Research Article

Hyun-Bin Hwang*

Image-schema-based-instruction enhanced L2 construction learning with the optimal balance between attention to form and meaning

https://doi.org/10.1515/iral-2023-0015
Received January 21, 2023; accepted April 7, 2023; published online April 21, 2023

Abstract: This study investigated the effectiveness of L2 instruction on the learning of English caused-motion constructions (e.g., Jane put the balls into the box, Harry swam Peter to the beach). Korean EFL high school students (N = 156) were randomly assigned to three instructional conditions and one control group. While the meta-language group relied on explicit uses of grammar jargon, the input flood group read narrative stories in which the instances of a prototypical verb were presented with higher frequency than other verbs. The image-schema group, which adopted a balanced approach to form and meaning attention, studied the core meaning of the target construction. Mixed-effects logistic regressions on Korean-to-English translation tests revealed that the image-schema group was most effective in terms of learnability and generalizability. The benefits of image-schema-based instruction were attributed to simultaneous attention to form and meaning, which might lead to deeper processing. Image-schema-based-instruction is discussed as a viable alternative for L2 construction learning.

Keywords: cognitive-linguistics-inspired (CL-inspired) instruction; construction grammar; construction learning; image-schema; image-schema-based-instruction (ISBI); input flood

1 Introduction

Even with the advent of focus on form in the 1990s (Long 1991), English as foreign language (EFL) classrooms still tend to favor traditional types of grammar

*Corresponding author: Hyun-Bin Hwang. Second Language Studies, Michigan State University, East Lansing, MI, USA, E-mail: hwangh16@msu.edu. https://orcid.org/0000-0001-6403-0805

Open Access. © 2023 the author(s), published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.
instruction such as present-practice-produce (Larsen-Freeman 2015; Pawlak 2021). Furthermore, such explicit instruction, more often than not, occurs with a sustained use of grammatical terminology (Loewen 2020), but the question of whether metalanguage is beneficial for second language (L2) learning is still inconclusive (Toth et al. 2020). On the other hand, providing the large amounts of the language input over an extensive learning period, which is required for implicit learning, is not a realistic expectation in formal instructional settings (Butler 2017; Muñoz 2008). In this regard, EFL teachers shoulder the responsibility to implement L2 instruction that could maximize learning outcomes from the limited input. To accomplish this pedagogical objective, it is important for L2 teachers to seek the optimal balance between attention to language form and meaning, which is consistent with their own teaching contexts and goals (Leow 2018; Loewen and Sato 2017).

The present study aims to investigate the ways to balance attention to form and meaning for more efficient L2 construction learning from a pedagogical perspective. The instructional conditions in the present study differed with respect to whether learners’ attention was primarily directed to language form or meaning. The meta-language group, typifying EFL classrooms, represented form-oriented instruction in which explicit grammar explanation made formal features of a linguistic target salient. On the other hand, the input flood group represented meaning-oriented instruction in which learners were implicitly exposed to multiple instances of a target construction while reading narrative stories for comprehension. The image-schema group, inspired by the principle of embodied meaning in cognitive linguistics, was presented with an abstract core meaning that underlied a target construction as an effort to seek the middle ground in-between the aforementioned two groups. The research findings are discussed with regards to how the three types of instruction can influence the ability to learn a target construction and generalize the learned constructional knowledge to a novel usage.

## 2 Literature review

### 2.1 Input flood and frequency distribution

Constructions, defined as learned pairings of form and meaning or function (Ellis and Wulff 2020), are the basic units of our linguistic knowledge (Goldberg 2006). The usage-based model of language learning holds that constructions must be learned by processing many exemplars in language input, and our language system (i.e., a network of constructions) emerges gradually on the basis of generalizing over item-specific constructional knowledge (Ellis 2006; Goldberg 2003, 2006; Tomasello 2000a, 2000b). In this regard, input flood, which is one of the most implicit form-
focused instruction (Loewen 2018), can aid frequency-based construction learning by artificially increasing the frequency of a target construction. Furthermore, repetitive exposure to multiple exemplars seeded in the input can enhance the salience of a linguistic target (Han et al. 2008; Liu et al. 2021), assisting learners in noticing linguistic features within meaningful contexts (e.g., stories), and thereby facilitating the form-meaning association of a construction. Such theoretical motivation notwithstanding, a large body of empirical research into the effectiveness of input flood has found mixed results. Some studies found significant effects of input flood alone on L2 learning (e.g., Hernández 2011; Loewen and Inceoglu 2016; Reinders and Ellis 2009), but other studies failed to observe such benefits (e.g., Jahan and Kormos 2015; Loewen et al. 2009; Szudarski and Carter 2016). These findings suggest that input flood by itself might not be salient enough to draw learners’ attention to targeted linguistic structures (Loewen 2020). In this regard, several attempts have been made to investigate the effects of input flood in combination of other techniques such as textual enhancement and explicit instruction. Overall, the benefits of compound input flood over simple input flood or a control group have been reported (e.g., Issa and Morgan-Short 2019; Liu and Hao 2021; Rassaei 2015), with the caveat that overenhancing target constructions could be counterproductive (Han et al. 2008; White 1998).

