University of Southampton

Faculty of Arts and Humanities

Languages, Cultures, and Linguistics

Genericity Acquisition by Arabic-speaking Learners of English: An Intervention Study

by

Manal Ali A Jallalah

ORCID ID [0009-0008-8168-2846]

Thesis for the degree of Doctor of Philosophy

June 2025

University of Southampton

Abstract

Faculty of Arts and Humanities

Department of Languages, Cultures and Linguistics

Doctor of Philosophy

by

Manal Ali Jallalah

Genericity, a universal semantic property, encodes complex form-meaning mappings. Learning genericity in a second language (L2) is challenging due to the nature of mapping the semantic meanings and their morphophonological expression, as well as the learner's native language (L1) and the L1-L2 crosslinguistic differences. The literature suggests that Generative Second Language Acquisition findings can be beneficial for the L2 classroom and, therefore, call for bridging the gap between GenSLA and L2 instruction by applying SLA findings in the L2 classroom (Slabakova, 2019; Whong et al., 2014; Marsden, 2018; Ionin & Montrul, 2023). This thesis seeks to contribute to bridging this gap by investigating the effect of addressing the learnability concerns predicted by acquisition research through instruction on the acquisition of genericity by Arabic–speaking learners of English in a classroom context, within the framework of the Feature Reassembly Hypothesis (FRH) (Lardiere, 2009) and Slabakova's (2009) cline of difficulty and Bottleneck Hypothesis. To this end, it conducts a cross-linguistic analysis of how genericity works in English and Modern Standard Arabic in preparation for predicting the difficulty that Arabic-speaking learners of English may face. Then, it tests how teaching intervention informed by SLA findings can affect learning generic form-meaning mappings in characterising and kind generic meanings.

This thesis predicts that mapping indefinite singulars and bare plurals onto characterising generics is challenging and that L2 learners will use the L1 form-meaning mappings with this meaning. Also, it predicts that the participants may face similar challenges in mapping bare plurals onto kind generic meanings. Mapping definite singulars onto kind generics is predicted to be less challenging in light of the similarity between the L2 learners’ L1 and L2 in this condition. Finally, instruction is predicted to support the L2 learners’ acquisition of generic form-meaning mappings. This thesis follows an intervention study design to test these predictions with a pretest, intervention, post-test and delayed post-test. The study included two groups of low-intermediate L2 learners divided into experimental and comparison groups (total n = 64), and a native control group who provided a baseline (n = 20). The experimental group received instruction on genericity for eight weeks. The study tasks included a written elicited production task, an acceptability judgement task with contexts, and a forced-choice task.

The pre-test results revealed that Arabic–speaking learners find the generic form-meaning mappings challenging even when mapping definite singulars to kind generic meaning where the L1 and L2 are similar. The immediate post-test results revealed a significant improvement in the experimental group’s mappings of bare plurals to both generic meanings and mapping definite singulars onto kind generic meaning. However, even after instruction, mapping indefinite singulars to characterising generics remained challenging for the experimental group. The comparison group did not show improvement in all conditions in the post-test. The experimental group maintained the improvement in mapping bare plural to kind generic meaning in the three tasks after being tested twelve weeks later. This study suggests that explicit instruction that considers the reassembly requirement in a learning context in depth and maximises the L2 learners’ engagement with input through practice might positively impact feature reassembly in L2 learning. The results support FRH and BH and highlight the benefit of using SLA findings in operationalising instruction to support L2 acquisition in the L2 classroom-based acquisition context.

Table of Contents

Table of Tables

Table of Figures

Research Thesis: Declaration of Authorship

Print name: Manal Ali A Jallalah

Title of thesis: Genericity Acquisition by Arabic-speaking Learners of English: An Intervention Study

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission .

Signature: Date:4/6/2025

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Roumyana Slabakova, for her invaluable support and guidance throughout the journey of writing this thesis. Her exceptional teaching, encouragement, feedback, and unwavering support over the years have been instrumental. I am indebted to Roumyana for her time and effort in reading various draft chapters of this thesis and for the extremely thorough and helpful feedback she provided. Additionally, her patience, understanding, and administrative support were crucial in helping me navigate the complexities of this project. I am also grateful for her guidance in professional development and academic growth, providing opportunities for advancement and introducing me to essential readings. Words cannot fully convey my appreciation. This thesis would not have been possible without her guidance and support. I am likewise thankful to my second supervisor, Glyn Hicks, for his valuable insights and support throughout this process. His feedback during progression reviews and research instrument development was a great help. I feel very fortunate to have had such a supervisory team.

In addition to my supervisors, I owe particular thanks to Laura Domínguez for her teaching during my first year and her guidance during my second progression reviews. Continuous discussions with Laura, Roumyana Slabakova, and other members of the CLLEAR group have significantly enhanced my research skills and critical thinking. Being a member in the CLLEAR reading group was an enjoyable and fruitful experience. I would like to express my appreciation to all the members of the CLLEAR reading group at Southampton. The insightful discussions within this group have greatly contributed to my professional development and made my PhD journey more rewarding. Additionally, I would like to extend my sincere gratitude to the examiners of this thesis, Dr. Sara Rule and Dr. Dora Alexopoulou. I deeply appreciate the time, expertise, and thoughtful feedback you have provided. Your insightful comments and constructive suggestions have not only contributed to the refinement of this thesis but will also be invaluable for my future academic development.

I am profoundly indebted to my parents for their unwavering emotional support, unconditional love, and patience. Words cannot adequately convey my gratitude for the care and support they have provided throughout this journey. Their belief in me has been a constant source of strength and motivation. From countless daily conversations to endless encouragement, they have been my rock. I am eternally grateful for everything they have done and continue to do for me. I also extend my heartfelt thanks to my siblings for their continuous support and encouragement. Furthermore, I would like to acknowledge my partner, Waleed, for his companionship during a significant part of this journey. I am grateful for whatever support offered and for the meaningful experiences we shared along the way.

I could not have undertaken this journey without the unwavering joy and light of my life, my son, Abdulwahab (Hobi). His smile has brightened my hardest times, and his presence has been an ongoing lesson in time management and project planning. His love, laughter, and constant need to play have helped me become more organised, making this journey not only fruitful but also joyful and filled with wonderful memories. This achievement is as much his as it is mine.

I would like to extend my heartfelt thanks to Suha, who has been with me through the highs and lows of this journey, from the very beginning to the submission of this thesis. Her encouraging words, patient listening, and understanding have been invaluable. We have celebrated each small achievement together, leading to the completion of this thesis. I am also grateful to my colleagues at Southampton. Special thanks to James Corbet, Lewis Baker, E. Jamieson, Eleonora, and Ania for their support, for answering many of my questions, and for sharing their experiences. Discussions with them were a pleasure.

Lastly, I would like to thank the English Language and Translation department at King Khalid University for their administrative support during the intervention. Their understanding and willingness to provide resources were incredible. Finally, I am thankful to all the participants of this study for being helpful and showing a high sense of commitment.

Definitions and Abbreviations

ACP .... Article choice parameter

AJT .... Acceptability Judgment Task

BH .... Bottleneck Hypothesis

DP .... Determiner phrase

EWPT .... Elicited written production

FAFT .... Full Access\Full Transfer

FH .... Fluctuation Hypothesis

FRH .... Feature Reassembly Hypothesis

GaD .... Generics-as-default

GenSLA............................ Generative Second Language Acquisition

ISLA .... Instructed Second Language Acquisition

L1 .... Native Language

L2 .... Second Language

MP .... Minimalist Program

MSA .... Modern Standard Arabic

NP .... Noun Phrase

P&Ps................................ Principle and Parameters

PI .... Processing instruction

PLD................................. Primary Linguistic Data

RDH................................ Representational Deficit Hypothesis

SLA .... Second Language Acquisition

TL .... Target language

TVJT .... Truth Value Judgement Task

UG .... Universal Grammar

Chapter 1 Introduction

1.1 L2 learnability in Generative SLA

Second language acquisition (SLA) refers to learning a second language (L2) after the native language (L1) has been learned (Gass et al., 2020). It is assumed that learning an L2 involves constructing a mental grammar system that enables L2 learners to communicate and comprehend the target L2 successfully. Gass and Mackey (2012, p. 94) define the interlanguage system as “the mental system developed by L2 learners that enables them to produce and understand utterances of the [target language] TL”. White (2020) clarifies that interlanguage refers to the L2 learner's linguistic competence, which involves unconscious mental representations. Combining information from multiple sources, including the learner's L1, Universal Grammar (UG), and L2 input, is necessary to construct this mental grammar (White, 2020; Rothman & Slabakova,2018). In this section, I give an overview of what acquiring an L2 means from various epistemological vantage points, elaborating on the Generative SLA perspective (GenSLA) as the epistemological paradigm for the acquisition part of this study.

Different theoretical perspectives have served as the epistemological foundation for SLA studies throughout time. These include the generative perspective, which is grounded in linguistic theory; the Interactionist Approach, which emphasises learner-related variables such as language, culture, emotion, and cognition; and the Emergentist Approach, which posits that SLA emerges “solely based on the linguistic experience and is in a sense created by that usage.” (Slabakova, 2016, p.7). In this research, I employ a Generative Approach to L2 acquisition (GenSLA).

GenSLA perceives language as a system of signs and the rules governing how those signs combine; a grammar that generates all the acceptable sentences in a language while excluding unacceptable ones from that language. A sign is a symbolic representation of form–meaning mapping. The human mind stores information about this sign system as a complex web of linguistic representations (Slabakova et al., 2020). Rothman and Slabakova (2018) discussed the main tenets of GenSLA. According to them, the aim of GenSLA research is to elucidate learners’ mental grammar. GenSLA seeks to explore L2 learners’ mental grammar, how L2 learners can acquire L2, what makes the acquisition process harder for L2 learners, and where the difficulty lies in cases where acquisition fails (Rothman & Slabakova, 2018). White (2020) summarises the main observed findings about L2 acquisition in this perspective, among which is that exposure to input is essential to L2 learning; L2 learners can know more than what the input gives, such as knowing about ungrammaticality based on grammaticality in the input, there are limits of the L1 and input effect on L2 acquisition, and that second language learning is different across different linguistic modules and interfaces between these modules. To evidence the latter, it has been found that “syntax acquisition is largely successful in contrast to inflectional morphology” (White, 2020, p. 32). In addition, GenSLA describes acquisition as developmental stages in the learner’s mental grammar. It uses formal linguistic theory to predict the factors that affect L2 learnability. In GenSLA, linguistic competence is explored by adopting various performance measures to capture the essential characteristics of the L2 learners’ mental grammar (White, 2020).

What we know about L2 learnability, from a GenSLA perspective, comes from accounts of what language is and how it is acquired in the generative linguistic framework developed by Chomsky (e.g., 1995, 2000,2005) and the evolving GenSLA hypotheses that make use of these accounts. According to Chomsky (2005), the emergence of language is the result of the interaction of three factors: the genetic endowment (UG), learner experience or primary linguistic data (PLD), and a non-linguistic cognitive factor (general learning principles and strategies). In a feature-based model of L1 acquisition, children must select which features of their genetic endowment to use to map these features onto the semantic concepts and lexical items of the target language, taking the PLD into account. The job of the L2 learner is twofold: choose novel L2 features (for those not chosen in the L1) and reassemble familiar L1 ones (in the case where features are selected differently in the L1 and the L2).

Different assumptions have been made about L2 learners depending on various positions towards UG access. On the one hand, some linguists believe that adult L2 learners do not have access to UG after the critical period. Advocates of that view have proposed the Representational Deficit Hypothesis (RDH), arguing that some features of the L1 become fixed after a critical period and that adult L2 learners are unable to adjust these features in the acquisition of an L2 (e.g., Hawkins & Chan, 1997). On the other hand, other linguists adopt a full-access view of UG, arguing that L2 learners can acquire the features that were fixed in their L1. The Feature Reassembly Hypothesis (FRH) (Lardiere, 2009) and the Bottleneck Hypothesis (BH) (Slabakova, 2009, 2019) are examples of feature-based hypotheses that investigate the L2 learning task and the sources of difficulties in L2 acquisition, assuming full access to UG. In what follows, I give an introductory account of FRH and BH, and in 3.2.2, I delve further into the assumptions underpinning these hypotheses and their relevance to this study.

Lardiere (2009) argues that acquiring an L2 grammar goes beyond the availability of features for selection from UG; rather, it requires considering the ways in which these features are bundled into lexical items or functional categories, as well as the language-specific conditions for phonological feature realisation. Lardiere (2007) stated that reconfiguring the bundles of formal and semantic features and determining the conditions under which they are expressed morphophonologically are both necessary steps in learning L2. Therefore, the acquisition task for L2 learners includes finding similarities between the functional meaning in L2 and L1 and mapping this meaning from L1 to L2 lexical items. Then, learners need to start checking the features and assembling them based on the L2 evidence in the input. She adds that the acquisition challenge varies depending on the context. That is, if a feature of the L2 is available in the L1, but the context of that feature is not related to its L2 context, acquiring it is more challenging, though possible.

Slabakova (2009) explores the successes and challenges faced by L2 learners in acquiring the semantic domain, introducing the Bottleneck Hypothesis. This hypothesis posits the universality of semantic primitives and syntactic operations. However, difficulties in acquisition arise when these two domains fail to interface seamlessly, particularly in mapping grammatical meaning and morphological realisations. As Slabakova (2009, p.282) asserts, “learning a second language entails mastering the new configurations in which various…features are mapped onto the target language’s inflectional morphology.” Slabakova (2019) catalogued some factors that complicate the L2 acquisition of functional morphology (other than L1 transfer): the morphosyntax–semantics mismatch between L1 and L2; the need for feature reassembly; L2 functional morpheme redundancy; the opacity of form–meaning mapping; and frequency of use. Chapter 3 elaborates on this hypothesis.

With this in mind, I define L2 acquisition in this study as the learner’s ability to develop an L2 mental grammar as they learn L2 form-meaning mappings and overcome any difficulty that arises during learning. [1] In this study, SLA focuses on the contrast within the learners' mental grammar and whether they overcome the L1 influence and acquire the L2 form–meaning mapping in their development of the linguistic property under investigation. In addition, the focus is on what difficulty they may face based on theoretical linguistic analysis of this property.

