFR scale (Council of Europe, n.d.) was well established. It should also be noted that 2 participants had to be excluded from the final results due to frequent delinquency during the intervention and delayed testing stages. Thus, only data gathered from 8 participants (4 in the CS condition and 4 in the NCS condition) were considered for further analysis.

In order to ensure that the results of the study could be reasonably attributed to the instructional intervention and not intergroup differences, the Vocabulary Size Test (VST) (Nation and Belgar, 2007) was administered one week prior to instruction. Results from an independent samples t-test (see **Table 3.1**) between the CS group (n = 4, Mean = 58.03, SD = 7.18) and the NCS group (n = 4, Mean = 65.53, SD = 5.45) indicate that, although the NCS group performed slightly better, the difference between groups in terms of prior receptive vocabulary knowledge was not statistically significant, $t(6)= -1.66$, $p > 0.05$. Thus, it is suggested that the results of subsequent tests can be reasonably attributed to the intervention.

3.3 Materials

In this section, the selection of target vocabulary words for this study, the assessment instruments used, and the assessment scoring method used will be presented.

3.3.1 Preliminary Vocabulary Test

As stated above, all participants were given the 14k bilingual (English/Japanese) version of the VST (140 multiple-choice items with 10 items per 1000 word frequency level) one week prior to the intervention. Nation (2012) strongly recommends that the bilingual version of the VST be used whenever possible as L1 equivalents make answer choices more readily comprehensible by avoiding potential complications found in the monolingual version of the test (i.e. L2 grammar). Due to schedule limitations, the test was given before or after regularly scheduled lessons. Participants were given approximately 40 minutes to complete the test, although many finished much earlier. They were encouraged to answer all the questions to the best of their ability and to guess for words they weren't sure about. The VST not only provided a baseline measurement of participants' receptive vocabulary knowledge, but also provided a source of target words for the study as will be outlined below. Since the target words were embedded within the larger 140 item VST, this helped minimize the risk of participants being sensitized to them, which therefore likely prevented them from studying the words prior to the intervention.

3.3.2 Selection of Target Words

As the VST provided a readily available indication of participants' receptive vocabulary knowledge (or lack thereof), 24 target words (see **Table 3.2** for a detailed breakdown) were selected for the study from the lesser known items on the test (very few words were completely unknown to participants). The words were then distributed into 4 sets of 6 words, each containing 3 nouns, 2 adjectives, and 1 verb. Given the limited time available for each instructional session, it was determined that keeping the number of words per session small would promote more concentrated learning and avoid putting unnecessary stress on participants. Although a few participants had provided correct answers for some of the target words on the VST, the researcher's extensive experience working with the study group gave strong reason to suspect participants' had only guessed these items correctly. This was quickly confirmed when participants were shown the words at the beginning of each session during the intervention. For instance, while 4 participants in the CS condition had answered correctly for 'communiqué' on the VST, none could produce its meaning when given the opportunity. The same was true of 'devious' for 3 participants in the NCS condition. This inconsistency could be attributed to the

Table 3.2. Target words broken down by VST level and instructional session

Word frequency level		Session		
1st 1000	figure		ruck	allege
4th 1000	allege; candid	1	hallmark	candid
5th 1000	compost		counterclaim	gauche
6th 1000	devious; malign; strangle; veer		compost	mumble
8th 1000	authentic; mumble	2	hutch	authentic
9th 1000	hallmark; regent; whim		crowbar	jovial
10th 1000	crowbar; egalitarian; peasantry; ruck		whim	veer
11th 1000	counterclaim; hutch	3	peasantry	malign
12th 1000	refectory		regent	egalitarian
13th 1000	communiqué; didactic; jovial		refectory	strangle
14th 1000	gauche	4	communiqué	didactic
			figure	devious

fact that participants' were given adequate time during the VST to consider the meanings of the answer choices, which may have made it possible to guess the answer by process of elimination without actually having known the target word. In other words, the English meanings of the incorrect answer choices may have been known by some participants making it easier to eliminate them.

3.3.3 Post-instruction Assessments and Scoring Methodology

Immediately following each session, participants were given a simple receptive test on the 6 words they had encountered, based on that used in Tian & Macaro (2012), to assess quality of learning (see **Appendix I**). The fact that the words would be tested was not disclosed to participants in the first session, but they expected the tests in the sessions that followed. These tests required participants to provide answers as short written (limited production) responses and to circle a number corresponding to the level of confidence in their answers (5 meaning very confident and 1 meaning not confident). Participants were permitted to write their answers in either English or Japanese so as to avoid any bias stemming from instructional conditions. As it is essentially impossible to guess correct answers for limited production tasks (Bachman and Palmer, 1990: 208), it was determined this format would eliminate the potential recurrence of problems experienced with the initial VST. As expected, participants in the CS condition provided all answers in Japanese. However, most of the answers given by participants in the

NCS condition were in English (with the exception of two participants who wrote some of their answers in English, Japanese, or both).

To measure participants' retention of the 24 target words, cumulative delayed posttests (see **Appendix II**) were administered at 2 weeks and 8 weeks after the last intervention session. The delayed posttests were the same format as the posttests. It should be emphasized that these tests contained items which participants had only received limited exposure to as early as 12 weeks prior to testing (in the case of the 8 week delayed test). As with the posttests, participants in the CS condition answered in the L1. Those in the NCS condition also provided most of their answers in the L1 for both delayed posttests despite having written most answers in English in the posttests. All tests were written and administered by the researcher.

Although the limited production response format of the posttests minimized the possibility for guessing, it was necessary to employ an interpretive partial-credit scoring method to address the diversity of participants' responses. The sole criterion for correctness was based on whether or not participants could demonstrate a satisfactory receptive understanding of the target words' meanings in the L1 or L2 (spelling and grammar errors were ignored as long as responses were comprehensible). Since participants could use their L1, it was necessary to consider a multitude of semantically acceptable responses as many of the target words did not share clear one-to-one L1 translations. This was particularly important for those in the NCS condition because their L1 responses were products of form-meaning mappings they had internally generated independently of the researcher. Those in the CS condition simply had to supply the translation equivalents they had been taught (although a few participants provided acceptable responses which had not been taught). In consideration of these issues, a scoring rubric was developed as follows: an answer was awarded 2 points if it demonstrated a semantically appropriate understanding of the target word's core meaning, 1 point if it failed to fully capture the target word's meaning (i.e. partially related/incomplete) or contained superfluous, unrelated words which detracted from the meaning, and no points if it was completely unrelated or left blank. With regard to the delayed tests, responses for which participants had not received instruction during the intervention (i.e. because of delinquency) were excluded from the final scores regardless of correctness.

Given the interpretive scoring system, it was necessary to test its reliability by having other qualified individuals use it to score the tests. In addition to the researcher, 3 experienced EFL teachers were recruited to score test data (2 native Japanese speakers and 1 native English speaker with near-native Japanese language proficiency and some translation experience). As a majority of the participants' answers were provided in the L1, it was imperative that the markers had a strong understanding of both languages to ensure their scoring was reliable. To guide the markers, the aforementioned scoring rubric and an answer key (see **Appendix III**) were provided with the test data. Bachman and Palmer (1990) recommend the inclusion of an answer key as it reduces the likelihood of scoring inconsistencies across different markers. The key was by no means an exhaustive list of every semantically acceptable answer, but instead was to be used as a starting point for makers to gauge their own judgments. Lists of acceptable answers were compiled by searching the target words and their synonyms using a popular online bilingual dictionary, Weblio (http://ejje.weblio.jp). The resulting Japanese terms were searched in reverse to confirm that they were indeed semantically appropriate translations. Finally, a native Japanese speaker was consulted on the consistency of the lists' answers before the final answer key was decided.

Before test data was sent to the other markers, it was modified to ensure participants' anonymity and compiled into an excel file containing separate spreadsheets for each delayed posttest data set (immediate posttest data was excluded to reduce the markers' burden). It was decided that separating the data sets in this way would help markers focus on each answer individually and prevent other answers from influencing scoring decisions. Upon retrieval of the scoring data, inter-rater reliability was calculated in SPSS using the ICC (Intra-class Correlation Coefficient). A very high degree of inter-rater reliability was found among the markers with the ICC average measure reaching .986 with a 95% confidence interval from .964 to .995, $F\ (15, 45) = 97.176$, $p < .001$. Given these figures, it was determined that the immediate posttest data could be reliably scored by the researcher without the other markers, and only the researcher's scores for all 3 data sets would be necessary for further analysis.

3.4 Procedure

While other comparable studies have utilized receptive reading and listening activities in heavily teacher-fronted arrangements to present target words imbedded in a wider context (e.g. Laufer & Schmueli, 1997; Tian & Macaro, 2012), in addition to explicit vocabulary instruction, the present study sought to expose participants to target words through collaborative activities and communicative games that encouraged them to not only read or listen to the words in context, but to produce the words and their meanings through spoken output. Therefore, a series of activities was devised to provide both receptive exposure (written and aural) and opportunities to engage in productive spoken output.

3.4.1 Intervention Schedule

The instructional intervention spanned 4 weeks with once a week, 60 minute lessons. Actual sessions were shorter, however, lasting approximately 30 to 45 minutes (participants were not given time limits for most activities). Each session was recorded with a digital video camera for later observation. As the primary goal of the study was to measure the effect of instruction and communicative interaction on learning and retention, it was hoped that inserting time at the end of lessons for unrelated activities would covertly distract participants from studying the target words outside of the classroom. To this end, participants were also forbidden from removing any materials from the classroom, taking notes, or using their dictionaries during class time. Despite these efforts, participants had to be requested more explicitly not to study the target words once it became apparent the distraction had failed. One particular participant in the CS condition protested the request as it contradicted his normal tendencies. As will be shown later, this participant ignored subsequent requests from the researcher to abstain from studying the target words outside of class, which made it necessary to exclude his delayed posttest data. This represents another critical limitation to the study, which warrants further discussion later.