Given the repetitive nature of input flood (Szudarski and Carter 2016), the inconsistent results of previous studies might be due to large variations in the number of occurrences of a target construction. For example, Hernández (2011) and Loewen and Inceoglu (2016), which found the positive effects of input flood, presented 20 and 28 times of target constructions each, whereas the number of encounters with target items were 5 and 12 times in Jahan and Kormos (2015) and Szudarski and Carter (2016), respectively. While no clear evidence has been obtained for the optimal number of exemplars for input flood (Wong 2005), a small group of studies have compared the effects of the input with different numbers of exemplars in a systematic way (e.g., Leow et al. 2003; Shook 1999). Most recently, Liu et al. (2021) found that providing contextual cues for inferring a word’s meaning had more profound effects on L2 lexical learning than manipulating the number of instances of the target item from 2 to 10. These results indicated that merely the higher frequency of a construction in input flood might not necessarily lead to better learning outcomes because learners become desensitized and pay less attention to the later encounters of the same target words (e.g., Winke 2013).

In tandem with the frequency of exposure to a construction, frequency distribution is another methodological factor that has remained underinvestigated in input flood (e.g., Denhovska et al. 2016). There is good evidence that type/token frequency distributions can play a substantial role in construction learning (Ellis and Ferreira-Junior 2009; Goldberg 2006), specifically facilitated by the presence of
one prototypical member that accounts for the lion’s share of a particular construction. For example, Goldberg et al. (2004) examined extensive corpus data of children’s and mothers’ speech. The analysis of fifteen mothers’ speech to their 28-month-old children revealed that one prototypical verb appeared in a particular construction with very high frequency (e.g., the verb put accounting for almost 38% tokens of caused-motion constructions) and this tendency was mirrored in their children’s speech. The authors concluded that such skewed input might produce a cognitive anchoring effect which may induce children to form a tentative category organized around the central member. This initially formed category, in turn, might serve as a stronger category representation for easier assimilation of other members to that category (Elio and Anderson 1984). Goldberg et al. (2007) further manipulated the order of presentation of instances. In the experiment, the skewed-first group was first provided with eight instances of one prototypical verb before those of four other verbs (i.e., 8-2-2-2-2) whereas the skewed-random group met the same eight instances later than the other instances (i.e., 2-2-2-8-2). It was found that the skewed-first group identified new instances of the target construction more accurately than the skewed-random group. Their results suggested that shared concrete similarity in the initial stage of construction learning allowed learners to form an abstract category that could account for most of the instances of the target construction. In a recent study, Zhang and Mai (2021) reported an additional facilitative effect of skewed-first input on the learning of English present and past counterfactual conditionals, which shared subtle differences in both form and meaning. Their findings highlighted that an initial low variance of the skewed input might reduce negative interference that arises when two similar constructions are learned together. With these benefits in mind, the present study manipulated frequency distribution in input flood to be skewed disproportionately towards a single prototypical verb of a target construction. It was hypothesized that such skewed input flood would facilitate frequency-based construction learning (e.g., extracting regularities in language usage).

### 2.2 Image-schema-based-instruction in L2 pedagogy

Mark Johnson, known for his early contributions to cognitive semantics, criticized the objectivist view of meaning, which assumed that the nature of meaning was independent of human experiences with their surrounding physical and social environment. Instead, Johnson argued that “meaning cannot be separated from the structures of our embodied perceptual interactions and movements” (Johnson 1989, p. 109). In a critical response to the objectivist view, cognitive semantics introduced
image-schema in order to explain this embodiment of meaning (e.g., Lakoff and Johnson 1980).

Image-schema, as one of the key terms of cognitive linguistics, is “a relatively abstract conceptual representation that arises directly from our everyday interaction with and observation of the world around us” (Evans 2007, p. 106). To put it simply, image-schema is the abstract generalization of everyday experiences of the physical and social world. For instance, infants as young as six months can distinguish caused-motion from self-motion (Mandler 2004). Their cognitive system begins to develop the conceptual structure of caused-motion through repeated encounters with the spatio-physical events in which someone exerts force to move an entity from one place to another. The image-schema exemplified in Figure 1 emerged through generalization over perceived similarities of such recurring experiences. That is, it is our embodied experiences with the surrounding environment that give rise to such image-schema. Mandler (2004) further argued that such image-schema, in turn, could shape our cognition and provide fundamental frames of reference for linguistic concepts and meanings. For example, the caused-motion image-schema in Figure 1 serves as a conceptual tool for the interpretation of caused-motion constructions. In this respect, language is firmly grounded in our experience, and meanings are inseparable from the ways we perceive and categorize the real world around us (Ellis and Cadierno 2009). Therefore, meaning is embodied (Johnson 2007; Tyler 2012).