[1] Krashen (1981) proposed that acquired L2 knowledge, which becomes part of a learner’s unconscious competence, is distinct from learned L2 knowledge, which is gained through conscious study and effort. A fuller discussion of the different types of knowledge is provided in Chapter 4. Since GenSLA research and instruction share the ultimate goal of improving the implicit acquired knowledge and since the views on the possible move from one knowledge type to the other is assumed as will be shown later, the terms L2 learning and L2 acquisition are used interchangeably in this study, following Ionin and Montrul (2023, p.15).

In this sense, this study assumed a GenSLA perspective in accounting for the acquisition and learnability of genericity.

1.2 Research context and research problem

In this section, I establish the theoretical and empirical context which informs this thesis and then introduce the linguistic property and research problem. It discusses the syntax–semantic interface in L2 acquisition and the difficulty arising from L2 form-meaning mappings from a GenSLA perspective. It introduces genericity form-meaning mappings to exemplify an L2 interface. Then, it highlights recent calls in the literature that consider relating L2 instruction to the GenSLA-based account of acquisition difficulty at the interface level, especially in an L2 classroom-based acquisition context.

1.2.1 Form–meaning mappings in L2 acquisition

Learning a language involves pairing linguistic forms with meanings to communicate and understand messages. Meanings, whether lexical (concepts), semantic (phrases), grammatical (inflectional morphemes meanings such as present tense), or pragmatic (implied), are all universals, and languages may differ in the ways of encoding these meanings (Slabakova,2016). The study of the syntax-semantics interface focuses on the elements of syntactic structure that influence meaning composition and the aspects of meaning that systematically impact syntactic structure (Hackl, 2013; Ionin et al., 2024). White (2011) clarifies that in L2 acquisition, L2 learners need to grasp the linguistic phenomena related to an interface and may represent these differently from native speakers. This does not necessitate acquiring the interface itself (White, 2011). According to Lardiere (2000), it is the L2-correct mappings that the learners need to acquire, and these mappings prove problematic for some interfaces.

Slabakova (2013) argues that the essence of language acquisition is the acquisition of form-meaning mapping in functional morphology, which captures the syntactic and semantic cross-linguistic differences between languages. “When universal grammatical meanings are captured by different grammatical morphemes…it is not enough to just learn the new words and put together the L2 sentence. L2 learners have to do some special restructuring in their grammar.” (Slabakova, 2016, p.290). In other words, when form-meaning mappings in the learner’s L1 and L2 are not the same, the L2 learner needs to go through functional morphology, i.e., acquire appropriate form-meaning mappings, to acquire syntax and semantics in the L2. As a result, functional morphology is located at the bottleneck of L2 acquisition, as mentioned above. Positioning this functional morphology in the bottleneck of L2 acquisition resulted from the fact that it “hosts a bundle of morphosyntactic and semantic features, which may or may not be overtly expressed by a functional morpheme” and therefore influences the acceptability and meaning of the entire sentence (Slabakova, 2019, p. 2).

White (2011) provides an example of this interface by examining the acquisition of English articles by L2 learners whose L1s lack articles, such as Mandarin, Russian, or Turkish. The universality of semantic primitives suggests that these languages encode definiteness but do not use articles as the morphological realisation of this meaning. According to her, learners from these L1 backgrounds must understand that the [± definite] feature is mapped to articles in English in order to acquire English articles successfully. Due to space limitations, for empirical evidence on the difficulties arising from form-meaning mappings in different interfaces, see White (2011).

White (2011) argues that the linguistic properties related to language module interfaces reflect varying difficulty levels during the acquisition process or at the end state. This difficulty varies across interfaces and linguistic properties. Regarding the explanation of difficulties arising from interfaces, White highlights two main perspectives in the literature. Some scholars attribute problems at an interface to the learners’ underlying linguistic representations. In contrast, others attribute them to processing pressure, referring to the cognitive load involved in learning and using the new language. White (2011) concludes that different interfaces and linguistic properties, acquired at different proficiency levels, have distinct sources of difficulty and, therefore, require multiple explanations.

1.2.2 Genericity as an example of form-meaning mapping at an interface

Consider genericity as an example of an interface in English L2 acquisition. The interface in genericity is between morphology (articles), clause-level semantics, and knowledge of the world (discourse). All languages can express genericity; hence it is a universal linguistic meaning (Carlson, 2011). However, linguistic differences can be seen in that languages may differ in mapping the generic semantic features to different linguistic forms. As far as English is concerned, generic form-meaning mappings are not straightforward and, therefore, are complex and interesting. Krifka et al. (1995) describe two types of generic meaning: kind generics within the noun phrase and characterising sentences[2].

[2] In the literature, characterising sentences and generic sentences are terms referring to sentence-level generics. Krifka et al. (1995, p. 3) included other terms, such as "gnomic", "dispositional", "general" or "habitual". Some researchers use the term "I-generics". Moreover,

A characterising generic sentence refers to the general regularity of the predicate, as in (1a, b), and genericity is expressed through the habitual or stative verbs, as in (1c and 1d).

(1) a. Green lamps are relaxing. (generic readings)

b. A green lamp is relaxing.

c. John smokes.

d. A potato contains vitamin C.

(2) The green lamps are not working. (episodic reading)

The example in (1a) encodes that a general characteristic of green lamps is relaxation. Hence, “green lamps” refers to green lamps in general rather than specific green lamps in the context, as in (2). The example in (2) expresses maximality and refers to all the green lamps in front of the speaker in the context, hence denoting an episodic reading. In English, the characteristic generic meaning is mapped onto different forms: bare plurals, as in (1a) and indefinite singulars, as in (1b and d). In this type of genericity, the generalisation is on the sentence level and there are no restrictions on the NPs[3] in the subject position, according to Krifka et al. (1995), as shown in (1c) where the sentence is giving a generalisation about the subject (John), who refers to an existing individual. Moreover, although limited with this reading, definite singulars can be mapped to this meaning in the existence of a well-defined kind as in (3), which gives a generalisation about coke bottles as having a narrow neck (Krifka et al., 1995, p. 11).

(3) The Coke bottle has a narrow neck.

Turning to kind generics, in English, NPs denoting a kind generic reading are selected by kind-predicates (such as become extinct/ be widespread, among others). Sentences that express kind readings encompass the entire genus (Krifka et al., 1995), and genericity is expressed at the NP level as shown in (4):

(4) a. The potato was first cultivated in South America.

b. Potatoes were first cultivated in South America.

the terms "kind-referring", "D-generics", and "generic NPs" are those that refer to NP-level generics. This study will use the terms "characterising generics" and "kind generics" to maintain clarity.

[3] Throughout this thesis, I used the term NP to refer to both NP and determiner phrase (DP). When the discussion is on theoretical proposals on generic interpretations, I clarify whether the nominals are considered as DP or NP, following Ionin et al. (2011a).

English maps this meaning onto bare plurals (4b) and definite singulars (4a). However, both forms can also express episodic reading, which refers to making a statement about individuals existing in a particular context, as in (5).

(5) a. Dogs are barking.
b. The dog is barking.

Although these meanings are universal, no linguistic construction is assigned to the expression of the universal meaning of genericity, and generic form-meaning mappings differ among different languages (Chierchia, 1998). In contrast to English, Modern Standard Arabic (MSA)—the L1 of the learners in this thesis—uses definite singulars and definite plurals to express characterising and kind generic readings as in (6).

(6) a. al–kalb–u y–anbaḥ –u
the–dog.SG-DEF bark.
'The dog barks./The dog is barking.'
b. al –kilaab –u t –anbaḥ –u
the –dogs.PL –DEF bark
'Dogs bark./Dogs are barking.'

According to Chierchia's (1998) Nominal mapping parameter, MSA NPs are predicative. Hence, they require an article or other determiner to function as arguments. Therefore, MSA disallows bare plurals with a generic interpretation, as in (7). Later in Chapter 2, I provide a detailed feature-based contrastive analysis of genericity in both languages.

(7) *dainasour –at –u mungaridh –at –un
dinosaurs.PL extinct.
'Dinosaurs are extinct.'

An L1 child can successfully acquire this complex form-meaning mapping at the age of four (Gelman & Bloom, 2007). In L2 acquisition, as suggested by White (2011), different sources contribute to the acquisition: how L1 form-meanings mappings differ from the L2 mappings, whether L1-L2 are similar or different, other L2 meanings mapped to the forms in L2, L1 transfer and level of proficiency as well as the quantity and quality of the L2 input. Therefore, the acquisition of form-meaning mappings is challenging in L2. Within Slabakova's BH and Lardiere's FRH, different factors affect the difficulty in acquiring L2 form-meaning mappings, including L1 transfer, the mismatch between mappings in L1 and L2, the complexity of the L2 input, and the quality and frequency of the evidence for the form–meaning mapping in the input.

In Chapter 3, I review key studies that consider difficulty in the acquisition of generic form-meaning mappings.

In light of these factors, a growing number of academics are advocating for applying the findings of GenSLA to facilitate acquiring L2 in classroom settings where L2 learners lack natural exposure to the target language and are limited to the instructed input in their L2 acquisition journey (Whong et al, 2013; Slabakova, 2013; Whong et al., 2014; Marsden, 2018). Acknowledging that there is no direct relation between the research agenda of GenSLA and that of L2 instruction, both paradigms seek a shared goal of developing the L2 learners' linguistic competence (Whong et al., 2013). In the following, I provide an introductory conceptualisation of L2 instruction and how it can benefit from GenSLA and vice versa. Later in Chapter 4, I elaborate on the discussion of the impact of L2 instruction, the explicit/implicit debate, and provide a detailed account of the impact of L2 instruction on the acquisition of genericity.

1.2.3 L2 instruction and its relation to GenSLA

The term 'instruction' involves various elements working together to achieve effective communication in a second language. It includes cognitive, social, and methodological elements. Marsden and Slabakova (2018) introduced a special issue about grammatical meaning and L2 classrooms. They highlighted that generative linguistics, which focuses on linguistic properties and language structure, has informed GenSLA. Therefore, GenSLA views L2 instruction as engaging with the grammar elements of language teaching and learning. They also emphasised that while grammar is important, other aspects of language teaching are equally valuable, and their view does not prioritise grammar over these other elements. Marsden and Slabakova (2018) highlight the importance of considering the implications of GenSLA findings for L2 classrooms. Marsden and Slabakova (2018) emphasise that while implicit learning is vital, explaining the rules and illustrating their application in various contexts can benefit L2 learners. They conclude with a call for cooperation between L2 teachers and SLA researchers to bridge the gap between theory and practice.

Recent years have seen interest in bringing together linguistic theory, SLA and L2 classroom research. Ionin and Montrul (2023) bring together linguistics, SLA and L2 classroom research in a book that aims to strengthen the link between the three fields. They provide a foundation for theoretically informed intervention research, which refers to intervention studies in which the starting point is the linguistic analysis of a linguistic property, and the focus is the learnability of the selected structure and the impact of instruction on its learnability. There is growing interest in discussing the links between these three fields (Whong et al.,2013; Whong et al., 2014; Marsden & Slabakova, 2018). A discussion of the direction and benefits of this link is provided by Whong et al. (2013, 2014), who suggest that theory–driven evidence from GenSLA—the cognitive process in L2 learning, the difficulty, and the role of UG, input and L1 in L2 acquisition—can inform L2 classrooms by helping L2 teachers understand how an L2 mental grammar is constructed and design teaching to support this development. Moreover, L2 instruction can also inform GenSLA; L2 classrooms provide rich data about L2 learners' mental grammar, which can be used to elaborate discussion on which aspects of UG are most challenging, which models are validated or which theoretical accounts need refinement. Although L2 instruction and GenSLA are different fields, the cooperation between these fields bridges the gap between GenSLA theory and practice. Therefore, there is a call in the literature for intervention research that can contribute to bridging the Gap between GenSLA theory and practice.

1.2.4 Research problem

The context discussed so far highlights form-meaning mappings as an area of difficulty for L2 learners and that the transfer of L1 mappings and how they differ from the L2 ones contribute to this difficulty. It shows that genericity is a universal property that shows three-way interface between syntax, clause–level semantics and discourse, and that English and MSA differ in mapping generic meanings to grammatical representations. Although English and MSA express genericity through articles as determiners in the noun phrase (NP), the NPs to which genericity is mapped differ in both languages. There is abundant evidence that the English article system is particularly challenging for L2 learners of English (Ionin et al., 2004; Snape, 2008; Sabir, 2015, among others). Previous research has established that the L2 acquisition of generic interpretations is challenging (Ionin et al., 2011b; Ionin & Montrul, 2010; Snape et al., 2013; Snape, 2018), especially when the L2 acquisition context is the L2 classroom (Snape & Yusa, 2013; Umeda et al., 2019). The investigation on the acquisition of generic form–meaning mapping difficulty has been extended to Arabic-speaking learners of English L2 (Alzamil, 2019; Abumelha, 2019; Hermes, 2020; Aboras, 2020; Sabir, 2015).

To date, there is a paucity of intervention studies which contribute to bridging the gap between GenSLA and L2 instruction considering genericity as the linguistic property. Regarding the acquisition of genericity by L2 learners, only four theoretically informed intervention studies exist. These intervention studies have shown various focuses and inconsistent results. In particular, intervention studies by Snape and Yusa (2013) and Umeda et al., (2019) were informed by linguistic theory and focused on the development of the mental grammar in L2 learners from an L1 without articles, Japanese. Studies that consider L1s with articles focus on

Arabic–speaking learners' acquisition of English article semantics, including genericity, and have compared explicit and implicit instruction (Sabir, 2015) and implicit and explicit input (Abumelha, 2019). Moreover, while some studies reported no impact of the intervention on acquiring generic form-meaning mappings (Snape & Yusa 2013; Sabir 2015), others showed a short-term impact of instruction (Umeda et al., 2019; Abumelha, 2019). Abumelha (2019) reported a month-long term impact of instruction on one condition, mapping bare plurals to generic interpretations. Umeda et al. (2019) reported that instruction on genericity did not show a long-term impact when testing the sample, a year after the intervention. They suggested that instruction did not improve the L2 learners' implicit knowledge of genericity.