3.4.2 Session Structure

Each instructional session contained two stages. The first stage was aimed at introducing the target words and allowing participants opportunities to construct initial form-meaning mappings. First, there was an introductory phase where pairs of participants were given 6 vocabulary words

Table 3.3. Examples of instruction given in CS and NCS conditions for 'refectory'

CS Condition	
R:	Refectory is like 食堂 (tr. cafeteria)
(Ss expressed surprise)	
R:	修道院とか学校の食堂みたいな. 特に修道院., (tr. like a monastery or school's cafeteria. Especially monastery.)
NCS Condition (NCS1)	
R:	Refectory is like uh a cafeteria, but very specifically it's a cafeteria in, for example, a school, or a church, monastery. Do you know what a monastery is?
NCSp1:	No--.
R:	A monastery is where monks live. So, it's a kind of monastery.
NCSp1:	So is that different from a church?
R:	Church, a church is for average people to go once a week or whenever they have services, but a monastery is where people in that particular religious institution actually live. They live there, and study or... So, yeah it's where monks live, basically. So, obviously they all eat together in the refectory or something like that.

printed on cards with pronunciations modeled by the researcher. They were then encouraged to make observations about the target words (e.g. identifying part of speech, comparing with known similar sounding words) and speculate about meanings with a partner (or the researcher if only one participant was in attendance). Following this was an instructional period in which the researcher taught the target words according to the research condition (i.e. CS or NCS). Examples of instruction can be seen in the transcription in **Table 3.3**. Next, pairs were given simplified English definitions on cards, which they matched with the target words. This was followed by a cloze-style matching activity in which pairs had to match the target words to contextually appropriate sentences printed on another set of cards. After each matching activity, participants were given time to confer with other pairs when possible (see **Appendix IV** for an example of materials). The second stage featured communicative activities encouraging the participants to collaboratively interact with the target words. With the exception of the 'story cloze' activity in Session 3, all the activities contained a communicative information 'gap', which required participants to elicit guesses from other participants by either explaining or demonstrating the target words by non-linguistic means (i.e., pictures or gesture). A list of the communicative activities and their descriptions can be seen in **Table 3.4.**

Table 3.4. Communicative activities and descriptions.

Activity	Description
Make a story	Participants had to work together to make a story or individual sentences using the target words.
Verbal guessing game 1	Participants were given a stack of cards which had the target words and some condition-specific written assistance printed on them (i.e., CS L1 equivalents vs. NCS L2 synonyms). Participants took turns drawing target word cards from the stack and had to describe the word in the L2 until another participant could successfully guess it.
Verbal guessing game 2	This activity was the same as verbal guessing game 1, but participants were given target word cards that didn't include any written assistance.
Picture/Gesture guessing game	This was a variation on the verbal guessing games, but required participants to use non-verbal strategies to elicit the target words instead. They could either draw pictures on the whiteboard, use gesture or make noises to demonstrate the target words, but were forbidden from using any spoken or written language to elicit guesses from others.
Story cloze	This activity was an extension of the sentence matching exercise in which a longer text, including all 6 target words, was written on the whiteboard and concealed prior to the session. Participants collaborated as a group to fill the gaps with the target words. To allow the participants to focus completely on the meaning of the text, the target word cards were left face-up on the desk making them readily available for reference during the task.

Although care was taken in selecting and preparing the activities for this study, the inability to pilot them due to contextual constraints (i.e. insufficient time and demographically similar participants outside the study group) necessitated a degree of latitude to address any complications once the intervention was underway. Regarding the communicative activities, it was determined some flexibility was needed in case participants lost interest or responded poorly to certain activities, in which case they could be subsequently altered or replaced. An example of such replacement was the 'make a story' activity (see **Table 3.4**). After having just learned the meaning of the target words, participants in both groups generally found it too difficult to make productive use the words in stories or even single sentences. Thus, the activity was dropped from further sessions in favor of activities that focused more on the target words' meanings.

Table 3.5. Breakdown of instructional session task progression

Session 1	**NCSs1**	**NCSs2***	**NCSs3**	**NCS Avg.**	**CS**
Introduction	02:31	01:54	01:43	**02:03**	04:00
Instruction	04:07	06:10	04:34	04:57	**01:45**
Definition matching	02:17	02:00	04:08	**02:48**	04:37
Sentence matching	10:35	07:44	14:30	**10:56**	12:38
Make a story	07:38	08:33	05:59	07:23	**05:15**
Verbal guessing game 1	06:12	03:42	07:21	05:45	**03:37**
Quiz	07:35	03:10	03:57	**04:54**	05:27
Total	40:55	33:13	42:12	38:47	**37:19**
Session 2	**NCSs1***	**NCSs2**	**NCSs3***	**NCS Avg.**	**CS**
Introduction	02:18	01:16	01:35	**01:43**	02:18
Instruction	04:46	03:51	04:17	04:18	**02:00**
Definition matching	03:39	03:43	02:55	**03:26**	03:59
Sentence matching	02:59	08:20	07:11	06:10	**05:23**
Verbal guessing game 1	05:05	03:48	02:01	**03:38**	03:50
Picture/gesture guessing game	04:59	05:40	04:45	**05:08**	08:19
Quiz	03:12	04:25	04:49	04:09	**02:14**
Total	26:58	31:03	27:33	28:31	**28:03**
Session 3	**NCS1**[1]	**NCS2**	**NCS3**	**NCS Avg.**	**CS**
Introduction		01:10	01:26	**01:18**	01:39
Instruction		03:17	04:07	03:42	**01:50**
Definition matching		02:20	01:44	02:02	**01:44**
Sentence matching		09:12	11:32	**10:22**	12:50
Picture/gesture guessing game		04:48	06:54	**05:51**	07:21
Story cloze		07:09	05:20	06:14	**01:44**
Verbal guessing game 1		06:47	03:24	**05:05**	06:58
Quiz		**[2]	03:09	03:09	**02:14**
Total		**34:43**	**37:36**	37:44	**36:20**
Session 4	**NCSs1***	**NCSs2***	**NCSs3**	**NCS Avg.**	**CS**
Introduction	02:02	01:31	01:45	**01:46**	05:26
Instruction	04:07	04:03	03:55	04:02	**02:29**
Definition matching	01:09	02:47	03:24	**02:27**	03:20
Sentence matching	07:57	06:02	08:53	07:37	**04:20**
Picture/gesture guessing game	05:52	02:00	02:08	**03:20**	07:17
Verbal guessing game 1	03:19	03:09	03:34	03:21	**02:58**
Verbal guessing game 2	05:31	01:15	02:32	**03:06**	03:31
Quiz	01:58	03:00	02:25	02:28	**01:39**
Total	31:55	23:47	28:36	**28:06**	31:00

* Single participant session

As participants' understanding of the target words and the contexts they were presented in were given the highest priority, the time allotted for each activity also required more flexibility (for a breakdown of time spent on tasks, see **Table 3.5**). The researcher allowed as much time for each activity as necessary until participants felt adequately comfortable moving on to subsequent

[1] This session was cancelled because the participant attended the NCS2 session.
[2] This session's recording stopped short of task completion due to technical issues with the recording device.

activities. For example, despite efforts to keep the definitions and sentences relatively simple, some unexpected comprehension difficulties required time for further elaboration during the definition and sentence matching activities. In such cases, support was given according to the research conditions (i.e. either L1 equivalents for CS or L2 explanations for NCS). As support was provided on an as-needed basis, variability in time spent on activities across different sessions resulted (e.g. Session 1's sentence matching took just under 8 minutes for NCS2 while it took over 14 minutes for NCS3, see **Table 3.5**). However, when such comprehension issues arose, they were generally caused by words other than the target words, and most of the other time discrepancies were related to the number of participants in a particular session (i.e. single participant sessions generally required less time for each activity).

There were some instances where participants required support during the communicative activities as well. For example, in the picture/gesture guessing game, some of the more concrete target words lent themselves very well to the activity, but the more abstract words occasionally proved too challenging for some participants, requiring assistance from the researcher in such cases. It is acknowledged that the differences in time spent on the activities, the variability of the activities themselves, and the variability in the amount of extra assistance provided by the researcher in different sessions may all have had an influence on participants' learning or lack thereof. These issues will be addressed in more detail when examining the results of the study.

4 Results

In this section, results from all tests given during and following the instructional intervention will be presented and analyzed along with an analysis of the effect of instructional time on learning.

4.1 CS vs. NCS for Vocabulary Learning

The first research question aimed to investigate whether or not codeswitching to teach target vocabulary facilitated learning better than monolingual strategies. Here, successful learning is defined as participants' ability to accurately recall target word meanings on immediate posttests. Descriptive statistics for pre, post, and the two delayed posttests can be seen in **Table 4.1**. Paired samples t-tests were conducted in SPSS to compare the pretest and combined posttest results (see

Table 4.2). The results for the CS condition showed that significant gains had been made from pretest (M = 32.29, SD = 7.88) to posttests (M = 95.83, SD = 2.05), t(3) = 18.91, p < .005 with a very large effect size, *d* = 9.44. Significant gains were also made in the NCS condition from pretest (M = 29.16, SD = 8.33) to posttest (M = 89.41, SD = 12.70), t(3) = 6.68, p< .05, with a very large effect size, *d* = 3.34, though these gains were somewhat less than the CS condition. Although the CS group tended to outscore the NCS group on the posttests (with the exception of posttest 4), results from an independent samples t-test found that the difference was not statistically significant, t (6) = .99, p > .05 (see **Table 4.5**). A second independent samples t-test was conducted to compare between-test gains for CS (M = 63.54, SD = 6.72) and NCS (M = 60.24, SD = 18.03) conditions, but again the results were not statistically significant, t (6) = .34, p > .05 (see **Table 4.6**). It should be noted that despite the lack of a significant difference between groups, there was still greater inconsistency (i.e. a larger standard deviation) among NCS participants, which gives cause for concern and will be examined further below.