With many variations in terminology (e.g., mental imagery, schematic diagram, core-image, proto-scene, schema of a complete orienting basis of action), the idea to utilize image-schema for L2 pedagogy is not new, and a number of quasi-experimental studies have investigated the effectiveness of image-schema-based instruction on the learning of various types of constructions such as (a) prepositions (e.g., Hung et al. 2018; Tyler et al. 2011, Wong et al. 2018; Zhao et al. 2020), (b) verbs (e.g., Morimoto and Loewen 2007; Mueller and Tsushima 2019; Yamagata 2018; Verspoor and Lowie 2003), (c) modal verbs (e.g., Tyler et al. 2010), (d) phrasal verbs (e.g., Lee 2016; White 2012), (e) figurative idioms (e.g., Boers et al. 2004; Szczepaniak and Lew 2011), and (f) syntactic structures (e.g., Jacobsen 2016). The common principle of such image-schema-based instruction is to take advantage of image-schema as a vehicle to visualize an abstract core meaning that underlies a particular construction. The increased salience of the central meaning in this way is expected to

---

**Figure 1**: Caused-motion image-schema (Mandler 1992, p. 595).
help learners internalize embodied meaning which lays conceptual and semantic foundations of the construction.

While the potential benefits of image-schema-based instruction for L2 construction learning are generally acknowledged (Yu 2022), its main rationale can be justified by the following two theoretical assumptions. First, image-schema can facilitate deeper processing of the form-meaning association of constructions. The levels of processing model (Craik and Lockhart 1972) indicates that the greater cognitive efforts involved when information is processed will create richer memory representations, ultimately leading to better performance and retention than otherwise shallow processing. In the scene of L2 learning, deeper levels of processing refer to the simultaneous attention to form and meaning in that it presumably requires higher amount of cognitive involvement and elaboration (Leow 2015; Rice and Tokowicz 2020). However, since our limited attentional resources tend to give priority to processing input for meaning over form, making the concurrent association of form and meaning would be more effortful and demanding for L2 learners (VanPatten 2004). In this regard, image-schema-based instruction could save L2 learners the trouble of implicitly extracting the shared semantic structure of a construction from multiple instances in the input, the process of which might be challenging in EFL classrooms due to limited amounts of the input (Morimoto and Loewen 2007). As a result, such meaning salience can reduce cognitive load for semantic processing and simultaneously allow more attentional resources to be allocated for a higher level of form processing (i.e., accurate understanding of the underlying rule).

Second, the abstract and schematic nature of image-schema can help learners make inferences on novel usages (Tyler and Evans 2004; Mahpeykar and Tyler 2015), resulting in better transfer to similar linguistic contexts. Since image-schema is established through generalization of similar and recurring experiences, they are “vague and flexible enough to be adjustable to a range of different contexts” (Littlemore 2009, p. 130). Pedagogically, this also means that image-schema can present the shared core semantics in an intuitively appealing way, preventing L2 learners from being constrained to possibly inaccurate L1–L2 equivalents or metalanguage (Morimoto and Loewen 2007). Therefore, the inherent flexibility of image-schema can help L2 learners have a deeper understanding of the cognitive mechanism behind seemingly unmotivated meanings of constructions, and, in turn, better generalize to the novel usages in different contexts in systematic and consistent ways.

It has been pointed out that previous studies on image-schema-based instruction are “largely confined to the pedagogical exploitation of figurative thought, and this mostly pertains to polysemous words (especially prepositions and particles) and idiomatic expressions” (Boers and Lindstromberg 2006, p. 305). In this regard, future
research is needed to see whether the potential benefits of image-schema could be transferred to more complex syntactic constructions (Boers and Lindstromberg 2008; Tyler 2012). To address this gap, argument structure constructions were chosen as the linguistic target for the study.

2.3 Argument structure constructions

Unlike a traditional view that constructions only refer to clause-level expressions (e.g., wh-construction), constructions in the usage-based approaches can vary in their size and complexity across all layers of language (Goldberg 2006). For instance, constructions can be morphemes (e.g., pre-, -ing), words (e.g., and, avocado, regular plural [N-s]), and idiomatic expressions (e.g., give the Devil his due, jog <someone’s> memory). Furthermore, abstract syntactic patterns by themselves, so-called argument structures, are also constructions. More specifically, the verb put in (1) is a predicate which typically takes three arguments: an agent (Jane), a theme (her toys), and a goal (into the box). Whereas the verb swim only requires an argument of someone who is swimming, it can also appear with the non-compositional meaning as in (2), roughly interpreted as ‘the dinosaur moved his friends to the mainland by means of swimming’.