As I detail in 4.4, different design-related factors are highlighted for further investigation in these studies, such as the length of genericity instruction and the need to use multiple measures to tap into different knowledge types. Thus far, conflicting results from studies suggest the need for new investigations that consider whether and how a more extended instructional intervention—centred on GenSLA findings on what is easy and what is challenging to acquire in generic form-meaning mappings and a linguistic theory-based cross-linguistic analysis of the L2 and the L2 learners' L1—affect the L2 learners' mental representations, which is the focus of this thesis as the following section explains further.

1.3 The empirical Study

1.3.1 Research aims and questions

The main objectives of this study are i) to explore what is easy/difficult in acquiring genericity form-meaning mappings for Arabic-speaking learners of English, and ii) to determine whether and how acquisition-informed instruction supports this acquisition, building on theoretical and empirical findings on how English and MSA express genericity. In particular, this thesis sets out to explore the interplay between genericity and the five count nominal forms (definite, indefinite, and bare singulars, and bare and definite plurals) that are used in kind generics and characterising generic sentences in English and MSA, and the effect of instruction on addressing the learnability concerns predicted by acquisition research within the framework of Lardiere's (2008, 2009) FRH and Slabakova's (2009) cline of difficulty in acquisition and BH. In this way, the goal of this study is different from purely pedagogical studies because genericity as a linguistic property is the starting point of this intervention study, which included an application of research on how L1 differs from L2, how the L1 affects L2 acquisition, and what is easy and hard to learn as the basis for the content and order of the instruction intervention to enable supporting the L2 learners' mental grammars, hence informed by GenSLA. The central questions this thesis asks are:

RQ1: Do intermediate Arabic-speaking English learners demonstrate target-like mapping of English articles to kind and characterising generic meanings? If not, what is the difficulty and how can it be accounted for according to FRH and BH?

RQ2: After being exposed to 8-week acquisition-informed explicit instruction on English genericity form–meaning mappings, can the experimental group improve their form-meaning mappings in kind and characterising generic meanings compared to the comparison group? Is any such improvement retained three months after instruction?

To address these questions, this study follows an intervention study with a quasi-experimental design, using pre-test, intervention, post-test and delayed post-test to investigate the impact of instruction. It started with an empirical account of how native speakers of English and MSA express genericity using Arabic and English acceptability judgment tasks (AJT) to account for variability in the L2 input, validate the semantic literature on MSA and make predictions on the difficulty in light of the FRH and BH. The sample included two groups of low-intermediate L2 learners (experimental and comparison groups, n = 64) and native controls who formed the baseline group (n = 20). After conducting the pre-test, the instruction intervention lasted eight weeks, in which the experimental group received instruction on genericity form-meaning mappings, and the comparison group received no instruction on genericity but enrolled in their regular grammar classes. After the intervention, a post-test was carried out, and three months later, the delayed post-test was conducted. This thesis used data from an elicited written production task (EWPT), an acceptability judgement task with contexts (AJT), and a forced–choice task (FCT) in each testing session, in addition to a language background questionnaire and a proficiency measure. The study predictions are discussed in Chapter 5, and the intervention study methodology in Chapter 6.

1.4 Research significance and contributions

GenSLA empirical research has had implications for formal linguistic theory and has had rich results on the accounts of the mechanism of L2 acquisition, investigating the role of UG, the L1 and L2 input in L2 acquisition, and which elements of grammar are easy and which are challenging in L2 acquisition. Concerning the question of the acquisition of genericity, genericity is expressed through functional morphology, but its interpretation is linked to discourse and world knowledge. As a result, genericity is complex and challenging for English L2 learners to acquire. There is an ongoing debate on which interfaces are more difficult to acquire in L2, and a call for more investigations into the L2 acquisition of linguistic properties that include interfaces (White,2011). The complexity of the interface itself served as inspiration for this research, the findings of which are expected to contribute to the existing debate by adding to the existing empirical evidence on interface-based difficulty in genericity acquisition.

GenSLA has extensively investigated the mechanism of L2 acquisition, L2 learnability issues and the role of input and L1 in acquisition. Therefore, it has the potentials to inform L2 instruction and to be informed and refined by data from L2 classrooms. However, there is a gap between theory and practice. Existing empirical investigations have achieved little to bridge the gap between GenSLA and L2 instruction regarding the acquisition of genericity as stated in 1.3 above. This thesis is an interdisciplinary study that focuses on relying on acquisition research findings on how genericity form-meaning mappings are acquired and what is easy and what is difficult in the learnability of genericity from a feature-reassembly perspective to inform the teaching intervention that aims to support the acquisition of generic from-meaning mappings. Therefore, it is hoped that findings from the current study can contribute in some way to bridging the gap between GenSLA and L2 instruction.

In addition to the theoretical contributions, this thesis is anticipated to have a practical significance for L2 intervention research and L2 instruction. This thesis provides an opportunity to advance the understanding of theoretically-informed instruction design by providing an empirical investigation that could highlight areas of challenge in conducting intervention studies. It hopes to give insights into methodological considerations and challenges that support L2 instruction. Its findings are expected to be of significance to L2 linguistics researchers, L2 teachers, and textbooks and curriculum designers as it aims to give insights into how genericity can be taught to L2 learners.

1.1 Organization of the thesis

The overall structure of this thesis takes the form of eight chapters. Chapter 2 provides a feature-based contrastive analysis of genericity form-meaning mappings in English and MSA. Chapter 3 presents the theoretical foundation of L2 acquisition from a Minimalist, feature-based perspective elaborating on the Feature Reassembly Hypothesis and the Bottleneck Hypothesis and their predictions for the L2 acquisition of form-meaning mappings. It reviews key GenSLA-based empirical studies on the acquisition of genericity by L1 children and L2 learners. Chapter 4 provides a detailed discussion of L2 instruction and how it can be theorised based on SLA research findings. In addition, it reviews intervention studies on the acquisition of genericity by L2 learners of English. Chapter 5 provides an empirical account of how native speakers of MSA and English express genericity. Chapter 6 outlines this study's methodology. Chapter 7 analyses the data gathered and presents the results. Chapter 8 interprets the results in light of the research questions and predictions. It concludes with a discussion of the study's implications, limitations, and directions for future research.

Chapter 2 Genericity in English and Modern Standard Arabic

2.1 Introduction

As shown in the previous chapter, this thesis investigates how Arabic-speaking learners of English acquire genericity form-meaning mappings and how explicit teaching impacts the acquisition of these mappings. Therefore, a clear account of genericity as a linguistic property and how it is expressed in both the learners' L2 and their L1 is essential for predicting potential challenges learners may face and informing instructional interventions. This chapter offers a linguistic account based on the semantic literature on genericity and how it is expressed in English and MSA.

This chapter is composed of four main sections. Section 2.2 defines genericity, presents diagnostic tests that differentiate between generic and non-generic meanings, and describes the main characteristics of generic meaning. Section 2.3 discusses the semantic interpretations of count noun phrases (NPs) in subject positions in both English and MSA. Section 2.4 examines the semantic account of cross-linguistic differences in the expression of genericity between MSA and English, employing the semantic frameworks of Chierchia (1998) and Dayal (2004).

2.2 Generic vs. particular readings: diagnostic tests and main characteristics

Carlson's (1977) view of the ontology of kind marked the beginning of scholarly interest in genericity. The literature shows different accounts for genericity; linguistic, psychological and philosophical. The psychological and philosophical accounts are beyond the scope of this thesis. However, more information about psychological and philosophical accounts can be found in Lazaridou-Chatzigoga (2019), and Nickel (2016). From a linguistic perspective, genericity refers to two distinct linguistic phenomena; one within the noun phrase, known as kind generics, and one within propositions that report general properties, known as characterising sentences (Krifka et al.,1995). What follows elaborates on each type of genericity, the diagnostic tests to differentiate generic statements from particular statements and the main characteristics of generics.

Starting with kind generics, in this type of genericity the NP refers to a genus or kind in general rather than referring to a specific object as in (1, a and b). While the NP 'the cat' refers to a particular cat in the context in (1a) it refers to cats as a kind in (1b).

(1) a. The cat is sleeping on the mat. particular sentence

b. The cat was first domesticated in Cyprus. ✓ **kind generic**

Kind generics are not marked. According to Krifka et al. (1995), a given NP can serve as a denotation of simple object reference or as a kind, as in (1a) and (1b), respectively. Therefore, the literature proposes a diagnostic test for distinguishing kind-denoting NPs from object-referring NPs. Krifka et al. (1995, p. 10) suggested the kind predicates test, which refers to using the predicates "be extant, die out, be invented, or be exterminated". These predicates only allow kind-denoting NPs. A basic property of kind generics is that the genericity comes from the NP, as in example (2), adapted from Krifka et al., 1995, p. 2:

(2) a. The potato was first cultivated in South America. ✓ **kind generic**

b. Potatoes were introduced into Ireland by the end of the 17th century. ✓ **kind generic**

The second type of genericity is characterising generic sentences, which refers to statements that represent "propositions which do not express specific episodes or isolated facts, but instead, report a kind of general property, that is, report a regularity which summarises a group of particular episodes or facts" (Krifka et al., 1995, p. 2). In other words, the basic characteristic of a generic sentence that differentiates it from a particular sentence is that it refers to the general regularity of the predicate. To clarify, consider the following examples from Krifka, et al. (1995, p.3-9):

(3) a. John smokes a cigar after dinner. ✓**Characterising generic**

b. A lion stood in front of my tent. Particular statement

Example (3a) is a characterising generic sentence and can be interpreted to mean that smoking after dinner is a general regulation about John. In contrast, example (3b) is interpreted as if there was a particular episode in which a specific lion was standing in front of my tent at a specific time.

Krifka et al. (1995) discuss some diagnostic tests to distinguish characterising generics from particular statements. The first test includes using the adverb "usually/typically". They stated that the insertion of this adverb into a characterising generic sentence leads to a slight change in meaning while applying this adverb to a particular statement causes a change of meaning from a particular event to a general rule as in "a lion usually stood in front of my tent". Another test is related to the distinction between stative and non-stative sentences. Characterising sentences are stative while particular sentences are non-stative. Transferring the characterising sentence into a progressive one causes it to lose its generic reading. For instance, turning (3a) to the progressive tense leads to losing the generic reading as in (4), which refers to the existence of a particular episode.

(4)	John is smoking a cigar.	Particular statement

Having demonstrated how to diagnose and characterise generics, we now consider the main characteristics of generics as discussed in the literature. It is true that "no language has a unique, unambiguous marker of genericity equivalent to a quantifier or determiner" (Dahl, 1995, p.425). Moreover, within a single language, different types of NP can occur in a generic statement (Krifka et al., 1995). Hence, a basic characteristic of characterising generic sentences is that the generic force comes from the sentence, according to Krifka et al. (1995) who stated that characterising sentences can be associated with different types of NP either with a kind reference or an individual reference. The range of NP types with characterising generics is exemplified in (5), which shows that definite singular NPs (5a), indefinite singular NPs (5b), bare plural NPs (5c), and proper nouns (5d) are all possible in a preverbal position with characterising generics.

(5)	a.	The whale breathes underwater.	Definite singular NP
	b.	A whale breathes underwater.	Indefinite singular NP
	c.	Whales breathe underwater.	Bare plural NP
	d.	Jack smokes a cigar after dinner.	Proper noun

Lazaridou-Chatzigoga (2019) discusses the main characteristics of generics naming the following characteristics: a) temporal unboundedness; b)resistance to contextual restriction; and c) exception tolerance. What follows is a brief discussion of each.

Temporal unboundedness

In the literature, it has been stated that a distinctive property of characterising generic sentences is that they express an unbounded truth. This is illustrated in example (6a-c), from Lazaridou-Chatzigoga (2019, p. 160):

(6)	a.	Cats are meowing today.	[particular statement, can be bounded]
	b.	?Cats meow today.	[adding adverb of time causes odd reading]
	c.	Cats meow.	[characterising generic, timeless truth]
	d.	In 1989, Mary played tennis.	[characterising generic, bounded truth]

The examples in (6) show that the generic sentence in (6b) has an odd reading when bounded by the adverb of time when compared to (6c). This adverb can be added to a particular statement as in (6a). Krifka et al. (1995, p.36) questioned this characteristic stating that generic sentences may be temporally located by adverbs as in (6d). Mari et al. (2013, p. 50-52) explore temporal unboundedness as a property of characterising generics. They account for (6d) by making a distinction between ‘strong generics’ which are unbounded as in (6c) and ‘weaker’ generics as in (6d), concluding that unboundedness is a property of characterising generics.

Resistance to contextual restrictions

Another property of characterising generics is that they are contextually unrestricted. This characteristic differentiates characterising generics from quantified statements. According to Lazaridou-Chatzigoga (2019, p. 161), it has been argued that generics cannot be contextually restricted. By way of example, consider lions as in example (7), where in (a) every lion refers to the lions inside the cage, but in (b) lions can be interpreted as lions in general rather than the loin in the cage, demonstrating that the interpretation of generics at the sentence level is not constrained to a certain provided context.

(7) (Context: There are tigers and lions in this cage)

a. <u>Every lion</u> is dangerous. Particular reading

b. <u>Lions</u> are dangerous. **Generic reading**

Tolerance for exceptions

Tolerance for exceptions is a property of characterising generics (Krifka, et al.,1995). This is illustrated by (8a), which means that birds generally fly, even if some birds, such as penguins, cannot.

(8) a. Birds fly. [Characterising generic, true even when penguins cannot fly]

b. Every bird flies. [Quantified statement, false because penguins cannot fly]

Lazaridou-Chatzigoga (2019,p, 160) highlights that some scholars question the amount of exceptions a generic statement allows without losing its truth value. The answer to this question ranges from 0% to 99% depending on the type of property predicated of the kind as in (9). Some properties are essential as in (9a), others are characteristics as in (9b), and some are accidental as in (9c).

(9) a. Snakes are reptiles. [0% exception]

b. Telephone books are thick. [50% exception]

c. Mosquitoes carry the West Nile virus. [99% exception]

However, the type of properties is beyond the scope of this thesis and therefore is not discussed further. Having discussed the meaning of the basic types of genericity and how to differentiate generic readings from particular readings for each type, an account for genericity form-meaning mappings in English and MSA follows.

2.3 Genericity in English and MSA

As stated in Section 1.2.2, there are two types of genericity: characterising generics and kind referring NPs (Krifka et al., 1995). The discussion of how genericity is expressed in English and MSA in this section follows Krifka et al.'s (1995) analysis.