4.2 Efficiency of CS vs. NCS Instruction

The second research question considered which mode of instruction (MoI) was a more efficient (i.e. took the least amount of time to effectively convey target word meanings). To determine this, the time taken for providing target vocabulary instruction was measured for each target word in each session and compared against participants' posttest scores. Times for each instructional instance were measured twice using a smart phone stopwatch application and averaged together to produce final results (see **Table 4.3)**. On average, it took the researcher over twice as much time to teach the vocabulary in the NCS condition (M = 41.05s, SD = 17.9) compared to the CS condition (M = 18.29s, SD = 9.88). Results from an independent samples t-test show the this difference was statistically significant, t (46) = -5.45, p < .005, with a large effect size, *d* =1.57. Pearson's correlation coefficient was used to test for a linear relationship between instruction time and posttest performance for each target word. Results found a moderate negative correlation between instruction time and average posttest performance, r (46) = -.35, p < .05. In other words, although the NCS group generally received more instruction time, this did not necessarily lead to better posttest performance for all participants. Two of the less proficient participants (NCSp2, and to a lesser degree, NCSp4), scored lower than their peers (see **Table**

Table 4.1. Descriptive statistics comparing vocabulary learning (posttests) and retention (delayed tests) for CS and NCS conditions

		Pretest	Posttest 1	Posttest 2	Posttest 3	Posttest 4	Combined Posttests	2-Week Delayed Test		8-week Delayed Test	
CS	Mean	32.29	86.11	97.92	100	97.92	95.83	43.92	31.33	29.52	25.66
	SD	***7.88***	***4.82***	***4.16***	***0***	***4.16***	***2.05***	***25.33***	***5.68***	***8.74***	***4.04***
	Range	16.67	8.34	8.33	0	8.33	4.17	56.25	11	20.84	7
	Max	37.5	91.67	100	100	100	97.92	81.25	36	41.67	28
	N	4	3	4	3	4	4	4	3*	4	3*
NCS	Mean	29.16	83.33	88.89	83.34	100	89.41	41.32		43.92	
	SD	***8.33***	***18***	***9.62***	***28.05***	***0***	***12.70***	***10.42***		***11.83***	
	Range	16.66	**41.67**	16.67	**58.33**	0	**27.09**	20.83		28.47	
	Max	33.33	100	100	100	100	97.92	50		56.25	
	N	4	4	3	4	4	4	4		4	

Note: * indicates the exclusion of data from the participant in the CS condition who cheated.

Table 4.2. Paired samples t-test results for CS and NCS groups

		Between test time comparisons				
		MD	**SD**	**SE**	**Sig. (2-tailed)**	**Effect size (Cohen's *d*)**
CS	**Pretest vs. Posttests**	**63.54**	**6.72**	**3.36**	**.000**	**9.46**
N = 4	Pretest vs. 2-week delayed	11.63	26.26	13.13	.441	
	Pretest vs. 8-week delayed	-2.78	12.38	6.19	.684	
	Posttests vs. 2-week delayed	-51.91	26.86	13.43	.031	-1.93
	Posttests vs. 8-week delayed	-66.32	10.47	5.23	.001	-6.33
	2-week vs. 8-week delayed	-14.41	16.87	8.44	.186	
CS*	**Pretest vs. 2-week delayed**	-0.46	12.53	7.23	.955	
N = 3	**Pretest vs. 8-week delayed**	-6.48	12.14	7.01	.453	
	Posttests vs. 2-week delayed	**-65.05**	**6.82**	**3.94**	**.004**	**-9.53**
	Posttests vs. 8-week delayed	**-71.06**	**5.40**	**3.12**	**.002**	**-13.16**
	2-week vs. 8-week delayed	**-6.02**	**2.12**	**1.22**	**.039**	**-2.84**
NCS	**Pretest vs. Posttests**	**60.25**	**18.03**	**9.02**	**.007**	**3.34**
N = 4	**Pretest vs. 2-week delayed**	12.16	16.56	8.28	.238	
	Pretest vs. 8-week delayed	14.76	15.18	7.59	.147	
	Posttests vs. 2-week delayed	**-48.09**	**9.17**	**4.58**	**.002**	**-5.27**
	Posttests vs. 8-week delayed	**-45.49**	**19.48**	**9.74**	**.019**	**-2.34**
	2-week vs. 8 week-delayed	2.60	11.20	5.60	.674	

Note: * indicates the exclusion of data from the participant in the CS condition who cheated.

4.4) despite having given the impression they understood explanations. For example, despite having taken nearly 4 times longer than the CS condition to explain target vocabulary words to NCSp2 in session 1, her posttest score was only 58.3%. Upon closer examination, scores for the words which required longer explanations appear to suffer more than those taking less time. For example, a 98.3s explanation for 'gauche' resulted in a score of '0' while explanations for 'hallmark' (63.5s), 'counterclaim' (92.5s) and 'allege' (82.9s) only resulted in partially correct scores of '1'. This is particularly alarming as NCSp2 was the only participant in this session, and was similarly given numerous opportunities to clarify her understanding of the target words. Some potential reasons for her difficulties will be considered later. Due to the CS group consistently exhibiting better posttest results despite the shorter instruction time they received, it can be argued that codeswitching was a more efficient form of instruction than

4.3 CS vs. NCS for Vocabulary Retention

The third question concerned what impact MoI had on retention of target vocabulary words. When scoring the 2-week delayed tests, it was discovered that one participant in the CS group

Table 4.3. Breakdown of target word instruction times (in seconds)

Session	Target word	NCSs1	NCSs2	NCSs3	NCS Avg.	CS
Session 1	**ruck**	12.1	17.3	**19.1**	16.2	13.1
	hallmark	**80**	63.5	58.6	67.4	12.5
	counterclaim	21.5	**92.5**	42.4	52.1	34.6**
	allege	39.3	**82.9**	47.3	56.5	14.6
	candid	24.4	16.1	**40.5**	27	16.8
	gauche	47.3	**98.3**	55.3	67	12.7
	Total	224.6	**370.6***	263.2	286	104.3
Session 2	**compost**	**85.2**	70.2	67.8	74.4	25.4
	hutch	16.3	24.2	**25**	21.8	8
	crowbar	39.2	45	**64.8**	49.7	21
	mumble	39.7	34.8	**42.9**	39.1	12.5
	authentic	**96**	38	23.9	52.6	34.8
	jovial	8.4	16.3	**30.7**	18.5	9
	Total	**284.8***	228.5	255.1*	256.1	110.7
Session 3	**whim**		**56.3**	52.5	54.4	14
	peasantry		19.5	**41.3**	30.4	11.7
	regent		**46.2**	32.5	39.35	43.1**
	veer		**27.1**	21.8	24.45	11.2
	malign		8.5	**21.2**	14.85	8.7
	egalitarian		39.5	**64.1**	51.8	21.7
	Total		197.1	233.4	215.25	110.4
Session 4	**refectory**	68	40.2	**74**	60.7	29.5
	communiqué	25.1	**43.1**	24.5	30.9	19.1
	figure	50.4	46.7	**61.9**	53	10.1
	strangle	14.6	**50.5**	25.7	30.3	7.8
	didactic	**26**	17.8	16.6	20.1	**31.4****
	devious	27.6	**39**	31.6	32.7	15.7
	Total	211.7*	**237.3***	234	227.8	113.6

Notes: * single participant sessions; ** participants had difficulties understanding the L1 equivalent.

had scored considerably higher (81.3%) than his peers (25 - 36%). As this participant had been outspoken against the prohibition on studying the target words outside of class, he was immediately suspected of violating the rule. When pressed, he reluctantly admitted to having studied outside of class time despite the researcher's repeated verbal and written requests. As the goal of the study was to examine the effect of *instruction*, this participant's data was excluded from further analysis regarding retention. Nevertheless, his data does present some interesting implications, which will be discussed later.