(1) Jane put her toys into the box.

(2) The dinosaur sawm his friends to the mainland (Hilpert 2019).

Goldberg (1995, 2006) suggested that it was the argument structure that overrode the meaning of the verb swim and coerced it into having a specific sense. In other words, the shared syntactic pattern (i.e., Subject Verb Object Obliquepath/loc) underlying in (1) and (2) yields the caused-motion interpretation (i.e., X causes Y to move along or towards Z). This is possible because additional theme and goal arguments are imposed on the verbs that otherwise cannot convey such non-compositional meaning by themselves. These examples well demonstrate that argument structures are also one kind of constructions in that their syntactic patterns alone can carry an inherent meaning, independently of the actual verb with which they combine. In this particular case, the argument structure constructions exemplified in (1) and (2) are called caused-motion constructions, which are the linguistic target of the present study. Especially, caused-motion constructions like (2) were defined as non-typical because the verbs (e.g., swim, fly, drive) that appeared in the non-typical construction did not have a caused-motion sense by default, and must be integrated with its syntactic structure to gain such interpretation.
3 Research questions

In order to find the optimal balance between attention to language form and meaning for L2 construction learning, three instructional conditions in the study were justified with respect to different degree of priority given to form and meaning in a learning phase. A metalanguage group represented explicit instruction in that metalinguistic explanation of the target construction first directed learners to pay attention to its formal features with L1 translation provided for meaning. An input flood group represented implicit instruction in that “neither rule presentation nor directions to attend to particular forms that were part of a treatment” (Norris and Ortega 2000, p. 437). Learners’ attention was incidentally directed to formal patterns by skewed input while reading stories primarily for meaning. An image-schema group was approached in a balanced way by presenting linguistic structures more implicitly than the metalanguage group (i.e., color-coding instead of direct rule representation), but by presenting meaning more explicitly than the input flood group (i.e., visualized core meaning instead of mere exposure). The control group, which followed a normal English class, was included as a baseline to compare the instructional effectiveness against. Research questions were thus formulated as follows:

1. Do three types of instruction with different degrees of explicitness (i.e., input flood, metalanguage, image-schema) have effects on the learning of typical English caused-motion constructions?

2. Do three types of instruction with different degrees of explicitness (i.e., input flood, metalanguage, image-schema) have effects on the generalization to non-typical English caused-motion constructions?

4 Method

4.1 Participants

A total of 156 high school students (aged 15–16) in South Korea participated in the study. The students’ average English proficiency was low-to-intermediate, as measured with their English grades in Korea National College Scholastic Ability Test. The majority belonged to or below the fifth bands on the Stanine scale (i.e., the bottom 60% of the nationwide testing population). Eight intact classes were randomly divided into (a) an input flood (IF) group (n = 37), (b) a metalanguage (MT) group (n = 38), (c) an image-schema (IS) group (n = 43), and (d) a control group (n = 38).
4.2 Target constructions

4.2.1 Typical caused-motion constructions (e.g., Jane put the balls into the box)

Typical caused-motion constructions in the present study contained 10 verbs: put, give, take, get, send, bring, throw, place, kick, and pass. These ten verbs were selected among the twenty most frequent verbs occurring in caused-motion constructions in corpus data (Hwang 2014). While all types of verbs were introduced to the three instructional conditions, the IF group received greater amounts of exposure to the typical caused-motion construction (108 tokens) in comparison to the MT and IS groups (23 tokens, respectively). The typical caused-motion constructions were provided in both learning and testing phases to examine whether and how much a typical usage of caused-motion constructions could be learned through instruction (i.e., learnability).

4.2.2 Non-typical caused-motion constructions (e.g., Harry swam Peter to the beach)

In order to prove the existence of argument structure constructions, it is necessary to show that “the particular combination of lexical items should not inevitably lead to the particular interpretation in question” (Goldberg 1995, p. 153). Learners, for instance, need to tap into knowledge of caused-motion constructions when processing or constructing the meaning of sentences like ‘Harry swam Peter to the beach’ since the verb swim does not independently carry a caused-motion sense. Justified by this non-compositionality principle, 10 verbs (i.e., talk, dance, swim, drive, fly, row, walk, sneeze, march, and hurry) were selected and caused-motion constructions combined with those verbs were defined as non-typical. The non-typical caused-motion constructions were presented only in the testing phases to investigate whether and how much a typical usage of caused-motion constructions could be generalized to its non-typical usage with the aid of instruction (i.e., generalizability).