2.3.1 Generic form-meaning mappings in English.

When a sentence in English denotes a characterising generic reading as in (10), only indefinite singulars and bare plurals denote characterising generic readings.

(10) a. A green lamp is relaxing. **indefinite singular [+generic**]

b. Green lamps are relaxing. **bare plural [+generic**]

c. *Green lamp is relaxing. bare singular, ungrammatical

d. The green lamps are relaxing. definite plural [−generic]

e. The green lamp is relaxing. definite singular [−generic]

In (10), examples (a) and (b) describe a generalisation about the characterising property of green lamps. The generic force in these examples is independent of the subject NP, as in both (a) and (b) the example makes a characterising statement about green lamps in general. In contrast, definite plural and definite singular NPs in (d) and (e) cannot express such generalisation and are interpreted as being a particular statement about an existing definite green lamp (e), or a group of green lamps existing in the discourse (d). English maps the characterising generic meaning onto bare plurals and indefinite singulars.

However, definite singulars are more limited in their application to characterising generic sentences. To elucidate, look at example (11), adapted from Krifka et al. (1995, p. 11):

(11) a. The Coke bottle has a narrow neck. **Definite singular [+generic**]

(a well-defined kind generic NP in sentence-level genericity)

b. # The green bottle has a narrow neck. Definite singular [−generic]

c. A green bottle has a narrow neck. **Indefinite singular [+generic**]

d. Green bottles have a narrow neck. **Bare plural [+generic**]

Krifka et al. (1995) explain that using the definite singular in characterising generic sentences is only possible for what they call "well-defined" or 'canonical' kinds, such as "the coke bottle". A well-defined object such as "the Coke bottle", as in (11a), is a real-world object, but the well-defined kind of "green bottle" does not exist in the real world. In other words, green bottles are not identified as clear-cut, typical objects; hence, only indefinite singular NPs (11c) and bare plurals (11d) are allowed in characterising generics, whereas a definite singular NP is not (11b).

Turning to kind generics within the NP, the generic NP refers to a genus rather than to a particular object or individual, as shown in underlined words in example (12), where "the potato" in (12a) does not refer to a particular potato. Also, potatoes in (12b) do not refer to a group of potatoes but to the general kind of potato itself (Krifka et al., 1995, p. 2).

(12) a. <u>The potato</u> was first cultivated in South America. **Definite singulars [+kind**]

b. <u>Potatoes</u> were first cultivated in South America. **Bare plurals,** [+**kind]**

c. ?A potato was first cultivated in South America. Indefinite singulars

d. ?The potatoes were first cultivated in South America. Definite plurals

e. *Potato was first cultivated in South America. Bare singulars, ungrammatical

As shown in (12), reference to kind at the NP level is available to definite singular and bare plural NPs as in (12a) and (12b), but not to indefinite singulars as in (12c), or definite plurals as in (12d). The fact that the indefinite singular in (12c) cannot have the generic reading as the indefinite singular has in (11c) indicates that the generic reading for indefinite singulars can only come from the sentence level, not from the NP level.

In short, English characterising generics allow indefinite singulars and bare plurals to express the characterising generic meaning of the sentence. The definite singular is possible in characterising generics with the restriction of a well-defined kind. When the generic reading is within the NP (i.e., kind generics) English allows definite singulars and bare plurals. Definite plurals can express existential meaning, while a bare count singular in the subject position is ungrammatical in English.

2.3.2 Generic form–meaning mappings in MSA.

It is well known that English and MSA are both languages with an article system[4]. However, they differ in expressing genericity using NPs. In this section, I illustrate how MSA expresses different types of genericity. Starting with characterising generic sentences, only definite singulars and definite plurals allow the generic reading of the sentence as in (13):

(13) a. Aladwa? al-xadra? Murihat-un. **Definite plural [+generic**]

The -lamps-PL.DEF the green relaxing.

'The green lamps are relaxing.'

b. Addwa? al-?xdru Murih-un. **Definite singular [+generic]**

The -lamp-SG.DEF the green relaxing
'The green lamp is relaxing.'

---

[4] Discussing how articles are distributed with count plural and singular forms is essential to understand how the languages differ in the syntactic distribution of articles. The English language uses three types of articles (the, a(n), and Ø or no article option). The distribution of articles with count nouns is that the definite article 'the' comes with singular and plural count nouns, 'a' with singular count nouns, and Ø with plural count nouns. MSA is a language with a definite article, but without an indefinite article. It is worth mentioning that MSA's article system includes the definite-indefinite distinction (Fassi Fehri, 2004). While the morphological form of the definite article is the overt bound morpheme 'al' (the equivalent of the English the), the morphological expression of indefiniteness is debated. While some scholars follow the view of expressing indefiniteness using a *nunation* (-n suffix, pronounced but not written and called *tanwiin* in MSA grammar), others reject this suffix as a marker of indefiniteness (Fassi Fehri, 2004; Sarko, 2009). They justify this rejection by the occurrence of –n with definite nouns, such as proper nouns, as in (A).

(A) Ahmad –un fareh –un.
Ahmad –DEF.PN happy.
'Ahmad is happy.'

This study adopts this view that suggests there is no indefinite morpheme in MSA. Hence, bare nouns in MSA "exhibit a syntactic behaviour which makes them closer to Romance and English indefinites, rather than to true BNs, since they are not readily interpretable as Gen" (Fassi Fehri, 2004, p. 45), as I elaborate later in the discussion of genericity in MSA. As in English, the definite article in MSA can be attached to both singular and plural count nouns.

c. adwa?-un xadra?-un Murihat-un. Bare plural [−generic]

lamps-PL.INDF green relaxing.

'Green lamps are relaxing.'

d. dwa?-un ?xadra?-un Murih-un. Indefinite singular [−generic]

lamp-SG.INDF green relaxing.

'*Green lamp is relaxing.'

Unlike English, the characterising generic interpretation in MSA is expressed through definite NPs (plural and singular), as shown in examples (13a-b). Interestingly, definite singulars and plurals in (13a-b) can also express existential meaning in addition to the generic one. In contrast, bare singulars and plurals (13c-d) only allow existential reading, meaning that "there are some green lamps which are relaxing" and have only indefinite interpretation.

Turning to kind generic NPs, MSA is similar to English in allowing definite singular NPs to denote kind reading to the NP as in (14a). MSA differs from English in also allowing definite plurals to refer to the kind, as in (14b). Another difference is that bare plurals and bare singulars (which are always indefinite) refer to objects rather than kinds in MSA (14c-d).

(14) a. Albuma-at –u nader-aat-un fe Antarktika. **Definite plurals [+ kind]**

the owls–DEF.PL rare in Antarctica.

'?The owls are rare in Antarctica.'

b. Albuma-tu nader-atun fe Antarktika. **Definite singulars [+kind]**

the owl–DEF.SG rare in Antarctica.

'The owl is rare in Antarctica.'

c. *Buma-at –un nader-aat-un fe Antarktika. Bare plurals (ungrammatical)

owls–INDF.PL rare in Antarctica.

'owls are rare in Antarctica.'

d. *Buma-tun nader-atun fe Antarktika Bare singulars (ungrammatical)

owl–INDF.SG rare in Antarctica.

'owl is rare in Antarctica.'

To sum up, in MSA kind reference in the NP level is expressed with definite singulars and definite plurals. Indefinite plurals and singulars are not grammatical in preverbal position unless modified. In what follows, I discuss the available semantic frameworks to account for the cross-linguistic difference in NP interpretations.

2.4 Cross-linguistic semantic framework for NP interpretations in English and MSA

Several semantic frameworks have been proposed in the literature to explain how NPs gain their interpretations in characterising and kind generics. Finding a single framework that fits all languages is quite challenging. Hence, some semantic frameworks aim to incorporate cross-linguistic principles. In this section, I introduce and evaluate various semantic frameworks to explore how they handle the differences in how genericity is expressed between English and MSA, namely, the modal approach of Krifka et al. (1995), and Chriechia's (1998) and Dayal's (2004) frameworks.

2.4.1 Krifka et al. (1995): the modal approach

An influential framework is the modal approach suggested by Krifka et al. (1995), which has become a widely accepted view of generics. According to this approach, characterising generic meaning is driven by a hidden generic operator "GEN", which works as a quantifying adverb giving the characterising generic reading to the sentence by quantifying over individuals or situations. It binds variables in its scope, namely the restrictor (the condition) and matrix (main clause). This operator is not selective and can yield existential and generic readings, as exemplified in the tripartite form in (15), from Krifka et al. (1995, p. 26).

(15) Typhoons arise in this part of the Pacific.GEN [restrictor][matrix]

GEN[x;y](x are typhoons; y is this part of the Pacific & x arise in y) **generic**

GEN [x;y] (x is this part of the Pacific; y are typhoons & y arise in x)

= GEN [x;] (x is this part of the Pacific; ∃y [y are typhoons & y arise in x]) existential

In (15), the first structure yields the meaning that, in general, typhoons arise in this part of the ocean. The meaning of the second structure is existential meaning that 'there are typhoons that arise in this part of the Pacific'. In their view, bare plurals can be ambiguous between the existential and the generic readings, as shown in the example above.

The GEN operator can also explain the generic reading of indefinite singulars in characterising generics according to Krifka et al. (1995). They argue that indefinite singulars can be interpreted as generic in characterising sentences due to the presence of this generic operator. This operator binds the variable introduced by the indefinite noun phrase, enabling it to represent a generalisation rather than a specific example in the discourse. The generic operator functions similarly to an adverb of quantification and binds the variable introduced by the indefinite noun phrase. Consider (16):

(16) A lion roars.

The indefinite NP 'a lion' introduces a variable that can be bound by the generic operator. This binding process allows the indefinite to be interpreted as referring to any member of the category it denotes, rather than a specific individual. However, this approach faces empirical challenges in accounting for exception tolerance and various interpretations of generics across languages (Mari, 2015).

2.4.2 Chierchia (1998) and Dayal (2004)'s semantic frameworks

Chierchia (1998) uses certain type shift operations to explain the different forms of NPs used for generic and kind reference in English and various other languages. Two operators in this type-shifting operation are relevant to this discussion: the down operator ∩ and the iota operator ι. The down operator denotes the meaning of kind while the iota operator denotes the meaning of a nongeneric definite (maximality). In this account, Chierchia (1998) proposes the Nominal Mapping Parameter (NMP) which offers a structured explanation for the three types of languages: those with determiners, those without, and those that mix both. According to this theoretical framework, NPs can be either argument ([+arg]) or predicative ([+pred]). Predicative ([+pred]) NPs require an article or other determiner to function as arguments in the subject position.

Chierchia (1998) classifies languages based on this parameter as [+arg, +pred] languages , [-arg, +pred] languages, [+ arg, -pred] languages, and [-arg, -pred] languages. Starting with [+arg, +pred], Germanic languages including English fall in this category. In these languages, bare noun NPs can serve as arguments only if they denote kind. To give the episodic specific meaning, the NP must be transformed into a determiner phrase, either by adding an overt determiner (as in English) or through a type-shifting operation if no determiner is available (as in Russian). In contrast, [–argument, +predicate] languages such as Romance languages, bare NPs cannot work as arguments, NPs need to be attached to a determiner and form determiner phrases (DPs) in order to project to subject position. In these languages, bare NPs are required to be licenced by articles. The third type [+ arg, -pred] refers to languages like Chinese and Japanese that rely on a classifier system in which bare arguments are generalised, all nouns are mass, there is no plural, and bare nouns directly express kinds. The fourth type [-arg, -pred] languages do not exist.

Central to this discussion are the [+arg, +pred] languages and the [-arg, +pred] languages. English as discussed above fits into the first kind and allows both arguments and predicates in a preverbal position. In contrast, MSA fits in the second category and behaves like Romance languages, allowing only predicates. According to NMP theory, English is marked as [+pred, +arg], which indicates that the down operator ∩ can be used in bare plural arguments as there is nothing to prevent this according to the avoid structure principle, which states that shift operation applies at the earliest level, as the bare plural NP comes earlier in the structure. Because of the avoid structure principle and the blocking principle—which states that if a language has an overt determiner that lexicalises a particular operation, this blocks the covert application of the same operation—definite plurals in English lexicalise definiteness, which blocks mapping the generic reading onto definite plurals in English. In other words, only an episodic interpretation is possible for definite plurals in English which express maximality as given by the iota operator. According to Chierchia (1998), bare plurals prevent definite plurals from being generics. Definite plurals are determiner phrases, while bare plurals are arguments. Therefore, the avoid structure principle prevents the DP from shifting to kind interpretation, allowing the NP to shift there as the shift should be applied at the earliest level, the argument.

Applying the NMP to MSA can explain why bare plurals in MSA denote only an indefinite existential reading and why the generic reading is assigned to definite NPs. According to Fassi Fehri (2012:179), "Arabic NP/DP interpretations as generic (Gen) or (only) existential (Ex) depend on whether they express overtly the definite determiner or article D, or whether they lack such an overt expression of D." According to this, bare plurals in MSA are determiner phrases with null determiners that need to be licensed. Licensing it leads to a complex structure and, therefore, its distribution is restricted in MSA (for more discussion of the syntactic distribution of bare plurals in MSA see Fassi Fehri [2012]). Applying NMP theory, MSA is a [-arg,+pre] language and therefore bare plurals lack both kind and definite readings, behaving like bare plurals in Romance languages. In other words, the MSA definite article lexicalises both maximality (iota operator) and kind-reference (down operator), and, therefore, it allows definite plurals to have both maximal and generic readings and prevents the bare plural from having either reading in light of the blocking principle; the overt definite article blocks bare plurals form having a generic reading.

However, Chriechia's account faces a problem in accounting for Italian modified bare plurals with kind reading, as in (17):

(17) Insegnanti davvero dediti nella scuola di oggi sono quasi estinti.

"Really devoted teachers in today's school are nearly extinct."