A series of paired samples t-tests were conducted to measure retention from posttest to 2-week and 8-week delayed posttests (see **Table 4.2**). The CS condition showed a significant decrease from posttest (M = 96.52, SD = 1.84) to 2-week delayed posttest (M = 31.48, SD = 5.78), t (2) = -16.52, $p < .005$, and to 8-week delayed posttest (M = 25.46, SD = 4.01), t (2) = -22.78, $p < .005$,

Table 4.4. Instruction and activity time vs. posttest scores by participant

						Immediate Posttest Scores (percentages)								
		Time (in seconds)				Codeswitching				No Codeswitching				
		Instruction	**Phase 1**	**Phase 2**	**Total**	**CSp1**	**CSp2**	**CSp3**	**CSp4**	**NCSp1**	**NCSp2**	**NCSp3**	**NCSp4**	**Averages**
	CS	**104.3**	**1380**	**532**	**1912**	**83.3**	**91.7**	**83.3**	**					**86.11**
	NCS1	224.6	1170	830	2000					91.67				91.67
Session 1	**NCS2***	**370.6**	**1068**	**735**	**1803**						**58.33**			**58.33**
	NCS3	263.2	1495	800	2295							100	83.33	91.67
	NCS Avg.	286.1	1244.3	788.3	2033									83.33
	CS	110.7	820	729	1549	100	100	100	91.7					97.92
	NCS1*	284.8	822	604	1426					100				100
Session 2	**NCS2**	228.5	1030	568	1598						83.33			83.33
	NCS3*	255.1	958	406	1364							**	83.33	83.33
	NCS Avg.	256.1	936.7	526	1463									88.89
	CS	110.4	1038	963	2001	100	100	**	100					100
	NCS1	**	**	**	**									
Session 3	**NCS2**	**197.1**	**959**	**1124**	**2083**					**100**	**41.67**			**70.835**
	NCS3	233.4	1129	938	2067							91.67	100	95.835
	NCS Avg.	215.25	1044	1031	2075									83.33
	CS	113.6	935	826	1761	93.8	100	100	100					97.92
	NCS1*	211.7	915	882	1797					100				100
Session 4	**NSC2***	237.3	863	384	1247						100			100
	NSC3	234	1077	494	1571							100	100	100
	NCS Avg.	227.8	951.7	586.7	1538									100

Notes: * single participant session; ** participant was absent

Table 4.5. CS vs. NCS independent samples t-test for between group differences

		Comparison between groups			
		MD	**SE**	**Sig. (2-tailed)**	**Effect size (Cohen's *d*)**
Pretest	**CS vs. NCS**	3.13	5.74	.606	
Posttests	**CS vs. NCS**	6.42	6.43	.357	
2-week delayed	**CS* vs. NCS**	-9.84	6.77	.206	
	CS vs. NCS	2.60	13.69	.856	
8-week delayed	**CS* vs. NCS**	**-18.46**	**7.26**	**.052**	**1.92**
	CS vs. NCS	-14.41	7.35	.098	

Note: * indicates the exclusion of data from the participant in the CS condition who cheated.

with very large effect sizes (-9.53 and -13.16, respectively). Results from another paired samples t-test also indicated a statistically significant decrease between the 2-week and 8-week delayed posttests, t (2) = -4.92, $p < .05$, with a very large effect size, $d = -2.84$. Similarly, for the NCS group, there were significant decreases from posttest (M = 89.41, SD = 12.69) to 2-week delayed (M = 41.32, SD = 10.42), t (3) = -10.49, $p < .005$, and to 8-week delayed posttests (M = 43.92, SD = 11.83), t (3) = -4.67, $p < .05$, with very large effect sizes (-5.27 and -2.34, respectively). However, somewhat surprisingly, there was a slight, though insignificant, increase between the 2-week and 8-week delayed posttests, t (3) = .46, $p > .05$. This increase was due to two participants (NCSp2 and NCp4) actually scoring over 15% and 6% higher on the 8-week delayed test than on the 2-week delayed test. The potential reasons for this will be considered in the discussion section.

The above results show considerable attrition took place for both groups, but the NCS group managed to retain the vocabulary words considerably better than the CS group, whose participants had forgotten over 70% of the target words 8 weeks after the intervention. To examine this further, independent samples t-tests were conducted for both delayed tests (see **Table 4.5**). Although not significant at 2 weeks, t (5) = -1.45, $p > .05$, the difference between groups narrowly reached statistical significance at 8-weeks, t (5) = -2.54, $p \leq .05$, with a large effect size, $d = 1.92$. Thus it can be suggested that overall, the NCS group retained significantly more of the target vocabulary 2 weeks after the intervention and the retention was durable up to 8 weeks.

Table 4.6. CS vs. NCS independent samples t-test results for between-test time differences

		Comparison between groups			
		MD	**SE**	**Sig. (2-tailed)**	**Effect size (Cohen's *d*)**
Pretest vs. Posttests	**CS vs. NCS**	3.30	9.62	.744	
Pretest vs. 2-week delayed	**CS* vs. NCS**	-12.62	11.52	.32	
	CS vs. NCS	-.52	15.53	.97	
Pretest vs. 8-week delayed	**CS* vs. NCS**	-21.24	10.73	.105	
	CS vs. NCS	-17.53	9.79	.124	
Posttests vs. 2-week delayed	**CS* vs. NCS**	**-16.96**	**6.35**	**.044**	**2.04**
	CS vs. NCS	-3.82	14.19	.797	
Posttests vs. 8-week delayed	**CS* vs. NCS**	**-25.58**	**11.82**	**.083**	
	CS vs. NCS	-20.83	11.06	.109	
2-week vs. 8-week delayed	**CS* vs. NCS**	-8.61	6.71	.256	
	CS vs. NCS	-17.00	10.13	.144	

Note: * indicates the exclusion of data from the participant in the CS condition who cheated.

To compare the magnitude of between-test time vocabulary attrition, independent samples t-tests were conducted (see **Table 4.6**). Results show a statistically significant difference between the CS (M = -65.04, SD = 6.81) and NCS conditions (-48.09, SD = 9.16) from posttest to 2-week delayed posttest, t (5) = -2.67, $p < .05$, with a very large effect size, $d = 2.04$. However, although the mean difference was greater between CS (M = -71.06, SD = 5.4) and NCS conditions (M = -45.48, SD = 19.47), results did not reach statistical significance when comparing posttest and 8-week delayed posttest performance, t (5) = -2.16, $p > .05$. It was also found that though the CS group's retention between delayed posttests was worse than the NCS group's, this difference was not significant, t (5) = -1.28, $p > .05$. In other words, while the CS group experienced a much higher rate of attrition than the NCS group after 2 weeks, at 8 weeks this trend slowed down and there was too much inconsistency among NCS participants to make null rejection possible.

Finally, a series of t-tests were administered to measure the overall impact of the intervention by comparing pretest data against data from both delayed posttests. Results suggested CS participants actually performed better on the pretest (M = 31.94, SD = 9.62) than both 2-week (M = 31.48, SD = 5.78) and 8-week (M = 25.46, SD = 4.01) delayed tests, t (2) = .06, $p > 0.5$; t (2) = -.93, $p > .05$. These results seemingly confirm that participants had indeed guessed words

correctly for which they had little or no knowledge. In contrast, results from paired samples t-tests (see **Table 4.2**) show that the NCS condition exhibited modest gains from pretest (M = 29.16, SD = 8.33) to 2-week (M = 41.32, SD = 10.41, and 8-week (M = 43.92, SD = 11.82) delayed posttests, but these differences were not significant, $t(3) = 1.47$, $p > .05$; $t(3) = 1.94$, $p > .05$. However, taking into consideration the CS group's outcomes, it is possible that the NCS group's gains could have been more had the pretests more effectively limited the influence of guessing. Unfortunately, due to the limitations posed by the pretest lacking the same format as the posttests, further exploration of this issue is difficult.

4.5 Summary of Results

In consideration of the various data gathered during this study, the results can be summed up in the following points:

1. Codeswitching produced more consistent and reliable learning of the target words than L2 exclusive strategies, regardless of L2 proficiency level.
2. Codeswitching was significantly more efficient than L2 exclusive strategies in terms of time taken promoting accurate responses on posttests.
3. Codeswitching advantages are significantly diminished over time while L2 exclusive instruction resulted in more durable, though less accurate, retention of the target words.

5 Discussion

The potential reasons for the above outcomes will now be considered within the context of relevant literature reviewed earlier.

5.1 CS vs. NCS for Vocabulary Learning

The first research question sought to investigate which MoI (i.e. CS or NCS) resulted in better receptive target vocabulary learning. Although both conditions demonstrated significant improvement compared with the pretest, it can be suggested that the participants under the CS condition outperformed those under the NCS condition. This finding echoes those in similar studies (Hennebry *et al.*, 2013; Lee & Macaro, 2013; Tian & Macaro, 2012; Zhao & Macaro, 2014). Although the results from the independent samples t-test did not indicate a statistically

significant difference between conditions on the posttests, it is difficult to ignore the intragroup consistency with which participants in the CS condition responded to posttest items compared to the NCS condition (i.e. the standard deviation was much smaller). The reason for this consistency is obvious, as participants simply had to provide the L1 equivalent they had been taught. However, that some of the NCS participants' answers were even slightly less accurate than the CS condition is concerning. That is, the fact that even a small potential exists for misunderstanding L2 exclusive explanations is troubling for any language teacher, especially when learners can seemingly give the appearance of having understood correctly when in reality they have not (see discussion of NCSp2 in **Section 4.2**).

It is apparent that in the case of the NCS condition, participants' overall proficiency had an impact on their ability to process target word explanations. As participants NCSp1 and NCSp3 were judged to have higher proficiency, it is possible that this aided in their capacity to better process explanations leading to posttest performance comparable to that of the CS condition. Additionally, it has been suggested that more proficient learners employ a wider range of metacognitive strategies to process L2 speech (e.g. O'Malley, Chamot & Küpper, 1989; Vandergrift, 2003). Although previous studies have focused on learners' internal strategies for listening comprehension tasks, it can be similarly argued that both NCSp1 and NCSp3 employed interactional metacognitive strategies by actively engaging with the researcher to negotiate the meaning of the target words through further inquiry, and comprehension checks. For example, in the first session for NCS3 there was an episode during the 'sentence matching' activity in which NCSp3 persisted until negotiation of a comprehensible explanation for the sentence corresponding to 'hallmark' was successfully reached. This effort paid off as his posttest response reflected the intended abstract meaning (i.e. 'distinctive features') despite his classmate, NCSp4, opting to answer with the unintended concrete meaning (i.e. 'stamp, show quality' [sic]). Furthermore, while the higher proficiency participants provided nearly all their responses in English, the lower proficiency of NCSp2 was reflected in over 80% of her posttest responses, which were given in Japanese or included Japanese in addition to English.