4.3 Learning materials

The IF group was asked to read four narrative stories (see Figure 2) in which multiple exemplars of the typical caused-motion constructions were embedded with boldfaced so that learners’ attention could be also given to form incidentally while processing meaning (e.g., Issa and Morgan-Short 2019; Winke 2013). Specifically, caused-motion constructions with the verb put appeared more frequently (26 times) than those with
Once upon a time there was a Mother Goat and her seven goat children. One day Mother Goat had to go to the market to buy food. She told her children to lock the door, and to only open it for her. She warned her children about the Big Bad Wolf, who would try to get inside the house and eat them.

“You will know it is me because of my sweet voice and white feet,” said Mother Goat.

She gave a kiss to each of her children. Then she placed her coat onto her shoulders, and she put her basket onto her arm. Mother Goat waved goodbye and went to town. The children locked the door like their mother told them.

Before long, a loud knock brought the children to the front door.

“Dear children, it is your mother. I have returned from the market,” said a low, gruff voice.

The Big Bad Wolf had seen Mother Goat walking to the town, and he knew her children must be alone. But the goat children were not fooled.

“You are not our mother!” cried the goat children. “Our mother has a lovely sweet voice!”

So the wolf ran to the market. He got some honey off the beekeeper. He brought the honey to his lips, and drank the golden liquid. The honey softened his voice, and made it high and sweet. The wolf ran back to the goats’ house.

Figure 2: Example of the narrative story.

nine other verbs (5.8 times on average) on the assumption that a prototypical verb of a particular construction could play a path-breaking role in the initial detection of pattern similarities and construction category (Goldberg et al. 2004).

Furthermore, 15 true-or-false comprehension questions were provided at the end of each story to encourage learners to spend enough time reading relevant caused-motion constructions. For example, learners should properly process the typographically highlighted sentence ‘Mother Goat put her scissors, needle, and thread into her pocket’ in the reading text to judge the truth of the statement ‘Mother Goat put her scissors, needle, and thread into her basket’. The instruction lasted about 1 h (i.e., 15 min for each story), and no explicit rule explanation on caused-motion constructions was provided to ensure, as much as possible, that learners remained unaware of what was being learned. The reading activities for the IF group were structured in such a way that caused-motion constructions could be learned implicitly during meaning-oriented activities (e.g., Godfroid 2016; Long 2017).

The MT group was opposite to the IF group in that the rules of caused-motion constructions were explicitly taught by using grammatical terminology (e.g., subjects, objects, and adverbials). For instance, learners in the MT group participated in decontextualized activities such as completing quizzes on the syntactic features of caused-motion constructions (e.g., how many nouns does a caused-motion construction
have?) and putting words into a right order to match given L1 Korean sentences (see Figure 3). In this respect, the MT group received form-oriented instruction where the formal aspects of caused-motion constructions were more emphasized than meaning. The lesson lasted about 1 h.

An instructional focus of the IS group was to balance between meaning- and form-oriented instructions. This goal was achieved by employing the image-schema that underlied caused-motion constructions. Unlike image-schema presented as static pictures in previous research, the image-schema in the present study was created with animation effects to highlight the dynamic nature of caused-motion. For example, learners in the IS group studied a typical caused-motion construction ‘Jane put apples into the basket’ by watching the apples depart from Jane to the basket moving along the arrow (see Figure 4). At the end of the animation, a color-coded word corresponding to each picture was shown one at a time so that learners’ attention could be paid to both form and meaning, which would promote deeper levels of processing (Craik and Lockhart 1972; Leow 2015; Rice and Tokowicz 2020). Replacing the explicit rule presentation in the MT group, the animated image-schema was presented as a part of 1-h activities such as explaining a given image-schema with learners’ own words and drawing the image-schema of caused-motion constructions.

### 4.4 Korean-to-English translation test

A translation test was used to measure a controlled use of caused-motion constructions (e.g., Jahan and Kormos 2015; Wong et al. 2018) due to low English proficiency level of learners. Key content words were supplied as stimuli to elicit target responses. As for each item, the first Korean sentence served as a situational context and learners were asked to translate the second Korean sentence into

![Figure 3: Example of the sentence-analyzing activity.](image)

![Figure 4: Example of the animated image-schema of caused-motion construction.](image)
English by using the words provided (see Figure 5). In the translation test, learners were required to (a) provide a proper preposition and (b) arrange words according to the syntactic pattern of caused-motion constructions.

The translation test consisted of 10 items for the typical and non-typical caused-motion constructions, respectively, and 10 filler items were included to mask the purpose of the test (see Supplementary Material A). One point was given for each correct answer. Inappropriate words arrangement or preposition selection were scored as an incorrect answer with 0 points. Errors not directly related with caused-motion constructions (e.g., spelling, tense, article) were not reflected in scoring. The translation test was conducted three times (i.e., pretest, immediate posttest, and delayed posttest) with a 15-min time limit for each. In order to prevent prior testing experience from providing extra language input that could affect a subsequent test, there was a two-week interval between each test. The items were presented in a randomized order for each testing phase. The experimental design of the study is summarized in Figure 6.