Chierchia (1998, p.392) stated that "...the nominalising operator ∩ is an intensionalised version of ι. It follows that in appropriate (intensional) contexts we will be able to obtain with ι (i.e. the definite article) what we can get in English with ∩ (i.e. bare nominals). In other words, Chierchia (1998) argued that the generic and kind reading of definite plurals are derived by the intensionalised iota. The modified bare plurals are predicates with a null determiner. Therefore, the down operator can apply as the D is empty; hence, the absence of an overt article prevents blocking the down operator from assigning the generic reading for the Italian modified bare plural. Ionin et al. (2011a) argue that the solution is not ideal as bare plurals and definite plurals are given the same meaning; this meaning is derived from different operators. If Chierchia's account for modified Italian bare plurals is correct, it would be predicted that MSA bare plurals can have a kind reading when modified. Fassi Fehri (2012) accounts for the bare plural in MSA and provides evidence that they cannot express generic reading on either the NP level or the sentence level even when modified, unlike Italian bare plurals. The first evidence was that MSA bare plurals give an existential reading even with kind predicates, as in (18):

(18) a. fiyalat-u-n bayḍaaʔ-u nqaraḍa-t.

elephants-nom-n white-nom became.extinct-f.

'(Some) white elephants became extinct.'

b. fiyalat-u-n bayḍaaʔ-u t-uṯiir-u ʔiʕjaab-a n-naas-i

elephants-nom-n white-nom f-attract admiration-acc the-people.

'White elephants attract the admiration of people. '(Fassi Fehri, 2012, p.183)

The second piece of evidence is that using bare plurals in MSA with individual level predicates turns the sentence ungrammatical, unlike in Italian, as in (19) adopted from Fassi Fehri (2012, p. 184):

(19) ?? kilaab-u ḥiraasat-in daatu ḥajm-in kabiir-in ʔaktar-u faaʕiliyyat-in

dogs-nom watching-gen of size-gen big-gen more efficiency-gen
'Watch dogs of big size are more efficient.'

Therefore, unlike in Italian, MSA bare plurals are ungrammatical with an individual-level predicate, hence they are unable to express genericity. Fassi Fehri (2012) states that bare plurals in MSA are quantificational, express indefiniteness, and always convey existential meaning[5].

Moreover, Chriechia's avoid structure principle faces challenges in accounting for languages in which both bare and definite plurals have kind interpretations such as German and Brazilian Portuguese (Dayal, 2004; Ionin et al, 2011a). Although these languages are not related to the languages investigated in this study, Ionin et al. (2011a) argue that Dayal's proposal was empirically supported to be a better account. Therefore, the following briefly discusses this proposal and explains how it accounts for MSA and English plural generic form-meaning mappings.

Dayal (2004) adopted Chierchia's semantic framework but not the avoid structure principle and derived testable predictions for the distribution of plural and singular kind NPs. According to this proposal, the first prediction is that every language has a definite article, lexicalises the iota operator and is predicted to have definite singular generics (Dayal, 2004). According to this prediction, both English and MSA have a definite article and both languages were similar in mapping the definite singular to generic reading, as shown in Section 2.3. This prediction holds for both English and MSA.

The second prediction is concerned with plural kind forms. According to Dayal (2004), plural kinds are derived by the down operator—not the intensionalised iota operator as claimed by Chierchia— and the blocking principle does not have to apply to the down operator. Languages are on three options: first, the iota is lexicalised by the definite article and the down operator is not. The definite article blocks definite plurals from having kind reading by the down

[5] According to Fassi Fehri (2004, 2012), bare plurals in Arabic differ significantly from English bare plurals in both distribution and interpretation. In Arabic, bare plurals are subject to stricter syntactic constraints compared to their English counterparts. While English bare plurals can appear freely in various syntactic positions—such as subjects, objects, or predicates—Arabic bare plurals are more restricted in distribution. They most naturally occur in predicate or existential contexts, as in *fī al-bayti qiṭaṭ-un* ("there are cats in the house"), but are marked or even awkward when used in subject positions with a generic interpretation, as seen in *qiṭaṭ-un yaʿīshna huna* ("cats live here"), which is only felicitous under an existential reading. Furthermore, the quantificational force of bare plurals in Arabic differs significantly from English. English bare plurals often imply universal quantification in generic contexts (e.g., "Cats hate water"), whereas Arabic bare plurals generally convey existential readings unless definiteness or other syntactic devices are used to force a generic interpretation. For instance, *qiṭaṭ-un takrah al-māʾ* is best read as "some cats hate water," whereas *al-qiṭaṭu takrah al-māʾ* with the definite article allows a generic reading. This reflects a deeper semantic distinction discussed by Fassi Fehri, who argues that English bare plurals can denote kinds directly, while Arabic requires definite plurals for kind reference. Thus, in Arabic, the semantic type associated with kind-denoting nominals is closely tied to definiteness, highlighting a fundamental difference in how each language encodes generality and kind-reference through plural forms.

operator which gives kind reading to bare plurals, as no blocking occurs, such as in English. Second, languages may lexicalise both iota and down operators. The blocking principle applies to both, allowing definite plurals to have definite and kind readings and blocking both readings from being assigned to bare plurals. This option applies to the Romance languages and MSA. Third, languages may lexicalise both iota and down operators, and the blocking principle is applied only to the iota operator. In this language option, definite plural and bare plurals can express kind readings as in German. Having discussed the interpretation of bare and definite plural NPs in both languages, the following deals with singular kind NPs.

For singular kind NPs, it has been claimed that the down operator cannot apply to singular count nouns and that the mechanism that derived the king reading from singular NPs is different from that of plural NPs (Chierchia, 1995; Dayal, 2004). However, various semantic frameworks have different explanations for how definite singulars have a kind interpretation. In Chierchia's (1998) framework, the definite singular has a kind reading as a result of a massifying operation in which the definite singular is turned into a mass noun before being used to express kind reading. However, this explanation is applied to English only and cannot account for other languages (Ionin et al., 2011a). In contrast, Dayal (2004) argues that kind definite singulars have the same semantic features as the non-generic definite singular. The kind reading is derived by the application of the iota operator to a taxonomic noun. in other words, the generic definite singular NP has the feature [+taxonomic], which ensures that the iota operator is combined with a taxonomic noun to establish a kind reference. Moreover, Dayal (2004) noted that the well-defined kind constraints on the definite singular explained earlier is a pragmatic consequence of the taxonomic noun and is universal. This claim can be illustrated by considering the following example:

(20 a. The Bengali tiger is dangerous.

b. # The wounded tiger is dangerous. (Ionin et al., 2011b, p. 249)

"The Bengali tiger" is taxonomic in certain pragmatic contexts, while "the wounded tiger" can refer to a particular tiger in the discourse. In this sense, definite singulars with a kind reading differ semantically from definite plurals with a kind reading; the iota operator applies to the former and the down operator applies to the latter in the derivation of the kind reading.

Considering the use of indefinite singulars, it has been stated that an indefinite singular can give a characterising generic reading when it is bound by the GEN operator to express habitual or general truth in English across different semantic frameworks. The semantic frameworks differ in emphasising different aspects; for instance, Chierchia (1998) stated that the binding of indefinite singulars by the GEN operator can give the indefinite singular a lawlike sense. Dayal (2004) highlights that the availability of definite singular and bare singular NPs in a language does not falsify the blocking principle as the definite article blocks the definite (familiar/anaphoric reading) for bare singulars. Moreover, if bare singular exists in a language that expresses singular kind with the definite article, the kind reading is ruled out for the bare singular, which is the case for MSA, or restricted if available. Dayal's (2004) predictions have been tested by Ionin et al. (2011a) by testing native speakers of English, Spanish and Brazilian Portuguese. The results support the suitability of Dayal's framework to account for English. However, no empirical investigation exists in the literature on how MSA expresses genericity. It is worth mentioning that MSA is like Romance in plural generics but differs from Romance in the absence of indefinite singulars. Fassi Fehri (2012, p. 186) stated "Arabic bare plurals behave ... like Arabic bare singulars." This means the indefinite singular lacks the definite and the generic readings in the subject position. While Romance indefinite singulars can express characterising generic reading, MSA indefinite singulars cannot (Fassi Fehri, 2012, p. 192)[6].

---

[6] Fassi Fehri's analysis of Arabic nominal syntax shows that the referential feature of the Determiner position in Arabic is fundamentally different from that in English. In Arabic, referentiality is not always overtly encoded through a determiner; it can be encoded via an empty D or noun movement to D (as in proper names). Moreover, Arabic lacks kind-referring bare nouns and does not permit determiners with proper names in referential contexts (e.g., *"this Bagdad"* is ungrammatical), indicating that Arabic treats referential DPs differently both syntactically and semantically compared to English.
Arabic uses bare nouns to express indefiniteness:

- Arabic:
  رأيتُ ولدًا
  raʔaytu waladan
  "I saw a boy"
  → *waladan* is bare singular, but it's interpreted as indefinite due to case/morphology, not a dedicated article.
- English:
  "I saw a boy"
  → Requires overt indefinite article "a" to signal the same interpretation.

So, in Arabic, indefiniteness is encoded via morphology/case or syntactic position, not with a dedicated D head like "a".

Another evidence for the fundamental difference comes from the syntax of proper names. In Arabic proper names can occur without a determiner and still be referential.

- Arabic:
  زارَ بغدادَ
  zaara Baghdād-a
  "He visited Baghdad"
  → No article needed; referential interpretation is retained.
- English:
  "He visited Baghdad"
  → Also no article, but note: English can also say "this Baghdad", while Arabic cannot:

In sum, this discussion showed that several attempts have been made to discuss how and why languages differ in expressing generic meanings (Chierchia, 1998; Dayal, 2004; Ionin et al., 2011). All accounts agree on the two types of genericity: characterising and kind generics. However, they differ in the mechanism of assigning kind reading to definite plurals. While Chierchia's account stated that the definite kind reading is assigned through the intentionalised iota, Dayal uses the down operator for definite and bare plurals.

2.5 Summary

This chapter provides an account of the linguistic property under investigation, genericity. It discusses different types of genericity and the diagnostic tests to differentiate between generic and non-generic interpretations. Then, this chapter discusses the form-meaning mappings in kind and characterising generics in English and Modern Standard Arabic. It discusses different accounts in the semantic literature that explain the NP generic and non-generic interpretations. The cross-linguistic comparison of NP in interpretations in English and MSA is summarised in Table 2.1.

Table 2-1 Cross-linguistic comparison of NP interpretations in English and MSA.

NP type	**English**	**MSA**
Definite singulars	Characterising generics [+generic, +definite, −plural] or kind generics (WDK) [+kind, +definite, −plural] or canonical reading [+definite, −plural]	Characterising generics [+ generic, +definite, −plural] or kind generics [+kind, +definite, −plural] or canonical reading [+definite, −plural]
Indefinite singulars	Characterising generics [+generic, −definite, −plural] or existential reading [−definite, −plural]	It is not available in generic reading. Existential reading [−definite, −plural]

- Ungrammatical in Arabic:
  hādhihi Baghdād → [ungrammatical]
  "this Baghdad"

This tells us that D in Arabic proper names does not tolerate a deictic specifier like "this," unlike in English—showing a deeper difference in the structure of the DP.

NP type	**English**	**MSA**
Bare plurals	Characterising generic [+generic, −definite, +plural] or kind generic [+kind ,−definite, +plural] or existential readings [−definite, +plural]	No generic reading [−definite, +plural]
Bare singulars	ungrammatical	No generic reading [−definite, −plural]
Definite plurals	No generic reading [+definite, +plural]	Characterising generics [+generic, +definite, −plural] or kind generics (WDK) [+kind, +definite, −plural] or maximal reading [+definite, −plural]

Chapter 3 Genericity in L2 Acquisition

3.1 Introduction

The aims of this chapter are: i) to set the theoretical framework of this study by elaborating on key concepts in L2 acquisition related to this study, ii) to review and evaluate minimalist theories of the acquisition of L2 form–meaning mapping and its difficulty and iii) to review a selection of empirical literature on the acquisition of genericity form-meaning mappings in English as an L1 and L2. The chapter starts by giving a brief overview of different perspectives on how the L2 learner's mental representations of the L2 are constructed. Section 3.2.1 provides a historical discussion of theoretical proposals that explain language and language acquisition, focusing on the Generative Approach to Second Language Acquisition (GenSLA). Section 3.2.2 elaborates on Lardiere's (2009) Feature Reassembly Hypothesis and Slabakova's (2014) Bottleneck Hypothesis. In section 3.3, this chapter presents and evaluates relevant studies on the acquisition of genericity, including studies on L1 and L2 acquisition with a focus on studies that consider Arabic-speaking learners of English as L2 in section 3.3.3. Section 3.4 provides an interim chapter summary with a commentary on the main findings and how they inform this study.

3.2 The Generative Approach to SLA: Feature Reassembly and the Bottleneck Hypothesis.

This section discusses the key elements of GenSLA related to this study, namely, a feature-based view of the acquisition of form-meaning mappings and a GenSLA-based account for the difficulty in the acquisition of these mappings. Before doing so, it provides a brief historical account of generative linguistic theory more broadly to set the stage for discussing the acquisition of form–meaning mappings from a GenSLA minimalist perspective.

3.2.1 Historical overview of generative linguistic theory

Learning an L2 entails building a system of mental grammar that enables the learner to communicate and comprehend the learnt L2 successfully. Constructing a mental grammar involves combining information from different sources, including the learner's native language, Universal Grammar and L2 input. The generative approach to Second Language Acquisition (GenSLA) research aims to illuminate the interaction between UG, L1 knowledge and L2 input in acquiring L2s (Rothman & Slabakova, 2018). Over time, various viewpoints have emerged to explain how the mental representations of L2s are formed, whether they differ from those related to L1, and why some aspects of language are more accessible or more challenging to learn than others. These different perspectives offer various accounts of language and language learning. The following is a brief note on these perspectives.

Chomsky's (1965) view of UG as an innate ability that allows language acquirers to create new sentences has guided research in SLA for many years. The Principles and Parameters model (P&P) and the Minimalist Programme (MP) are major perspectives in this area. For this study, I adopt a minimalist view of UG. Therefore, I will briefly discuss the theory of language and language acquisition from both perspectives on UG to lay the ground for more fully presenting feature-based proposals, namely, FRH and BH.