It has been argued that lower proficiency learners rely more heavily on their L1 to facilitate L2 processing than more proficient learners (e.g. Kroll & Stewart, 1994, also see **Section 2.3.2**), and

the narrow comprehension strategies they apply (e.g. searching for cognates) rarely lead to accurate understandings even when input has been pre-modified to ease comprehension and opportunities for clarification are given (Macaro, 2014). For example, learners often demonstrate a propensity for confusing novel words with phonologically similar ones (Laufer, 2005). For instance, CSp2 confused 'ruck' with 'lack', CSp4 confused 'hutch' with 'hatch', and NCSp3 confused 'peasantry' with 'pleasantry'. This, combined with other poor strategies, can potentially lead less proficient learners to inaccurately infer meanings due to their limited understanding of the context provided through L2 explanations. The potential for misunderstanding is further exacerbated by overloading less proficient learners with contextual information that exceeds their cognitive and metacognitive resources (Macaro, 2005). Indeed, too much context can be a source of distraction in which learners are inhibited from focusing on the meanings of individual words when priority is given to global comprehension (Laufer, 2005; Laufer & Shmueli, 1997). Similar to inferring the meaning unknown words from written context, it is likely the participants in the NCS condition were actively trying to infer L1 meanings from the spoken L2 context they received regardless of proficiency. Nevertheless, as the context provided to learners in the present study was primarily through spoken discourse, comprehension is potentially further constrained by the capacity limitations of learners' working memory, which is arguably limited by L2 proficiency (Bartolotti & Marian, 2013). Given the cognitively demanding nature of processing L2 spoken discourse, it is possible that the explanations may have exceeded the linguistic and extra-linguistic resources of some NCS participants, which resulted in inaccurate or erroneous understandings (Zhao & Macaro, 2014). This is apparent given the correlation found earlier in which length of explanations had a negative impact on posttest performance (see **Section 4.2**).

Out of all the NCS participants, NCSp2 exhibited the most difficulty in regards to understanding the L2 explanations, particularly in the first and third sessions. For instance, the lengthy explanation for 'counterclaim' in NCS2 (over 90 seconds) contained a considerable amount of problematic legal terminology (e.g. court, plaintiff, defendant), which led NCSp2 to make several attempts to negotiate meaning through requests for clarification and comprehension checks (similar to NCSp1 and NCSp3's behavior), but this only resulted in even more unfamiliar context for her to grapple with. Thus, unable to cope, she acquiesced to the most easily

accessible meaning, an inaccurate, though phonologically similar, L1 cognate of 'claim' (*kureemu wo iu* tr. to make a complaint) on the posttest. This tendency echoes that found in Macaro's (2014) study of learners' strategies for coping with L2 explanations. Had she been made aware of partial (non)equivalence (Pavlenko, 2009) of this cognate, she may have made more effort to negotiate meaning. Additionally, for 'gauche', the researcher offered Mr. Bean as an example when it became clear that NCSp2 could not understand 'socially awkward' and 'clumsy'. Mr. Bean is well known in Japan, and it was felt the participant's prior knowledge would benefit her understanding. Despite the intended focus being Mr. Bean's socially awkward behavior, NCSp2 ignored this focus by drawing on her own conceptual knowledge, leading to her inaccurate response, *fuugawarina* (tr. eccentric, strange). It should be noted that although corresponding materials (e.g. cloze sentence cards) also emphasized the behavioral aspect of 'gauche', this did not influence NCSp2 to rectify her initial misunderstanding. Consequently and more troubling is that this misunderstanding became durable as indicated in both delayed posttests (though *fuugawarina* was provided for 'strangle' on the 2-week delayed test). These episodes illustrate the inherent risk in allowing potentially incorrect inferencing to go unchecked (e.g. Hulstijn, 2008). Furthermore, no prompt signaling the need for the corrective feedback was available since NCSp2 gave the impression of having understood these explanations.

Further evidence of NCSp2's struggle can be found in her performance on the third posttest (scoring 41.67%) where she had correctly recalled but mistakenly interchanged target word meanings for 'whim' (*totsuzen omoitatsu* tr. suddenly decide to do) and 'malign' (very bad). The lengthy explanation for 'whim' (approx. 1 minute) may have forced her to prioritize global comprehension over attention to form while the brief explanation for 'malign' (8.5 seconds) did not allow enough cognitive engagement to facilitate the development of a strong form-meaning connection (Laufer, 2005; Schmitt, 2008). Additional interference may have been caused by participants focusing too much on contextual information used to introduce the target words instead of describing the actual meanings during communicative guessing activities (e.g. 'verbal guessing game') in this session. Specifically, NCSp1 described 'malign' as sounding 'like the name of a female singer' or using the context of mass shootings in the US instead of describing it as 'evil' or 'very bad' as had been intended. Since accurate initial form-meaning mapping can be inhibited by either providing excessive or insufficient context, which may result in confusion

and/or potentially durable misunderstandings, there is a strong argument in favor of ensuring the accuracy of initial understandings by immediately establishing a connection between novel L2 words and the learners' L1 through codeswitching so more attention can be devoted to the form and usage of the word.

In contrast, the only examples of poor posttest performance in the CS condition come from the first instructional session. One possible explanation for this is that participants were not informed they would be tested immediately following instruction, which may have caused them to approach the target words with less intention to learn. In all but the first session, the CS condition participants received perfect or near perfect scores (91-100%) while NCS participants scored as low as 41% (NCSp2 in session 3). Even though NCS participants were aware they would be tested after the first session, the results of some did not mirror the uniform increase in performance exhibited by CS participants (see previous discussion of NCSp2's difficulties). Once CS participants became aware of the posttests, they approached the subsequent sessions with a significantly stronger intent to learn the target words. This became clear when many participants, no doubt aware of their memory limitations, asked repeatedly for L1 equivalents during the instruction phase, and became visibly anxious to begin the tests towards the end of sessions, in some cases pressing the researcher to quickly pass out the test papers. However, it has been argued that the quality and frequency of lexical processing activities are far more important in influencing learning and retention than learners' being informed of post-instruction assessment or whether they processed lexical information with or without intent to learn (Hulstijn, 2001). In other words, the activities participants engaged in may have been a greater determinant of whether or not they learned the target words than having prior knowledge of being tested following the activities.

Following from the discussion above, a potential mismatch between activities used and the intended learning outcome is another potential explanation for poor CS posttest performance. While many of the activities were designed to produce the target words alone (e.g. 'picture guessing game') or within a provided context (e.g. 'cloze sentence matching', 'story cloze'), the inclusion of the production-oriented 'make a story' activity presented participants in both conditions with an extremely demanding cognitive challenge, which many of them struggled

with. If all the activities used in the sessions had uniformly focused on developing receptive vocabulary knowledge, there may have been better receptive learning outcomes (e.g. Mondria & Wiersma, 2004). However, there have been studies that demonstrated a positive effect on receptive knowledge despite engaging in productive tasks (i.e. sentence writing), though these gains depended on whether participants received adequate time for the task or not (Webb, 2005, 2009). While adequate time was given for other activities, in the case of the 'make a story' activity, time was insufficient for participants to create sentences with the target words, especially after having just learned their meanings. As Schmitt (2008: 345) suggests, reliable productive mastery of new vocabulary is not sufficiently predicated on receptive exposure alone. Moreover, as participants were not allowed to write anything down during the activity, the cognitive load may have been too heavy to make meaningful productive use of the target words. Nevertheless, it can be argued that by allowing CS participants to make direct L2-L1 connections through codeswitching, they were better equipped to take advantage of the other subsequent activities than those in the NCS condition who had less stable understandings.

5.2 Efficiency of CS vs. NCS Instruction

The second research question concerned which MoI was more efficient in teaching the target words. It clearly took significantly less time to teach the target words in the CS condition while at the same time promoting consistently more accurate responses compared to the NCS condition (see **Table 4.3**). On average it took over twice as long to explain the target words in the NCS condition, and there is evidence that the longer the explanation took, the less likely the participant would reach an accurate understanding (see **Section 4.2**). As can be seen from NCSp2's results in session one, her posttest score was the lowest despite having received instruction 4 times longer than the CS condition in a session by herself.

One possible explanation for the superior efficiency in the CS condition concerns the way in which participants in different conditions processed and recalled the target words. As indicated by Kroll & Stewart's (1994) RHM, the route for processing L2 exclusive explanations is potentially longer and more complex than being provided L1 translation equivalents. While participants in the CS condition could immediately establish direct links between the target words and their L1 lexical networks through the researcher's use of codeswitching, participants

in the NCS condition not only had to comprehend the explanations and integrate the lexical information, but also match it to prior conceptual knowledge before they could infer the target word meanings (Zhao & Macaro, 2014, but see also **Section 2.3.2**). In other words, if the initial explanation is not sufficiently comprehensible to learners, more time must be taken for input modification, but at the same time this runs the risk of potentially increasing learners' cognitive burden by forcing them to process meaning from increasingly elaborate contexts. Should L2 explanations exceed learners' processing capacity, they are more likely to show preference for more readily accessible meanings derived from their own conceptual store or whatever meanings they successfully extract from the speech stream. This relates back to the earlier discussion of NCSp2's preference for an inaccurate L1 cognate when the explanation for 'counterclaim' proved too difficult to process. This corresponds with another possible explanation related to word-type effects on processing (i.e. the concreteness/ abstractness of target words).