*Figure 5:* Example of the context-embedded Korean-to-English translation test.

*Figure 6:* Overview of the experimental design.
4.5 Analysis

Using the *lme4* package in R, separate mixed-effects logistic models were built in order to investigate the effectiveness of three types of instruction on the learnability of the typical caused-motion constructions (RQ 1) and generalizability to the non-typical caused-motion constructions (RQ 2). Fixed effects included Group (i.e., IF, MT, IS, and control groups), Time (i.e., pretest, posttest 1 and posttest 2), and the interaction of Group by Time. Treatment coding was used for Group and Time with the control group and pretest as a baseline, respectively. By-participant random slopes for Time were included to allow the effect of Time to vary for each participant. By-verb random intercepts were also requested to take the idiosyncrasies of individual target verbs into account. The inclusion of the bobyqa optimizer led to successful convergence of this model. Cronbach’s *α* for the internal consistency of the translation test were 0.95 and 0.90 for the typical and non-typical caused-motion constructions, respectively.

5 Results

5.1 The IS group showed the largest amount of improvement on the learning of the typical caused-motion constructions (RQ 1)

Figure 7 visualizes the results of the translation tests on the typical type of caused-motion constructions. Mean scores of individual participants were dotted around

![Figure 7: Mean scores and 95% CIs of the translation test on the typical caused-motion constructions.](image-url)
95% CIs of the group means across the testing times. It was revealed that all four groups showed improvement on two posttests in general, but they appeared to already be able to produce nearly half of the typical caused-motion constructions correctly at the time of pretest (see Supplementary Material B).

The results of the mixed-effects logistic regression indicate that 74% of the variance was explained by both fixed and random effects combined. It is also confirmed that the simple effects of posttests 1 and 2 on the control group were significant ($\beta = 0.94, 95\% \text{ CI} [0.12, 1.77], z = 2.24, p < 0.05$; $\beta = 1.08, 95\% \text{ CI} [0.29, 1.88], z = 2.69, p < 0.01$, respectively). In addition, the simple effects of the IS group on posttests 1 and 2 were significantly larger than the simple effect of the IS group on pretest ($\beta = 2.51, 95\% \text{ CI} [1.31, 3.71], z = 4.10, p < 0.001$; $\beta = 2.28, 95\% \text{ CI} [1.14, 3.42], z = 3.91, p < 0.001$, respectively) (see Supplementary Material C).

Figure 8 is a graphic illustration of the results in terms of the predicted probability of providing correct responses on the typical caused-motion constructions and 95% CIs by group over time. A considerable degree of overlap among the CIs at the time of pretest suggests that there were no significant group differences in the pretest scores. However, the IS group significantly outperformed the control group on both posttests 1 and 2, as indicated by the non-overlapping CIs between the two groups. More specifically, the learners in the IS group were predicted to produce the typical caused-motion constructions most correctly with about 0.95 (95% CI [0.85, 0.98]) chance on posttest 1 and 0.94 (95% CI [0.85, 0.98]) chance on posttest 2. On the contrary, the IF and MT groups failed to perform better than the control group on two posttests. Although it is true that the predicted probability increased over time for all
four groups, the varying widths of the CIs show the relative influence of each group on learnability. That is, the narrower CIs of the IS group on two posttests, which even do not overlap its own CI of pretest, could be interpreted as substantially reduced gaps between the individuals’ test scores in the IS group over time. To summarize, the image-schema-based instruction showed the largest amount of improvement on the learning of the typical caused-motion constructions and such benefits were maintained over two weeks.

5.2 The IS group showed the greatest degree of generalizability to the non-typical caused-motion constructions (RQ 2)

Figure 9 visualizes the results of the translation tests on the non-typical type of caused-motion constructions. Mean scores of individual participants were dotted around 95% CIs of the group means across the testing times. It was revealed that the IS group scored highest on two posttests, being able to produce nearly half of the non-typical caused-motion constructions correctly. However, the rest three groups showed only a slight upward trend in general (see Supplementary Material D).

The results of the mixed-effects logistic regression indicate that 60% of the variance was explained by both fixed and random effects combined. It is also confirmed that the simple effects of the IS group on posttests 1 and 2 were significantly larger than the simple effect of the IS group on pretest ($\beta = 2.14, 95\%$ CI

![Figure 9](image-url)  
**Figure 9:** Mean scores and 95% CIs of the translation test on the non-typical caused-motion constructions.
[1.34, 2.94], \( z = 5.22, p < 0.001; \beta = 2.27, 95\% \text{ CI } [1.42, 3.12], z = 5.24, p < 0.001 \), respectively) (see Supplementary Material E).