It is crucial to clarify that UG refers to a genetic and linguistic capacity for language and a theory of individual language grammar (Roberts, 2016). Chomsky (1981) claimed that each language comprises some universal principles common to all languages that everyone can possess and some language-specific properties that distinguish languages from one another (parameters). Structural dependency is a general assumption in this theory. According to this assumption, languages are structured in phrases grouped into larger structures to form sentences. The phrase structure rule is a principle shared by all languages, and the cross-linguistic differences are mainly parametric and refer to specific options that can vary between languages.

Considering UG, Chomsky’s (1981) development of P&Ps describes acquiring a language in terms of accessing massive information in the UG and selecting the appropriate parameter value considering the received input. In other words, children learn their first language by relying on access to innate UG principles with the help of L1 input, which guides the setting of their language-specific parameters. As for acquiring an L2, many hypotheses have been proposed considering UG access with a focus on the formulation of the mental realisation of the target-language grammar and whether it differs from that of the L1. Some research supports the claim that access to the UG is not possible for adult L2 learners, while others support the position that adult L2 learners can have either partial or full access to the UG (White, 2003). Schwartz and Sprouse's Full Transfer Full Access Hypothesis (FT/FA) (1994, 1996) posits that the learner’s L1 strongly influences L2 learning through transfer. It is argued that L2 learners transfer all the L1 parameters and then reset them for L2.

In alignment with the Full Access view, Ionin et al. (2004) examined parametric variation in the acquisition of articles in a seminal study. They proposed the Fluctuation Hypothesis building on the Article Choice Parameter (ACP) for two-article languages as either the definiteness or specificity setting. It claimed that having access to UG, L2 learners fluctuate between the two parameters of the ACP until L2 input enables them to make the correct choice. This example highlights that a crucial assumption of the P&P framework is the binarity of the parameters – that is, each parameter functions either one way or the other– which is claimed to reduce the adequacy of this framework (Chomsky, 2000; Roberts, 2016). Evidence of this stems from the fact that some languages have parameter values not allied with such a binary view, which a discussion of the head-direction parameter can illustrate. The head-direction parameter states that every structural phrase has a head that verifies the type of phrase (e.g., a noun is the head of a noun phrase). Languages can be head-final or head-initial, based on whether the head precedes or follows its complement. English is an example of a head-initial language, whilst Japanese is a head-final language. However, Chinese is an example of a language that does not have a binary value for this parameter. While it is head-final in the noun phrase, Chinese could be head-initial in the verb phrase and must be head-initial in the preposition phrase. This highlights that the descriptive and explanative adequacy of the P&P framework is limited[7].

3.2.2 Minimalist Theory of Language and Language Learning

The Minimalist Program (MP) (Chomsky, 1995, 2000, 2007) offers a more exploratory perspective that goes beyond the limitations of the P&P framework in explaining language by focusing on exploring “how well [the faculty of language] is designed” (Chomsky, 2000, p. 92). Within the MP, Chomsky (2005) describes language design by stating that language emerges from the interaction of three factors: the genetic endowment (UG), which is as large as needed; the experience or Primary Linguistic Data (PLD) for language acquisition; and non-language specific cognitive factors (in other words, general learning strategies). Therefore, the MP enables the move from total reliance on macroparameters (big trends in parametric linguistic variations) to assessing the input and role of non-linguistic factors in explaining acquisition and variation (Domínguez, 2014).

---

[7] As Roberts (2019) argues, the traditional notion of the head-parameter—as a single, binary setting determining whether heads precede or follow their complements—is too coarse to account for the attested variation across languages. Instead, more recent work proposes a distinction between macro-parameters and micro-parameters. Macro-parameters govern broad, cross-linguistically stable properties (e.g., head-initial vs. head-final tendencies), while micro-parameters capture more fine-grained, language-specific variation. This shift reflects the observation that languages tend not to be disharmonic in completely random ways; rather, variation often follows systematic patterns linked to particular syntactic domains or features. Under this view, apparent exceptions to head-directionality harmony can be understood as results of micro-parametric variation interacting with a more general macro-parametric setting. This framework allows for a more nuanced understanding of typological patterns and the limits of syntactic diversity.

Understanding basic assumptions of the MP that refine GenSLA views of language acquisition is crucial for understanding the language learner's mental grammar and linguistic variation. A basic assumption of the MP is that the language faculty can be reduced to a universal computational system, the lexicon, and two interface components: phonetic form and logical form. This system is responsible for generating linguistic expressions by using the basic derivational operations in this system (Merge, Move, and Agree) to map features to lexical items (Chomsky, 1995, 2000). Features are central to the derivation process; they are initially associated with functional heads and are later externalized through morphemes and words. According to the examples provided by Slabakova (2016, p.41), these features can be *phonological features* such as [± voice], *semantic features* such as [± definite] and *formal morphosyntactic features* such as [Number: {Sg, Pl}]. It is assumed that some features are semantically interpretable and enter the derivation with an inherently specific value (such as the Tense feature of T) and others are semantically uninterpretable, unvalued (in cases like Case features of NPs), and receive their value during the derivation before being deleted (Ionin et al., 2024). While some uninterpretable features, such as phi-features, may be valued, the focus here is on features that are both uninterpretable and unvalued at the relevant stage of derivation.

Returning to the derivational operations in the generation of linguistic expressions, the first operation, Merge, forms syntactic structures by combining two smaller linguistic elements. The second operation, Move, involves moving elements within a sentence to form different syntactic structures. The third operation, Agree, establishes syntactic dependencies by assigning values to uninterpretable features. This phase in feature mapping works as a feature-checking mechanism (Chomsky, 2001). Slabakova (2016) asserts that lexical expressions differ from features and are added to the message after feature-checking in the Agree operation. The output of these operations is a phrase structure (syntactic tree).

Another basic assumption of the MP is that parametric variation is situated in the lexicon. Chomsky (1991) adopted Borer's (1984) view of restricting interlanguage variation to the properties of lexical items, a position now known as the Borer–Chomsky conjecture. Baker (2008, p. 353) defined the Borer–Chomsky conjecture by stating that "all parameters of variation are attributable to differences in the features of particular lexical items (e.g., the functional heads) in the lexicon". According to this view of linguistic variation, linguistic differences lie in the mapping of semantic features and morphosyntactic features (formal features) in the functional category head and its associated morpho-lexical item. Baker (2008) questioned the sufficiency of this conjecture in explaining the complexities inherent in linguistic variation and argued against it.

It has been proposed that the linguistic variation is captured at the interfaces: phonetic form (PF) and logical form (LF). While variation may involve these interfaces, it is not necessarily limited to them. The phonetic form interface refers to the level of linguistic representation where abstract syntactic structures are converted into their phonetic realisations. Slabakova (2016) considers movement as an illustration of linguistic variation at the PF interface. When movement occurs in the syntactic structure, two copies of the moved item are in the phrase structure tree. One way of capturing linguistic variation is through PF features; languages may differ in choosing which copy to pronounce.

Others have proposed that linguistic variation is captured at the logical form interface. According to this view, since semantic features and functional heads are considered universal, the differences must be in how or when these functional heads are assigned value. This can happen either during syntax through functional morphology or after syntax at the logical form (LF) stage. In this stage, their meanings are determined by context, such as the semantic meaning of definiteness, which can be supplied by morphemes in some languages (e.g., articles in English) or by context in other languages (Slabakova, 2016). Understanding possible explanations for cross-linguistic differences is essential for this study, which considers the role of cross-linguistic variation in acquiring genericity by assuming an MP view.

Moving to the discussion of acquisition under Minimalism, Chomsky (2000) stated that acquiring an L1 involves selecting the L1 features provided by UG and assembling them into L1 lexical items under exposure to linguistic input. Different models of L1 acquisition have been developed, yet they are outside the scope of this study, L2 acquisition. Turning to the L2 acquisition of functional morphology, various theories have been proposed in the context of the MP to account for the acquisition of functional morphology based on access to the UG and the L2 initial state. Advocates of the view that adult L2 learners do not have access to UG during the post-critical learning period have proposed the Representational Deficit Hypothesis (RDH), arguing that some features of L1 become fixed after a critical period and that adult L2 learners are unable to adjust these features in the acquisition of L2 and cannot acquire features of L2 that are not available in their L1 (Hawkins & Chan, 1997).

In contrast, some scholars, assuming a Full-Access/Full-Transfer view, discussed above in section 3.2.1, reject the deficit view. Lardiere (2009) proposed the Feature Reassembly Hypothesis, which assumes access to the universal set of features through the UG and claims that "any feature contrast that is detectable is, in principle, acquirable" (2009, p.214). From this view, the L2 learner's task is to remap the L2 features – which may or may not exist in the learners' L1 or be bundled differently with other features – onto the L2 lexical representation. In line with the FRH, Slabakova (2009, 2019) argues that L2 learners can acquire the syntax and semantics of the L2, and the acquisition of the functional morphology causes difficulty. Hence, it is a bottleneck of L2 acquisition. Lardiere's FRH and Slabakova's Bottleneck Hypothesis propose explanations on how L2 is acquired and enable the building of testable predictions for the sources of variability in L2 acquisition, assuming the UG Full Access and Full Transfer. Therefore, this study adopts the FRH and the BH to establish predictions about the learnability of generic interpretations in English as an L2 by Arabic-speaking learners. What follows elaborates on these two hypotheses.

3.2.2.1 Feature Reassembly Hypothesis.

In consensus with the MP view of features mapping the core of cross-linguistic differences and L1 successful acquisition involving acquiring the language's formal features, Lardiere (2008, 2009) proposed the FRH. Lardiere (2009, p. 180) highlighted the importance of features and stated that "the basic unit of currency for describing differences between languages has been effectively 'exploded': i.e., reduced to the level of individual abstract features, to gain better empirical coverage of the data". In other words, formal features (phonological, syntactic, or semantic) are bundled together on the lexical items of a language and these features describe the cross-linguistic differences. Lardiere (2007) argued that acquiring an L2 grammar goes beyond the availability of features for selection from UG. It includes how these features are bundled into lexical items or functional categories and language-specific conditional environments in which they appear.

Lardiere (2009) claims that L2 learning involves acquiring a second set of features which may differ from the L2 learners' L1. She suggested that the L2 learner brings a complete set of L1-assembled features on the lexical items and functional categories to the acquisition task. Hence, the difficulty with L2 learning is located in the process of mapping the already assembled L1 syntactic knowledge onto the morphological or phonological modules of the L2. According to Lardiere (2007), acquiring an L2 requires the appropriate reconfiguration of the formal and semantic feature bundles and the appropriate determination of the conditions of the morphophonological expression of these features. Consequently, the acquisition task for L2 learners includes finding the similarities between the functional meanings in the L1 and L2 and mapping these meanings from the L1 onto L2 lexical items. Learners then need to check the features and assemble them based on the conditions of L2 evidence in the input, acknowledging that the acquisition challenge differs in different contexts. This means that if a feature of the L2 is available in the L1, but its context is not related to its L2 context, acquiring it is possible but more complex.

Thus, investigating the reassembly of features into L2 lexical items inclusive of all the L2-specific (re)assembly conditions plays a key role in understanding the challenges raised by cross-linguistic differences. According to Lardiere (2007, 2009), such conditions include whether the feature is morphologically realised as inflectional or free, covert, or overt, and interpretable or uninterpretable. The importance of understanding assembly conditions is revealed by discussing how the feature [+past] is assembled in English, Irish and Somali. Lardiere (2007) argued that although these three languages select the feature [+past], the way it is realized morphologically varies across languages, affecting its semantic and grammatical functions. In English, the morpheme -ed, which marks [+past], can also encode perfective aspect and contribute to the expression of the unreal hypothetical mood, in addition to its primary role in marking past tense, as seen in (1a and b).

(1) a. The cow jumped over the moon.

b. If I only had a brain... (Lardiere, 2008, p. 110-111)

Irish [+ past] is marked on complementisers in the phrase and agrees with the past tense of the verb in the embedded clause shown in example (2):

(2) Deir sé gurL thuig sé an scéal

Says he that understood he the story.

'He says that he understood the story.'

It occurs differently in Somali, as the feature [+past] appears on the determiner phrase and it implies different conditions in addition to the past tense. Example (3) shows a temporal habitualness expression in (3 a), whether the noun referent is visible or invisible to the speaker in (3b) and possession in predicative genitives in (3c).

(3) a. (Weligay) dúhur–kii baan wax cunaa

(always) noon–detM.past F.1S thing eat.pres

"I (always) eat at noon."

b. Inán–tii hálkée bay joogta?

girl–detF.past place–detM.Q F.3S stay.F.pres

'Where is the girl?'

c. Búug–gani waa búug–gíi Maryan

book–detM.dem Foc book–detM.past Maryan

'This book is Maryan's book.'

Moreover, Cho and Slabakova (2014) depicted the frequency and contradiction of the formal or semantic feature evidence in the input as affecting the speed of the acquisition and stated that "feature reassembly may be slow to occur or may not occur at all if the relevant evidence for the formal or semantic feature is rare or contradictory in the linguistic input" (p. 160). A reasonable approach to tackling acquisition success could be to provide L2 learners with more evidence of the features under investigation through instructional input. The FRH has been shown to be useful in L2 investigations because "it compels the researcher to build a more complex picture of the acquisition process and to pay attention to meanings, syntactic forms, conditioning environment of the functional morphology…" (Slabakova, 2016, p.201). Therefore, this investigation uses FRH to provide a feature-based linguistic contrast between the learners L1 and L2 and to define the L2 learning task and possible difficulty with the support of Slabakova's cline of difficulty, which is presented in the following part.

3.2.2.2 The Bottleneck Hypothesis.

Consistent with Lardiere's feature reassembly view, Slabakova (2009, 2013, 2019) contributed to the debate on what is easy and challenging to acquire in SLA by proposing the Bottleneck Hypothesis. Slabakova (2009) highlighted the importance of investigating what constraints the FRH faces to increase the predictive power of this proposal. To accomplish this, Slabakova (2009) considered the proposal of Ramchand and Svenonius (2008), which suggested that a language's semantic/pragmatic component is universally available in all languages and offers different meanings.[8] They added that linguistic parametric differences lie in expressing these universal meanings. Slabakova (2009, p. 316) provided clarification of the different ways, stating that "some languages have overt morphemes; other languages allow the context to fix the values of the specific features; still others leave functional morphology under-informative or vague, with the extra information filled in by context only when needed". In other words, some

[8] Slabakova (2009) mentions that the semantic/pragmatic component is called the conceptual–intentional system (C–I) by Chomsky (2004) and the conceptual structure by Jackendoff (2002).

languages employ explicit morphemes to convey grammatical information. For instance, in English, the suffix "-ed" denotes the past tense. In contrast, other languages rely on contextual cues to determine grammatical features. For example, in Chinese, verb forms do not change to indicate tense; instead, the context or additional words provide temporal information. Additionally, some languages possess minimal or ambiguous morphological markers, necessitating that listeners or readers infer grammatical details from the surrounding context (Slabakova, 2009). A primary implication of Ramchand and Svenonius' (2008) proposal is the continuous availability of the syntactic structure even in the absence of the morpheme realisation of this grammatical meaning.