Regardless of how concrete or abstract a word was, the provision of L1 equivalents made meanings immediately accessible provided they were known to participants. Indeed, learners' L1 knowledge should not be taken for granted as most CS participants did not immediately understand the equivalents for 'counterclaim', 'regent' and 'didactic', which required more time for elaboration, usually by referring to the words' written forms. On the other hand, for participants in the NCS condition, while it typically took less time to effectively explain more concrete/easily visualized words or words with easily accessible L2 synonyms (e.g. 'ruck/crowd', 'jovial/happy; friendly'), some abstract words took considerably longer for some participants (e.g. the explanation for 'ruck' took an average of 16 seconds while 'hallmark' took as long as 80 seconds), though there were some exceptions (e.g. 'compost' and 'crowbar' took as much as 85 and 64 seconds, respectively). As the set of target words in the first session contained more abstract words than subsequent sessions (e.g. hallmark, counterclaim, allege), this posed a cumulatively steeper cognitive challenge for participants in the NCS condition than the CS condition. For instance, although instruction and materials attempted to focus on the abstract meaning of 'hallmark' (i.e. a distinctive characteristic/feature), because the concrete meaning may have been more accessible (i.e. Japan has a culture of using seals and stamps for official purposes), half of the NCS participants' posttest responses reflected the concrete meaning (i.e. a seal certifying quality). In other words, the processing demands for the abstract meaning may

have triggered participants' recourse to the more familiar concrete meaning despite considerable amounts of time and steps taken to facilitate learning of the abstract meaning (i.e. through input modification). This trend persisted on the delayed tests, as well, where only NCSp1 had accurately retained the abstract meaning in English while the others' answers, given in Japanese, only partially captured the concrete meaning (e.g. NCSp2 wrote *in wo tsukeru* tr. to mark; NCSp3 wrote *shoninin* tr. seal of approval; NCSp4 wrote *insho* tr. seal). These results are discouraging since it appears well intentioned efforts to facilitate better learning by providing more context and/or alternative explanations may actually be obstructing or distracting learners from accurately grasping intended meanings. Thus, given the likelihood for misunderstandings to increase concurrently with the time and elaborative complexity of L2 explanations, it seems entirely logical to take the shortest, most efficient route to promote reliable understanding through codeswitching.

5.3 CS vs. NCS for Vocabulary Retention

Lastly, the third research question concerned which MoI facilitated better target word retention. As outlined in the results, while both groups experienced significant attrition overall, CS participants forgot significantly more of the target words (70%), as was evident from the t-tests comparing between-test time effects. One possible explanation for this comes from the level of engagement participants had with the target words. Given that the CS condition provided participants with easily accessible meanings through L1 codeswitching, this required considerably less time and effort on their part to process the lexical information in comparison to the NCS condition. Thus, although the target word meanings were more easily accessible through codeswitching, it is likely they were also more easily forgotten (e.g. Hulstijn, 2008; Laufer, 2005). Furthermore, it appears that because CS participants were given L1 equivalents for target words instead of having to independently infer their L1 meanings, they were less confident than the NCS participants in responding to target words which could not be recalled exactly as they had been taught. This is evident given that CS participants were less likely to provide partially correct or incorrect responses while the NCS participants took more risks by providing whatever information they could recall. Interestingly, while participants in the NCS condition provided nearly all their responses in English on the immediate posttests (with the exception of NCSp2), the majority of their responses on the delayed posttests were provided in

Japanese. This is not surprising in light of psycholinguistic research (e.g. Jiang, 2004; van Hell & Kroll, 2013) which demonstrates how even highly proficient learners still mediate L2 processing through their L1. Thus, while it may have been too difficult to reproduce the L2 explanations, some of the L1 meanings which had been inferred at the time of instruction were still accessible. For example, during the sentence matching activity, NCSp4 struggled with the sentence corresponding to 'ruck' because he was unfamiliar with the sense of 'squeeze' it contained (i.e. 'squeeze through the ruck') prompting the researcher to provide a gesture. Learning contexts such as these resulted in NCSp4's responses for 'ruck' and 'whim' on the 2-week delayed test (*konzatsu wo surinukeru* tr. to slip through a crowd; *shoudougai* tr. impulse buying), which indicate evidence of mental translation as a strategy for accessing episodic memory (Bartolotti & Marian, 2013). Thus, it can be argued that the level of engagement was greater for the NCS participants who had to go through a much more cognitively demanding process to infer the target word meanings, leaving a more durable trace in their memories. While this effort seemingly led to better retention, there was a trade-off in terms of accuracy, as inaccurate understandings were also better retained (e.g. NCSp2's inaccurate inference of 'gauche').

A particular point of interest uncovered in the results concerns an observable increase in performance between the delayed posttests for NCSp2 and NCSp4 (up 15% and 6%, respectively). NCSp2 was questioned to provide insight into the reasons for the increase, and she recounted how the incredible frustration she experienced during the 2-week delayed posttest led her to immediately consult a dictionary after leaving the classroom. However, she insisted this was an isolated incident. Although her behavior represents a violation of the study conditions, it provides promising evidence for the power of supplementing classroom instruction with intentional learning beyond class time. This is echoed by CSp1's violation discussed in **Section 4.3**. His studying of the target words outside of class time during the intervention clearly led to his superior performance on the 2-week delayed posttest (81.2%), although it is unclear to what extent he engaged in this behavior. This provides further support for developing a vocabulary learning program which encourages learners to take intentional learning beyond the classroom through repeated rehearsal to facilitate the development of stronger connections between L2 vocabulary and their existing L1 lexical networks.

In summary, it is clear that explicit, form-focused vocabulary instruction through codeswitching was not only more efficient, but led to more accurate learning of the target words compared to L2 exclusive strategies. However, this learning was considerably more fragile because CS participants did not need to expend nearly as much effort processing lexical information to generate form-meaning mappings as the participants in the NCS condition did. These results have considerable implications for language teaching, which will be discussed below.

6 Conclusion

The present study sought to investigate the effect of codeswitching compared to L2 exclusive explanations on Japanese adult EFL learners' learning, short-term, and long-term retention of novel English vocabulary words. The results suggest there are benefits for early provision of L1 equivalents in terms of accurately learning word meanings regardless of learners' level of proficiency, but these benefits are potentially diminished without subsequent opportunities for rehearsal or exposure. On the other hand, evidence also indicates benefits for both learning and retention in the case of L2 exclusive strategies, provided learners have the capacity to effectively process, store and retrieve novel lexical information. With particular respect to retention, the cognitive and metacognitive demands required to infer target word meanings from more elaborate contexts provided in the L2 explanations most likely led to greater retention, although the retention of inaccurate or erroneous understandings by less capable learners represented a considerable trade-off.

While these results are convincing, they must be interpreted in light of several critical limitations. The most obvious limitation concerns the sample size. With such a small sample size, it is not possible to generalize the results without framing them within the context of similar larger-scale studies. Another limitation concerns the disparity between the CS and NCS conditions in that conditions were not mirrored. While all CS participants received the same instruction, the individual needs of participants across the 3 separate NCS sessions forced the researcher to deviate from planned explanations when comprehension difficulties arose leading to an unavoidable degree of inconsistency. These issues could have been mitigated if the materials used in the intervention (i.e. definitions, cloze sentences) had been more adequately piloted and revised to maximize comprehensibility. However, as this was not possible, the researcher's

piecemeal measures to address these issues as they arose potentially caused further disparity. Another critical limitation comes from the lack of a pre-test which could reliably assess participants prior vocabulary knowledge. Given that the VST is a multiple-choice test, guessing becomes a viable option even when particular items are unknown. Considering how CS participants' 8-week delayed test scores actually fell below those of the initial VST gives cause for concern. Without a reliable baseline, the value of the final results is diminished since it becomes more difficult to determine which target words were actually unknown, partially known, or known by participants prior to the intervention. Finally, although the present study sought to investigate the effect *instruction* on vocabulary learning, it was impossible to control for participants' behavior outside of the classroom during the course of the intervention. Even though efforts were made to prevent participants for studying the target words outside of class, it was impossible to ascertain how effective these efforts were until participants were assessed with the delayed tests. The only way to mitigate this limitation would have been to consolidate instruction into one session per condition, but this would have necessitated a reduction in the number of target words to be taught so that sufficient time for communicative activities could still be obtained. Nevertheless, participants who engaged in this behavior give support for calls to provide more comprehensive vocabulary learning programs (e.g. Nation, 2001).

Just as other studies have concluded (e.g. Hennebry *et al.*, 2013; Tian & Macaro, 2012), there is clearly a value in using codeswitching to aid in explicit vocabulary instruction. However, while codeswitching is clearly advantageous for establishing initial form-meaning mappings, it is by no means a replacement for other potentially useful strategies. Given the sheer volume of L2 vocabulary and potentially variable uses they may have, is virtually impossible to explicitly teach every aspect of every word. Thus, it makes sense to use codeswitching to develop reliably accurate understandings of core word meanings so learners are better equipped for deeper and more elaborate processing of their behavior in context (i.e. in idiomatic expressions, or collocations). Indeed, beyond individual word meanings, languages often diverge in the way words combine to encode both culturally specific and cross-cultural conceptual knowledge. For example, while it is possible to 'lend a hand' in both English and Japanese (i.e. *te wo kasu*), it is only possible to '*borrow* a hand' (*te wo kariru*) in Japanese. Learners need to be made aware of these critical distinctions so that appropriate conceptual restructuring can take place over time

(Pavlenko, 2009). If the learners' L1 is banned from the classroom, opportunities to develop deeper understandings and metacognitive knowledge of these cross-linguistic differences are denied. Thus, conscientious teachers would serve their learners well by organizing a vocabulary learning program that promotes strong initial learning through explicit form-focused instruction (including the judicious and principled use of codeswitching) with plenty of opportunities for repeated, subsequent incidental exposure in diverse contexts.

References

Antón, M. & DiCamilla, F.J. (1998). Socio-cognitive functions of L1 collaborative interaction in the L2 classroom. *Canadian Modern Language Review* 54 (3), 314–342.

Atkinson, D. (1987). The mother tongue in the classroom: A neglected resource? *ELT Journal* 41(4), 241–247.

Atkinson, D. (1993). *Teaching monolingual classes*. London: Longman.

Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests* (Vol. 1). Oxford University Press.

Brookes, F.B. & Donato, R. (1994). Vygotskyan approaches to understanding foreign language learner discourse during communicative tasks. *Hispania,* 77(2), 262–74.

Bartolotti, J., & Marian, V. (2013).' Bilingual memory: structure, access, and processing.' In J., J. Altarriba, & L. Isurin (Eds.) (2013). *Memory, Language, and Bilingualism: Theoretical and Applied Approaches.* Cambridge: Cambridge University Press. 7-47.

Brooks-Lewis, K. (2009). Adult learners' perceptions of the incorporation of their L1 in foreign language teaching and learning. *Applied Linguistics*, 30(2), 216–235.

Brown, H. D. (2001). *Teaching by Principles: An Interactive Approach to Language Pedagogy*. (2nd ed.) New York: Longman.

Butzkamm, W. (2003). We only learn language once. The role of the mother tongue in FL classrooms: Death of dogma. *Language Learning Journal*, 28(1), 29–39.

Butzkamm, W. & J. Caldwell (2009). *The Bilingual Reform: A Paradigm Shift in Foreign Language Teaching*. Tubingen: Narr Studienbücher.

Carless, D. (2002). Implementing task-based learning with young learners. *ELT Journal*, 56(4), 389– 396.

Carless, D. (2004). Issues in teachers' reinterpretation of a task-based innovation in primary schools. *TESOL Quarterly*, 38(4), 639–662.

Cook, V. (2001). Using the first language in the classroom. *Canadian Modern Language Review* 57(3), 402–423.

Council of Europe (n.d.). *Common European Framework of Reference for Languages: learning, teaching, assessment* [online]. http://www.coe.int/t/dg4/linguistic/cadre1_en.asp (Accessed March 18, 2016)

Cummins, J. (2007). Rethinking monolingual instructional strategies in multilingual classrooms. *Canadian Journal of Applied Linguistics (CJAL)*, 10(2), 221–240.

de la Campa, J., & Nassaji, H. (2009). The amount, purpose, and reasons for using L1 in L2 classrooms. *Foreign Language Annuals*, 42(4), 742–759.

Duff, P. & Polio, C. (1990). How much foreign language is there in the foreign language classroom? *The Modern Language Journal*, 74(2), 154–166.

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition*, 23(2), 166-180.

Edstrom, A. (2006). L1 use in the L2 classroom: One teacher's self-evaluation. *Canadian Modern Language Review*, 63(2), 275–292.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27(2), 305-352.

Ellis, R. (1994). *The Study of Second Language Acquisition.* Oxford: Oxford University Press.

Ellis, R. & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition,* 21, 285–301.

Hall, G., & Cook, G. (2012). Own-language use in language teaching and learning. *Language Teaching*, 45(3), 271-308.

Hennebry, M., Rogers, V., Macaro, E., & Murphy, V. (2013). Direct teaching of vocabulary after listening: is it worth the effort and what method is best?. *The Language Learning Journal*, 1-19.

Howatt, A. (1984). *A History of Language Teaching*. Oxford: Oxford University Press.

Hulstijn, J.H. (2001). 'Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity.' In P. Robinson, (Ed.), *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press. 258–286.

Hulstijn, J. H. (2008). 'Incidental and intentional learning.' In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition*. Oxford: Blackwell. 349–381.

Hulstijn, J.H., Hollander, M. & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: the influence of marginal glosses, dictionary use and reoccurrence of unknown words. *The Modern Language Journal*, 80(3), 327–39.

Hummel, K. (2010). Translation and short-term L2 vocabulary retention: Hindrance or help? *Language Teaching Research*, 14(1), 61–74.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21(1), 47–77.

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, 24(4), 617–637.

Jiang, N. (2004a). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, 88(3), 416-432.

Jiang, N. (2004b). 'Semantic transfer and development in adult L2 vocabulary acquisition.' In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a Second Language: Selection, Acquisition and Testing*. Philadelphia/Amsterdam: John Benjamins. 101–126.

Kern, R. (1994). The role of mental translation in second language reading. *Studies in Second Language Acquisition*, 16(4), 441–461.

Kim, S. H. O., & Elder, C. (2005). Language choices and pedagogic functions in the foreign language classroom: A cross-linguistic functional analysis of teacher talk. *Language Teaching Research*, 9(4), 355-380.

Kim, S.H. O. & Elder, C. (2008). Target language use in foreign language classrooms: Practices and perceptions of two native speaker teachers in New Zealand. *Language, Culture and Communication*, 21(2), 167–185.

Knight, S. M. (1994). Dictionary use while reading: the effects on comprehension and vocabulary acquisition for students of different verbal abilities. *Modern Language Journal*, 78(3), 285–299.

Krashen, S.D. (1985). *The Input Hypothesis: Issues and Implications.* London: Longman.

Kroll, J.F. & Stewart, E. (1994). Category interferences in translation and picture naming: evidence for asymmetric connection between bilingual memory representation. *Journal of Memory and Language,* 33(2), 149–74.

Larsen-Freeman, D. (2000). *Techniques and principles in language teaching* (2nd ed.). Oxford: Oxford University Press.

Laufer, B. (2005). Focus on form in second language vocabulary learning. *Eurosla yearbook*, 5(1), 223-250.

Laufer, B. & N. Girsai (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics,* 29(4), 694–716.

Laufer, B. & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal*, 28(1), 89–108.

Lee, J. H., & Macaro, E. (2013). Investigating age in the use of L1 or english-only instruction: Vocabulary acquisition by Korean EFL learners. *The Modern Language Journal*, 97(4), 887-901.

Littlewood, W., & Yu, B. (2011). First language and target language in the foreign language classroom. *Language Teaching*, 44(1), 64-77.

Liu, J. (2008). L1 use in L2 vocabulary learning: facilitator or barrier? *International Education Studies*, 1(2), 66–70.

Liu, D., Ahn, G., Beak, K. & Han, N. (2004). South Korean high school English teachers' code switching: Questions and challenges in the drive for maximal use of English in teaching. *TESOL Quarterly*, 38(4), 605–638.

Long, M. (1981). Input, interaction and foreign language acquisition. In Winitz, H. (ed.) *Native Language and Foreign Language Acquisition*. Annals of the New York Academy of Sciences 379. 259–278.

Lotto, L. & De Groot, A.M.B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, 48(1), 31–69.

Lucas, T., & Katz, A. (1994). Reframing the debate: The roles of native languages in English-only programs for language minority students. *TESOL Quarterly*, 28(3), 537-561.

Luppescu, S. & Day, R. R. (1993). Reading, dictionaries and vocabulary learning. *Language Learning*, 43(2) 263–79.

Macaro, E. (1997). *Target Language, Collaborative Learning and Autonomy.* Clevedon: Multilingual Matters.

Macaro, E. (2000). 'Issues in target language teaching.' In K. Field (Ed.) *Issues in Modern Foreign Language Teaching*. London: Routledge. 171–189.

Macaro, E. (2001). Analysing student teachers' codeswitching in foreign language classrooms: Theories and decision making. *The Modern Language Journal,* 85 (4), 531–548.

Macaro, E. (2005). 'Codeswitching in the L2 classroom: A communication and learning strategy.' In E. Llurda (Ed.) *Non-Native Language Teachers: Perceptions, Challenges and Contributions to the Profession*. New York: Springer. 63–84.

Macaro, E. (2009). 'Teacher use of codeswitching in the second language classroom: Exploring "optimal" use.' In M. Turnbull & J. Dailey–O'Cain (Eds.), *First Language Use in Second and Foreign Language Learning*. Bristol, UK: Multilingual Matters. 35–49.

Macaro, E. (2014). Students' strategies in response to teachers' second language explanations of lexical items. *The Language Learning Journal*, 1-16.

Macaro, E. & Mutton, T. (2002). Developing language teachers through a co- researcher model. *Language Learning Journal*, 25(1), 27–39.

Min, H. (2008). EFL vocabulary acquisition and retention: reading plus vocabulary enhancement activities and narrow reading. *Language Learning*, 58(1), 73–115.

Mitchell, R. (1988). *Communicative language teaching in practice*. London: CILT.

Mondria, J. (2003). The effects of inferring, verifying and memorizing on the retention of L2 word meanings: an experimental comparison of the "Meaning-inferred Method" and the 'Meaning-given Method'. *Studies in Second Language Acquisition* , 25(4), 473–499.

Mondria, J-M. & B. Wiersma (2004). 'Receptive, productive, and receptive + productive L2 vocabulary learning: What difference does it make?' In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language*. Amsterdam & Philadelphia: John Benjamins. 80−100.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language.* Cambridge: Cambridge University Press.

Nation, I. S. P. (2012). The Vocabulary Size Test: Information and specifications. Retrieved from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/VocabularySize-Test-information-and-specifications.pdf (Accessed February 18 2016)

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. The Language Teacher, 31(7), 9–13. Retrieved from http://jalt-publications.org/tlt/issues/2007-07_31.7 (Accessed February 18 2016)

Nation, P. & Wang Ming-Tzu, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, 12(2), 355–379.

O'Malley, J.M., Chamot, A.U., & Küpper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics*, 10(4), 418-437.

Pavlenko, A. (2009). 'Conceptual representation in the bilingual lexicon and second language vocabulary learning.' In A. Pavlenko (Ed.), *The Bilingual Mental Lexicon: Interdisciplinary Approaches*. Clevedon: Multilingual Matters. 125–160.

Phillipson, R. (1992). *Linguistic imperialism*. Oxford: Oxford University Press.

Pica, T., Young, R., & Doughty, C. (1987). The impact of interaction on comprehension. *TESOL Quarterly,* 21(4), 737–58.