Figure 10 is a graphic illustration of the results in terms of the predicted probability of providing correct responses on the non-typical caused-motion constructions and 95\% CIs by group over time. A considerable degree of overlaps among the CIs at the time of pretest suggests that there were no significant group differences in the pretest scores. However, the IS group significantly outperformed the IF, MT, and control groups on both posttests 1 and 2, as indicated by its CIs that were slightly overlapped at the end or not overlapped at all by those of the rest three groups. More specifically, the learners in the IS group were predicted to produce the non-typical caused-motion constructions most correctly with about 0.56 (95\% CI [0.33, 0.76]) chance on posttest 1 and 0.64 (95\% CI [0.40, 0.82]) chance on posttest 2. However, its CIs with a wide range possibly indicate considerable individual variations in the test scores. On the contrary, the IF and MT groups failed to perform better than the control group on two posttests. To summarize, the image-schema-based instruction showed the greatest degree of generalizability to the non-typical caused-motion constructions on both posttests, but such benefits should be interpreted with caution due to its more widened CIs than those for the typical caused-motion constructions.

6 Discussion

The purpose of the study was to find the optimal balance between attention to form and meaning for L2 construction learning in EFL classrooms. The attentional balance
was manipulated in terms of the differential ratio of attention paid to language form and meaning. This manipulation was achieved through the three instructional groups: the form-oriented MT group, the meaning-oriented IF group, and the IS group, which allowed for simultaneous attention to form and meaning. To summarize the findings, the IS group was the most effective instruction both for the learnability of the typical caused-motion constructions (RQ 1) and generalizability to the non-typical caused-motion constructions (RQ 2). Learners in the IS group were able to produce more correct caused-motion constructions than the other three groups at the translation-based posttests, and such learning benefits were maintained over the 2 weeks after the treatment. It is worthwhile to note that such robustness of the image-schema-based-instruction was more evident in the non-typical caused-motion constructions, which were more conservative and challenging than its typical equivalent in that the non-typical usage was not taught during the learning phase, and its caused-motion sense could not be derived compositionally (Goldberg 1995). In this respect, it could be argued that learners’ constructional knowledge of caused-motion constructions was more deeply tapped into when processing the non-typical usage. These findings suggest that pedagogical values of image-schema could extend to more complex syntactic constructions like argument structures. In contrast, no conclusive evidence for such effects of the input flood and metalanguage was obtained in the current study. The relative advantages of the IS group are explained in comparison to the IF and MT groups.

The effectiveness of image-schema-based-instruction on L2 construction learning could be ascribed to the fact that image-schema might enable learners to pay their attention to form and meaning of the target construction at the same time, which presumably achieved the higher depth of processing (Craik and Lockhart 1972; Leow 2015; Rice and Tokowicz 2020). The intriguing question, then, pertains to the underlying mechanism that allowed for such attentional balance between form and meaning. In terms of the meaning process, image-schema might serve as an intuitive pedagogical tool to foreground embodied meaning of the target construction for L2 learners. This meaning salience might pose less attentional demand required for extracting semantics underlying the target construction. Such freed-up attentional resources could be allocated to process the formal aspects as well, which might have been otherwise too attention-demanding for simultaneous processing of form and meaning.

In contrast to the IS group, the present study failed to show the advantageous effects of the skewed input flood on the learning of argument structure constructions. One possible explanation is that the input flood in the present study might be overenhanced (Han et al. 2008) and perceived unnatural (Szudarski and Carter 2016). The target constructions in the reading texts were manipulated with respect to token frequency (i.e., 108 occurrences for 1 h), frequency distribution (i.e., 23 occurrences of
one prototypical verb), and textual highlighting (i.e., boldfacing). Although these devices were implemented to increase the salience of the target construction, it seemed that the expected effects were canceled out or neutralized by each other’s presence (Han et al. 2008). Given that the mechanism of skewed input is more likely to be activated in implicit learning condition than explicit (Elio and Anderson 1984), possibly overenhanced salience might make it difficult for skewed input to play its facilitative role in forming an initial category of the target construction, which would further serve as a basis for extracting regularities in the input (e.g., McDonough and Nekrasova-Becker 2014; McDonough and Trofimovich 2013; Year and Gordon 2009). In this regard, overenhancing might cause the problem that construction learning in the IF group occurred only at the level of shallow processing (Craik and Lockhart 1972; Leow 2015). That is, although learners might notice the target construction while reading the stories for meaning, they became desensitized to the input too early (Liu et al. 2021). As a result, further attempts to decode such perceived forms might not ensue beyond brief noticing at the surface level (Leow and Martin 2017; Winke 2013). This inefficient processing might be worsened in conjunction with the fact that, unlike the IS group, the IF group should extract the core semantic information all by themselves across multiple exemplars in reading texts. Such cognitive overload might exhaust the attentional resources available for decoding the form of the target construction, being dumped from a further higher level of form processing like rule formation and internalization (VanPatten 2004). Consequently, meaning encoding at the expense of form decoding might inhibit a deeper level of processing. In this sense, the IS group might bring considerable attention-benefits to learners by allowing them to allocate limited attentional resources in a more efficient way.