Slabakova (2009) asked an important question about the different mapping types between morphemes and their semantic meanings. Using articles as morphemes and the different ways of expressing the meanings related to these morphemes, such as (in)definiteness and specificity in different languages, Slabakova (2009) supported the argument of the universality of these meanings, as they are available in all discussed languages seen in Table 3.1. Parametric variation permits these languages to code these meanings differently. The table shows that some languages use an overt morpheme, whilst others allow the context to do so.

Table 3-1 Cross-linguistic options for mapping meanings encoded by definite articles onto morphemes or discourse-pragmatics, adopted from Slabakova (2009, p. 319)

Meanings	**Norwegian**	**English**	**Lillooet Salish**	**Russian**
Argument hood	morpheme	morpheme	morpheme	null morpheme
Familiarity	morpheme	morpheme	discourse	discourse
Specificity	morpheme	discourse	Morpheme	discourse
Referent tracking	discourse	discourse	discourse	discourse

Slabakova (2013) argued that syntactic and semantic cross-linguistic differences between languages are reflected in functional morphology. Arguing for the need to understand the language faculty components and interactions in exploring how language is acquired, Slabakova (2013) built on Reinhart's (2006) model and assumed that functional categories occur in the computational phase (see Figure 3.2) in which the syntactic operations of constructing phrases by using the lexicon occur and continue until the process of lexicon exhausting and feature checking are completed.

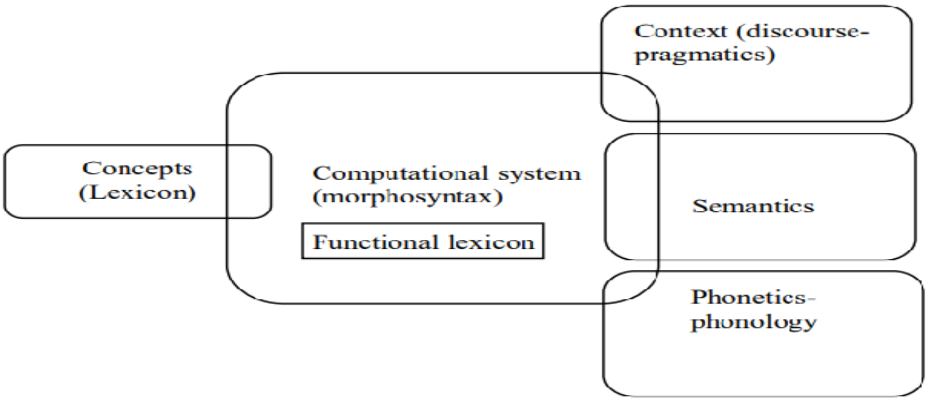


Figure 3.1 Modular design of the language faculty (Slabakova, 2013, p. 7)

Moreover, the essence of language acquisition lies in acquiring form–meaning mapping in functional morphology. Learners need to go through functional morphology to acquire syntax and semantics in an L2. As a result, Slabakova (2013) argued that functional morphology is the bottleneck of language acquisition, meaning that it is the most difficult to acquire. She added that once functional morphology is acquired, the task of acquiring syntax, semantics and pragmatics goes smoothly. Positioning functional morphology at the bottleneck of L2 acquisition resulted from the fact that it “hosts a bundle of morphosyntactic and semantic features, which may or may not be overtly expressed by a functional morpheme” and therefore influences the acceptability and meaning of the entire sentence (Slabakova, 2019, p. 2).

In an elaboration of the role of feature reassembly in L2 acquisition difficulty as discussed in Lardiere’s FRH, Slabakova (2009) proposed a cline of difficulty (Figure 3.1) to account for difficulties in grammatical features acquisition. Assuming Full Access to UG, this cline classifies grammatical features according to their relative difficulty for language learners and shows that if the L1 and L2 features match completely, less difficulty is predicted in the mapping process from L1 to L2. The difficulty increases as the need for reassembly increases alongside mapping.

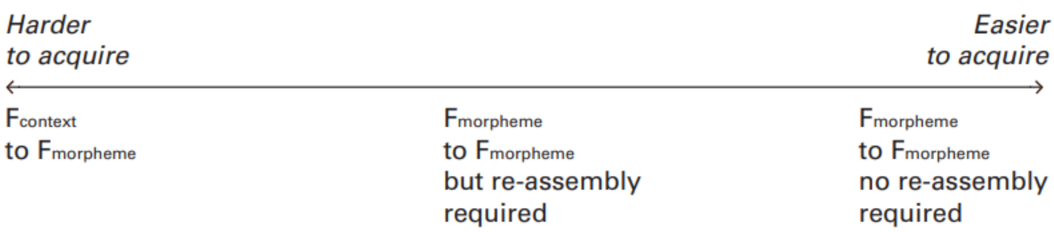


Figure 3.2 A cline of difficulty in grammatical feature acquisition, adopted from Slabakova (2009, p. 321). *F* refers to the functional feature.

In a development of this cline, Cho and Slabakova (2014) investigate how certain linguistic features are expressed differently in L1 compared to L2. They focus on whether these features are presented overtly (using morphemes) or covertly (using context) and directly or indirectly. Additionally, they examine the complexity of the relationship between form and meaning, such as whether a single feature has multiple morphophonological representations or if a single morphophonological form encompasses multiple features (Figure 3.2). This cline suggests that features which are clearly marked using morphological representation (overt) in both L1 and L2 are simpler to learn compared to those that are implied (covertly by means of discourse). Cho and Slabakova (2014) illustrated this using the expression of definiteness in English and Russian.

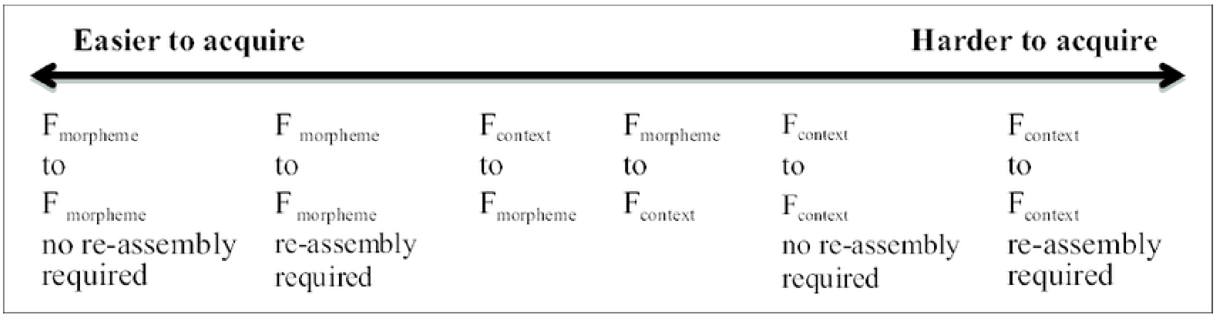


Figure 3.3 Cline of difficulty in feature acquisition. Adopted from Cho & Slabakova (2014, p.166)

Regarding direct versus indirect expression, English explicitly marks definiteness with articles ("the" for definite and "a/an" for indefinite), making it overt. In contrast, Russian lacks articles, so definiteness is conveyed indirectly (covert) and must be inferred from context. For English speakers learning Russian, this indirect expression of definiteness is challenging because they need to reconfigure their understanding of how definiteness is encoded. They can no longer rely on explicit articles and must instead use contextual clues. Conversely, Russian speakers learning English must adapt to using articles to express definiteness which is a feature not directly marked in their native language (Cho and Slabakova, 2014). This cline of difficulty is valuable in providing predictions and explanations of the learning difficulty which can be used in supporting acquisition through teaching.

According to Slabakova (2016), although it has been argued that the reassembly requirements contribute to the difficulty in the acquisition of functional morphology, the frequency of the grammatical feature has a vital role in the explanation of this difficulty; less frequent features are claimed to be more challenging to acquire because of insufficient evidence in the input received by L2 learners.

In an updated version of the BH, Slabakova (2019) proposed implications for adult L2 acquisition, explaining that acquiring L2 semantics requires little or no exposure in the absence of a syntax-semantics mismatch; therefore, it is predicted to be easier to acquire by L2 learners. When languages vary in the grammatical realisation of semantic features, difficulties arise. She catalogued some factors that complicate L2 functional morphology acquisition (other than L1 transfer): the morphosyntax–semantics mismatch between L1 and L2, the need for feature reassembly, L2 functional morpheme redundancy, the opacity of form-meaning mapping and frequency of use. Slabakova (2019) highlighted that the first two factors have a robust effect on predicting and exploring L2 learnability issues. Acknowledging that frequency, redundancy, and salience can explain slight variations in the acquisition process, Slabakova (2019) mentioned these factors, although they originate from a usage-based approach. Considering Slabakova's view of the vital role of mapping mismatch and assembly, a brief discussion of these two factors follows.

The mismatch in morphosyntax–semantic mapping in L1 and L2 is considered a core factor in the difficulty of functional morphology acquisition according to the BH. Slabakova (2019) exemplified this by discussing the generic and specific meanings and the definite plural in studies by Ionin and Montrul (2010). Ionin and Montrul (2010) found that Spanish-speaking learners faced difficulty, as the result of a mismatch between form meaning mapping in L1 Spanish which mapped the definite plural to the generic meaning in addition to the specific one and L2 English which does not allow this form to be mapped to the generic meaning. In a follow up study, they found that advanced Spanish-speaking learners of L2 English, who study abroad and received more naturalistic input and practice, successfully mapped the definite plural to existential meaning in English and overcame L1 transfer (Section 3.3.2 provides a detailed review of this study). Slabakova (2019) emphasized that learners must not only acquire new grammatical forms but also correctly associate these forms with their meanings. This requires understanding the subtle nuances and contexts in which these forms are used.

Slabakova (2019) supported the view of the feature assembly process as a factor affecting the acquisition difficulty. By discussing Hwang and Lardiere's (2013) results regarding the acquisition of Korean L2 by English-speaking learners, Slabakova illustrate the role played by the need for feature reassembly. Hwang and Lardiere's (2013) investigation focused on the Korean plural morpheme "–tul," which functions both intrinsically and extrinsically. Intrinsically, it marks plurality on nouns, similar to English plural morphology. However, unlike English, Korean also uses "–tul" extrinsically — it can attach to predicates, adverbs, and locative phrases to express distributive meanings. The acquisition challenge lies in the need to reassemble the feature bundle associated with "–tul". English learners must go beyond their L1 understanding of [plural] as a noun-bound feature and reconfigure it to include additional syntactic environments (e.g., predicates, adverbs) and semantic roles (e.g., distributivity). Therefore, what is being reassembled is the mapping between the morpheme "–tul" and its broader functional features, which include both plurality and distributivity, across multiple syntactic domains in Korean.

According to Slabakova (2019), the complexity and lack of transparency in the relationship between linguistic form and meaning play a role in acquisition difficulty. When learners encounter obscure structures, where the mapping between the form and its meaning is not clear, it can increase their acquisition difficulty. Opacity in form–meaning mapping arises when the relationship between a linguistic form and its semantic interpretation is not transparent, consistent, or straightforward. In such cases, learners cannot easily infer the meaning of a morpheme or structure based solely on its surface form or context.

For example, in English, the bare plural can express both generic and specific readings depending on context ("Dogs are friendly" vs. "Dogs chased me"), while definite and indefinite articles interact with number and specificity in nuanced ways. These variations obscure the mapping between a particular form (e.g. a bare noun or article) and its meaning. From the learner's perspective, such ambiguity requires extra processing effort and often results in delayed or incomplete acquisition, especially if the L1 encodes the same semantic features differently—or not at all.

Slabakova (2019) argues that this opacity complicates the acquisition process because learners may not receive enough consistent input to induce the correct mapping. The challenge is not only about noticing a form but also understanding what conceptual information it encodes across varied contexts. When this mapping is opaque, learners may misinterpret or fail to acquire the intended semantic contrasts altogether.

Similar to opacity is the functional redundancy, which refers to the presence of multiple grammatical markers that convey the same or similar information. In languages with high functional redundancy, learners may face difficulty due to the need to learn and differentiate between several forms that serve similar functions.

Generally, Slabakova (2019) clarified the pedagogical implications of understanding what is easy and difficult in language acquisition, explaining that it enables the language instructor to devote enough time and effort to what is challenging rather than wasting time on what comes more easily. In other words, overcoming the difficulty in functional morphology requires 'smart practice', referring to the focus on difficulties in classroom practice as an essential element in language learning.

Both the FRH and the BH offer a detailed explanation of the L2 acquisition process and propose a thorough discussion of possible reasons for the difficulty an L2 learner may face. They highlight functional morphology as the fundamental area of difficulty and name the L1-L2 form–meaning mismatch and reassembly requirement, the complexity of L2 form-meaning mappings, the frequency of functional morphology in the input and accessing and processing lexical items as possible explanations of the difficulty which may differ among languages combinations and language learning contexts. With this background in place, this study will review the empirical studies that investigate genericity acquisition in L1 and L2 and account for the difficulties faced by L2 learners.

3.3 Empirical evidence on the acquisition of genericity

This section reviews key empirical studies of the acquisition of genericity with a focus on investigations of the L2 acquisition of genericity within the FRH framework. This review is structured as follows: Section 3.3.1reviews the evidence on genericity acquisition by L1 children, Section 3.3.2 accounts for genericity acquisition by L2 learners from a variety of L1 backgrounds, Section 3.3.3 provides a summary of research on Arabic-speaking learners' acquisition of English articles as a prelude to discussing how Arabic-speaking learners of English acquire a generic meaning mapping to articles by reviewing studies on the acquisition of genericity by Arabic-speaking learners of English in 3.3.3.2.