Polio, C. & P. Duff (1994). Teachers' language use in university foreign language classrooms: A qualitative analysis of English and target language alternation. *The Modern Language Journal* 78(3), 313–326.

Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *The Modern Language Journal*, 80(4), 478–493.

Prodromou, L. (2002). 'The liberating role of the mother tongue.' In S. Deller & M. Rinvolucri (Eds.), *Using the Mother Tongue: Making the Most of the Learner's Language*. London: ETP and DELTA Publishing, 5.

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.

Rolin-Ianziti, J. & S. Brownlie (2002). Teacher use of learners' native language in the foreign language classroom. *Canadian Modern Language Review*, 58(3), 402–426.

Rolin-Ianziti, J. & R. Varshney (2008). Students' views regarding the use of the first language: An exploratory study in a tertiary context maximizing target language use. *Canadian Modern Language Review*, 65(2), 249–273.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329-363.

Swain, M. (1985). 'Communicative competence: Some roles of comprehensible input and comprehensible output in its development.' In S. Gass & C. Madden (Eds.), *Input in second language acquisition*. Rowley, MA: Newbury House. 235–253.

Swain, M. & S. Lapkin (2000). Task-based second language learning: The uses of the first language. *Language Teaching Research*, 4(3), 251–274.

Swain, M, Kirkpatrick, A., & Cummins, J. (2011). How to Have a Guilt-free Life Using Cantonese in the English Class: A Handbook for the English Language Teacher in Hong Kong. Hong Kong: Research Center into Language Acquisition and Education in Multilingual Societies, Hong Kong Institute of Education.

Tian, L., & Macaro, E. (2012). Comparing the effect of teacher codeswitching with English-only explanations on the vocabulary acquisition of Chinese university students: A lexical focus-on-form study. *Language Teaching Research*, 16(3), 367–391.

Turnbull, M. (2001). There is a role for the L1 in second and foreign language teaching, but... *Canadian Modern Language Review*, 57(4), 531–540.

Turnbull, M. & K. Arnett (2002). Teachers' uses of the target and first languages in second and foreign language classrooms. *Annual Review of Applied Linguistics,* 22, 204–218.

Turnbull, M. & Dailey-O'Cain, J. (2009). 'Introduction.' In M. Turnbull & J. Dailey-O'Cain. (Eds.) (2009), *First Language Use in Second and Foreign Language Learning*. Bristol: Multilingual Matters. 1-14.

Van Hell, J. G., & Kroll, J. F. (2012). 'Using electrophysiological measures to track the mapping of words to concepts in the bilingual brain: A focus on translation.' In J. Altarriba & L. Isurin (Eds.), *Memory, Language, and Bilingualism: Theoretical and Applied Approaches.* New York: Cambridge University Press.126–160.

Vandergrift, L. (2003). Orchestrating strategy use: Toward a model of the skilled second language listener. *Language Learning*, 53(3), 463-496.

Vygotsky, L.S. (1978). 'Mind in Society: The Development of Higher Psychological Processes.' In M. Cole, V. John-Steiner, S. Scribner and E. Souberman (Eds). Cambridge, MA: Harvard University Press.

Watanabe, Y. (1997). Input, intake, and retention. *Studies in Second Language Acquisition*, 19(3), 287-307.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27(1), 33-52.

Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC journal*, 40(3), 360-376.

Zhao, T., & Macaro, E. (2014). What works better for the learning of concrete and abstract words: teachers' L1 use or L2-only explanations?. *International Journal of Applied Linguistics.* http://dx.doi.org/10.1111/ijal.12080.

Appendices

Appendix I. Immediate posttest example

Name:_____

Date:_____

Please look at the words on the left. If you don't know them, please circle "0"; if you know their meaning, please write down either the Japanese or English meaning and, and rate how confident you are about your answer, from "1" (not confident) to "5" (very confident).

compost	**0**	_____	**1**	**2**	**3**	**4**	**5**
Hutch	**0**	_____	**1**	**2**	**3**	**4**	**5**
crowbar	**0**	_____	**1**	**2**	**3**	**4**	**5**
mumble	**0**	_____	**1**	**2**	**3**	**4**	**5**
authentic	**0**	_____	**1**	**2**	**3**	**4**	**5**
Jovial	**0**	_____	**1**	**2**	**3**	**4**	**5**

Appendix II. Delayed posttest

Name:_____

Date:_____

Please look at the words on the left. If you don't know them, please circle "0"; if you know their meaning, please write down either the Japanese or English meaning and, and rate how confident you are about your answer, from "1" (not confident) to "5" (very confident).

ruck	**0**		**1**	**2**	**3**	**4**	**5**
hallmark	**0**		**1**	**2**	**3**	**4**	**5**
counterclaim	**0**		**1**	**2**	**3**	**4**	**5**
allege	**0**		**1**	**2**	**3**	**4**	**5**
candid	**0**		**1**	**2**	**3**	**4**	**5**
gauche	**0**		**1**	**2**	**3**	**4**	**5**
compost	**0**		**1**	**2**	**3**	**4**	**5**
hutch	**0**		**1**	**2**	**3**	**4**	**5**
crowbar	**0**		**1**	**2**	**3**	**4**	**5**
mumble	**0**		**1**	**2**	**3**	**4**	**5**
authentic	**0**		**1**	**2**	**3**	**4**	**5**
jovial	**0**		**1**	**2**	**3**	**4**	**5**
whim	**0**		**1**	**2**	**3**	**4**	**5**
peasantry	**0**		**1**	**2**	**3**	**4**	**5**
regent	**0**		**1**	**2**	**3**	**4**	**5**
veer	**0**		**1**	**2**	**3**	**4**	**5**
malign	**0**		**1**	**2**	**3**	**4**	**5**
egalitarian	**0**		**1**	**2**	**3**	**4**	**5**
refectory	**0**		**1**	**2**	**3**	**4**	**5**
communiqué	**0**		**1**	**2**	**3**	**4**	**5**
figure	**0**		**1**	**2**	**3**	**4**	**5**
strangle	**0**		**1**	**2**	**3**	**4**	**5**
didactic	**0**		**1**	**2**	**3**	**4**	**5**
devious	**0**		**1**	**2**	**3**	**4**	**5**

Appendix III. Answer key

Word	Acceptable English answers	Acceptable Japanese answers
Ruck	crowd; mob; throng	人ごみ; 混雑; 群集; 群; 雑踏
Hallmark	characteristic; distinctive feature; stamp or seal of quality	純分認証極印; 太鼓判; 特徴; 特性
Counterclaim	a claim made against a previous claim; rebuttal	反訴; 反論; 反証; 反対意見
Candid	honest; straightforward; frank; direct	率直な; 正直な; 素直な
Gauche	awkward; inept; clumsy; graceless; unrefined; unsophisticated	ぎこちない; 不器用な; 上品でない; 品のない; ぶざまな
Allege	claim; assert; argue	主張する; 断言する; 言い張る
Compost	organic fertilizer; decayed organic material; humus	堆肥; 肥料; コンポスト;腐葉土; 配合土
hutch	small house for domesticated animals	小屋; おり; 小さな家
crowbar	iron bar; pry bar	釘抜き; くぎぬき;金てこ; バール
mumble	speak indistinctly; mutter; murmur	もぐもぐ言う; もごもご言う; ぶつぶつ言う; つぶやく
authentic	real; genuine; original	本物の; 本当の; 本格的な; 真正の
jovial	happy; cheerful; merry; friendly; joyful; sanguine	陽気な; 快活な; 楽天的な; 愉快な
whim	impulse; urge; caprice	気まぐれ; でき心; むら気; 衝動
peasantry	low-class farmers; sharecroppers; laborers	小作農; 農民; 農夫; 小作人;
regent	one who governs a state when the sovereign is too young, absent, or disabled; temporary/substitute governor	摂政; 君主の代わりに国を治める人
veer	(suddenly) change direction; swerve	(急に)方向を変える; 曲がって進む; それる
malign	evil; bad; malevolent; destructive	悪意のある; 有害な; 悪い
egalitarian	believing in social equality; equitable; equal	平等主義の; 平等な
refectory	church/ school cafeteria; dining hall	(修道院・大学などの)食堂
communiqué	official announcement/statement; press release; bulletin, proclamation; declaration	公式声明; 声明; 声明書; 公式発表; 公報; 宣言
figure	number; statistic	数字; 数; けた; 単位
strangle	choke; throttle; stifle; asphyxiate	絞め殺す; 窒息死させる
didactic	educational; educative; instructive; pedagogic	教訓的な; 説教的な; 教育的な
devious	dishonest; dishonorable; unethical; underhanded; deceitful; tricky; immoral; dubious; scheming; wicked	率直でない; 素直でない; よこしまな ; 腹黒い; ずる賢い; 邪悪な

Appendix IV. Example of Phase 1 Materials (Session 3)

Target words	Definitions	Cloze sentences
whim	**a sudden desire or change of mind, especially one that is unusual or unexplained.**	**On a ___ , Brenda bought herself a new designer bag while windowshopping downtown.**
peasantry	**people of low social status and education, typically farmers or laborers.**	**The governor's irresponsible actions during the famine caused the local ___ great hardship.**
regent	**a person appointed to govern a state when the monarch is too young, absent or disabled.**	**The boy's mother was named ___ until he was old enough to take the throne.**
malign	**evil in nature or effect.**	**The ___ influence of the NRA has prevented the passing of common sense gun control laws.**
egalitarian	**the principle that all people are equal and deserve equal rights and opportunities.**	**The US justice system is not ___ because rich white people are often treated less harshly than minorities or poor people.**
veer	**to change direction suddenly.**	**The bus ___ out of the way to avoid a fallen tree on the highway.**