Similarly, the present study also did not lend support to the use of metalanguage for L2 construction learning. This might be due to the fact that semantically-opaque grammatical terminology (Berry 2006), such as subject or adverbials, might not benefit learners, and the excessive use of such unintelligible jargon could impose unnecessary cognitive burden to decipher the terms themselves (Yang and Ahn 2015; Yoo 2016). This monopoly of attentional resources might induce learners to be obsessed with analyzing only overt surface patterns at the expense of underlying meaning. As a result, a heavy reliance on L1 grammar terminology might hinder simultaneous form-meaning processing, leaving little window of opportunity open for high depth of processing. On the contrary, the language-neutral nature of image-schema could minimize such negative interference of L1 metalinguistic explanation (Morimoto and Loewen 2007), and the rather simple color-coding technique in the IS group worked enough to effectively draw learners’ attention to formal features of the target construction while mapping with its meaning.
7 **Pedagogical implications**

The pedagogical implication of the present study is that the study provides empirical evidence for the potential benefits of image-schema-based instruction on learning constructions as complex as argument structure. In this regard, image-schema can be an efficient pedagogical tool for teachers to manipulate salience of form and meaning of complex constructions. The overuse of metalanguage could place an unnecessary burden on learners, immersing them in analyzing formal features of constructions on the surface level only. Solely relying on frequency-driven construction learning also might not be a realistic approach in EFL classrooms because this type of implicit learning usually involves a very slow and long process (DeKeyser 2003; Leow 2009). Given both quantitatively and qualitatively limited input in EFL classrooms, teachers could take advantage of image-schema as a strategic approach for more efficient L2 construction learning by increasing meaning salience and simultaneously helping learners process formal features within a larger meaning-focused context. In this regard, image-schema, as a reflection of embodied meaning, could reduce learners’ attentional demand needed to extract semantic similarities across multiple exemplars by presenting the core meaning in an intuitively-appealing way. As a result, learners could have more room in their attentional capacity to make inferences on formal features of target constructions. This formal processing could be promoted by the aid of textual enhancement (e.g., color coding). To summarize, image-schema can be pedagogically structured to provide the optimal balance between form and meaning attention by preventing limited attentional capacity from being unnecessarily overtaxed for only one end of the form-meaning continuum.

8 **Future directions and conclusion**

The present research suggests several future directions. First, some of the non-typical caused-motion constructions were not thoroughly corpus-based. Although all the examples employed in the study went through five English L1 speakers’ examination, it was not the case that they unanimously agreed on all the items (e.g., Peter talked himself into the pond). Even though such less authentic examples were included in the study in that they could clearly show what the researcher wants to see, more corpus-based studies might be needed “to investigate the frequency and distribution of constructions across large amounts of authentic language data in order to establish which ones to focus on in the language classroom” (Littlemore 2009, p. 170). Second, the possible interaction between image-schema-based instruction and learners’ individual differences might be an avenue open for
future research. Due to the highly imagery nature of image-schema, there might be a possibility that the effectiveness of image-schema-based instruction could be dependent on whether learners prefer to process information via image or not (Boers et al. 2008). In this regard, it would be worth investigating whether lower imagers are able to benefit from image-schema-based instruction as much as high imagers. Third, it could be argued that image-schema could cause overgeneralization of learners’ constructional knowledge. For example, although it is not the case for the present study, there still exists the possibility that learners might overgeneralize rules by simply plugging any types of verbs into argument structure constructions in an effort to be communicatively functional. More research is needed to understand how such overgeneralization process could be constrained with pedagogical intervention such as corrective feedback (e.g., Wong et al. 2018).

With respect to L2 construction learning in EFL classrooms, the study could contribute to debates on varying level of attention paid to language form and meaning. The study showed that image-schema-based instruction could be one viable option for L2 teachers to achieve the optimal balance between attention to form and meaning for L2 pedagogy. Further studies are warranted to draw any conclusions about the facilitative role of image-schema in the learning of other types of constructions and different methodological factors that could mediate its effectiveness.

Acknowledgments: This research is a part of the author’s M.A. thesis. I am deeply grateful to Dr. Sang-Ki Lee for his comprehensive guidance and Drs. Hyunsong Chung and Ji-Hye Kim for their insightful comments. I also express my sincere thanks to Dr. Shawn Loewen, Dr. Ryo Maie, Matt Coss, Vashti Lee, Joanne Koh, Jieun Kim and Botagoz Tusmagambet for their invaluable feedback and suggestions on earlier versions of the manuscript.

References


**Supplementary Material**: This article contains supplementary material (https://doi.org/10.1515/iral-2023-0015).