3.3.1 The acquisition of genericity by L1 children

It has been stated that generics are a universal property and that generics are complex because differentiating generic from non-generic meanings goes beyond the interaction between morphology and syntax to cover the interface with semantics, and world knowledge. However, despite its complexity children acquire this complex property early. Although the focus of this study is L2 learners of English, acknowledging how native speakers acquire generic interpretations and the challenges that they may face in form-to-meaning mapping could inform an understanding of how genericity is acquired in L2 acquisition.

Central questions in the literature on genericity acquisition concern when and how children acquire genericity and what challenges they may face. Many studies have revealed that preschool children (age 4) can produce and comprehend generic statements that differ from those of their parents (Gelman & Bloom, 2007; Gelman & Raman, 2003; Gelman & Tardif, 1998;

Khemlani et al., 2012). The literature shows two possibilities for explaining how children acquire a complex linguistic property that includes multiple form-meaning mappings. On the one hand, some scholars maintain that generic interpretation is innate (Gen operator). Hence, children acquire it freely (Pérez-Leroux, 2016). On the other hand, children may rely on their experiences with generics in the input.

The availability of generic statements in the input has attracted the attention of many studies. Gelman and Tardif (1998) considered the degree to which generic expressions could be recognised in Chinese and their recurrence compared to English. They audio-taped an hour of adult-to-child talk for Chinese and English parents during mealtime, playing time and dressing time. The sample age range is from 20 to 30 months. They found that genericity is universally salient in natural languages, although it may be more frequent in one language (e.g. English) than in another (e.g. Mandarin Chinese). Moreover, they conducted another study that focused on child-to-child talk. The sample contained 24 Mandarin Chinese children and their mothers and 24 English children with their mothers. The data was collected by audiotaping for reading picture books, matching games, and ordering toy sessions. They found that cross-linguistic differences may affect the frequency of producing generic statements and they found that generics are more frequent with animal domains in English compared to Mandarin.

Considering how children acquire generic interpretations and what role morphosyntactic (+/- articles) and pragmatic cues (anaphoric reference) play, Gelman and Raman (2003) investigated the role of these cues in children's acquisition of genericity, conducting two studies. In the first study, they examined children's sensitivity to definite and bare plurals by showing them pictures of two unusual entities (e.g. two penguins, unusual birds that do not fly) and asked generic and non-generic questions (Do birds fly? versus Do the birds fly?). Participants included adult college students, children above four years old and two to three-year-old kids. They found that children differentiated between bare and definite NPs by giving generic answers using bare plurals ("Birds fly") and non-generic ones for the definite plurals ("The birds do not fly").

Khemlani et al. (2012) suggested that children and adults rely on beliefs to make a judgment that a specific member of a type will have the relevant feature, a point of view known as the generic hypothesis. In another account, Gelman and Bloom (2007) conducted a study to investigate how generic nouns are understood by using scenarios followed by questions with the sample including four-and-five years old kids and undergraduates. They found that children and adults grasp the generic structure in similar ways, but that they differ in their mental representations of properties. They added that while children consider generics to be communicating any sort of regularity, adults consider generics to be communicating more essential properties.

In a discussion of the learnability challenges faced by children in acquiring genericity in their target language, Pérez-Leroux (2016) investigated the learnability challenges faced by children at various stages of their target language development (pre-school to early school age) and discussed possible problems, including semantic complexity and morphosyntax (form-meaning mapping). The findings suggested that despite its semantic complexity, children use generic expressions correctly from an early age. They might initially apply general meanings too broadly, but they soon learn to use the relationships between form and meaning. They also use the necessary contextual and grammatical clues to understand generic statements accurately.

The literature proves that neither the Gen operator nor complex semantic considerations are problems for children, as they are either innately acquired or provided by the input, but morphosyntax showed some evidence of generic overgeneralisation. Despite this difficulty, with experience, children successfully use generic expressions in spontaneous speech and develop accurate form-meaning mappings. It is quite possible that a learnability issue in L1 can inform L2 acquisition investigations where genericity form-meaning mappings cross-linguistic differences add to the acquisition difficulty. The following section reviews studies on genericity acquisition by L2 learners.

3.3.2 The acquisition of genericity by L2 Learners from various L1 backgrounds

Much of the literature on the acquisition of generic meaning highlights the influence of cross-linguistic differences in bilingual children. Studies have shown that bilingual children often interpret generic expressions differently from monolinguals, with these differences attributed to the interaction between their two languages (see Kupisch and Snape, 2024, for a review). While this line of research is highly relevant, a detailed discussion of bilingual child acquisition is beyond the scope of this section. Instead, I focus on adult L2 learners, who, according to the Full Transfer, bring the full representational system of their L1 to the L2 learning task. As with bilingual children, cross-linguistic differences may affect the acquisition of genericity in L2 learners, alongside the influence of L1 transfer.

The effects of cross-linguistic differences have been investigated in L2 acquisition, and the current literature has argued that cross-linguistic differences between the learner's L1 and L2 are one factor influencing L2 acquisition difficulty (Lardiere, 2009; Slabakova, 2013,2016,2019; Rothman & Slabakova, 2018). Researchers attempted to use the hypotheses discussed earlier in 3.2.2 to explain L2 learners' difficulty acquiring genericity. Ionin & Montrul

(2010) examined the role of the learner's L1 in the acquisition of the definite plural in English, as well as the possibility of overcoming the effect of the L1. Using an acceptability judgement task (AJT), a truth-value judgement task (TVJT) and a picture–matching task, they tested 24 Spanish-speaking learners (article language) and 29 Korean-speaking learners (article-less language). Spanish uses the definite plural NP in generic expression. Therefore, it was hypothesised that Spanish-speaking learners would transfer the L1 form and interpret the definite plural as generic. Korean expresses generics using a bare NP with no plural markers and the study hypothesised that Korean learners would acquire the specific interpretation of the definite plural with no difficulty, but they may face difficulty in interpreting bare plurals as generic in English under the effect of L1.

The results confirmed these hypotheses and revealed distinct patterns of L1 transfer. Despite being matched for overall proficiency and article accuracy, the Korean speakers were significantly more accurate in interpreting definite plurals as specific, whereas Spanish speakers often interpreted them as generic, reflecting influence from Spanish. However, for bare plurals, the pattern reversed: although both groups performed relatively well, Korean speakers were less accurate than Spanish speakers, at times interpreting bare plurals as specific—a pattern that mirrors how genericity is expressed in Korean. Thus, the findings show that while Spanish-speaking learners transferred a generic interpretation onto English definite plurals, Korean-speaking learners transferred a specific interpretation onto English bare plurals. This result was explained in terms of L1 transfer, suggesting that the Spanish learners were not at the end state of their acquisition and therefore faced difficulty in resetting the features of the L1 to those of the L2 under the effect of less exposure as being in an EFL context.

In a follow-up study, they tested advanced Spanish-speaking learners immersed in the United States to assess recovery from L1 transfer. The results confirmed that L1 Spanish advanced–level L2 learners showed recovery from L1 transfer: They were able to interpret both definite and bare plurals in a target–like manner. They concluded that with extensive exposure, recovery from L1 transfer is possible. Although this study provides empirical evidence of the recovery from L1 transfer, its scope was limited to the definite plural, where the differences between the learners' L1 and L2 are one-to-one. Lardiere (2009) suggests that the learning task may become more difficult in situations where the reassembly conditions (the feature is morphologically realised as inflectional or free, covert, or overt, and interpretable or uninterpretable) highlighting the need for more considerations of such conditions. MSA behaves like Spanish in allowing the definite plural with a generic interpretation but differs in allowing bare plurals to have indefinite and nonspecific interpretation in subject position; hence, the effect of MSA as learners' L1 is worth more investigation. The results and discussions in Ionin and Montrul's study can inform the discussion of the plural generics' learnability by Arabic-speaking learners of English.

In a wider-scope study, Ionin et al. (2011b) investigated the acquisition of the difference between the definite, indefinite singular and bare plural generics in English by Korean and Russian learners. The study focused on investigating whether learners from article-less languages can differentiate between the two types of genericities in English as L2. The sample included 45 Korean-speaking learners and 33 Russian-speaking learners as well as a native control group. The L2 learner groups included participants at different proficiency levels. In addition to a cloze proficiency test and a language background questionnaire, the participants filled out a Likert scale AJT. This task consisted of generic and anaphoric singular and plural contexts, and the latter context was used as a control category. The function of this control category was to provide information about the learners' acquisition of the basic article system in English.

The results of the control native speakers validated the AJT; they accepted bare plurals and definite singular in kind generics and bare plurals and indefinite singular in characterising generics. The L2 learners' results revealed that both Korean and Russian learners of English were sensitive to the morphological differences between the two types of generics. Moreover, while sentence-level generics caused no difficulty for both groups, they faced difficulty in the kind generics interpretation. Both groups' L2 learners were target–like in accepting bare plural and indefinite singular in sentence-level generics and non–target–like in their interpretation of the definite singular in the context of NP–level genericity. Very few advanced learners showed target-like performance allowing the definite singular in NP-level generics. They interpreted this difficulty as informed by the FRH, stating that the struggle results from the learners' need to select two features [+definite, +taxonomic]. This study provides an exemplary investigation of both singular and plural generics with participants from article-less languages. However, the conclusion about the possibility of recovery from L1 is limited to learners whose L1 does not have articles. This highlights the need for studies that consider L2 learners with an L1 with an article system.

Considering learners from both article and article-less L1 backgrounds, Snape (2013) replicated Ionin et al. (2011b). The study investigated the role of L1 in the acquisition of both NP-level and sentence-level generics by examining how well learners can rate definite singulars for NP-level generics, indefinite singulars for sentence-level generics, and bare plurals for both NP-level and sentence-level generics in L2 English. The study sample included 24 advanced Japanese-speaking learners, 18 advanced Spanish-speaking learners and 35 native English controls. The study predicted that Japanese learners would struggle with definite singular generics for NP-level generics but might find indefinite singular generics easier if they understand English articles. Bare plurals should be less challenging as they don't require an article. Spanish learners will easily identify the correct article for definite singular NP-level and indefinite singular sentence-level generics but will struggle with plurals since English uses bare plurals for generics, unlike Spanish, which uses a definite article.

The researcher tested these predictions using an adapted version of the Likert scale AJT from Ionin et al. (2011), which included control anaphoric singular and plural contexts and NP-level and sentence-level generic contexts. The results of the control anaphoric contexts revealed that the L2 groups rated the target sentences much higher than the non-target sentences. They also recognised the difference between the singular and plural control categories. As for the generic contexts, as expected Japanese learners found definite singulars for NP-level generics harder than indefinite singulars for sentence-level generics. Their ratings for definite singular NP-level generics were almost the same as their non-target ratings. Bare plurals were generally less problematic, but Japanese learners rated bare plurals higher for NP-level genericity than for sentence-level genericity. Spanish-speaking learners accurately rated the correct singular NPs higher than the non-target NPs in both levels, as expected. Surprisingly, they did not show the predicted difficulty with bare plurals for both NP-level and sentence-level generics, giving significantly higher ratings to bare plurals than to definite plurals. The researcher suggested that the lack of articles in Japanese might explain the difficulty in using definite singulars at the NP-level generics as the learners need to acquire the definite article and assemble the [+taxonomic] feature to it which has no parallel in the L1.

Snape et al. (2013) elaborated the investigation on the role of L1 by increasing the number of L1 backgrounds and using a forced choice task. The aim was to investigate the role of the L1 in the acquisition of the definite singular in NP-level generics and the indefinite singular, bare plural, and mass nouns in sentence-level generics. The sample included upper-intermediate and advanced L2 learners of English, of whom 33 were Japanese-speaking, 50 Spanish-speaking, and 88 Turkish-speaking, in addition to 17 English native speakers as the control group. Based on cross-linguistics differences, the study tested the same predictions of L1 Japanese and L1 Spanish learners in Snape (2013). As for the L1 Turkish learners, the learners L1 can express generic meanings using bare plurals, bare singular, and the indefinite article "bir," but it does not have a definite article. Therefore, the study predicted that these learners may face challenges in choosing "the" and often mistakenly omit the definite article in NP-level generics under the effect of L1 transfer.

To test these predictions, participants were given a forced-choice task in the classroom. The results revealed various patterns created by the effect of the L1s. The Spanish-speaking learners performed similarly to native speakers in using bare plurals, and their L1 aided in the acquisition of definite singular at the NP–level generics. Whilst Spanish–speaking advanced learners of English showed a target–like use of articles in generic interpretation, Turkish and Japanese speakers had more problems with definite article usage in NP–level generics. The indefinite article for sentence–level genericity was less problematic for the Turkish group, showing the effect of L1 positive transfer. The researchers reported the need for more investigation of the difficulty in using the definite singular in NP-level generics. They acknowledged that their results were limited to using one measure (FCT) and that unbalanced sizes for the groups may affect the findings. This study provides methodological considerations for future research on genericity.

Focusing on definite singular generics, Snape (2018) explored the acquisition of the definite singular generic and unique meanings by 47 Japanese-speaking learners of English. The study predicted that although the sample's L1 has no articles, the meanings expressed by the generic and unique definite singular are available, as these are universal, acknowledging that they may face challenges due to the required feature reassembly. Moreover, the study focused on whether the Japanese-speaking learners' responses differed between different tasks. Therefore, the study used a picture-matching task and a forced–choice elicitation task. The sample included upper-intermediate and intermediate Japanese-speaking learners and 26 English native controls.

The results of the first task showed that Japanese learners of English can comprehend the universal concepts of genericity and uniqueness. Some Japanese L2 learners performed comparably to native speakers in their selections in the picture matching task. The results of the second task reflected the opposite pattern; participants failed to choose the definite article for singular kind generics at a rate of around 83%. The results of this study support the view that difficulty results from the need to map semantic features to new morphemes in L2. This study highlights the effect of task type, and this informs this study to combine different task types in the investigation of the acquisition of genericity.

Taken together, this section presents empirical evidence highlighting the challenges L2 learners encounter in acquiring generic form-meaning mappings and the influence of L1 transfer on this process. Several studies have examined L2 learners from article-less languages, such as Korean and Russian (Ionin et al., 2011b), and Japanese (Snape, 2018). Other studies have compared learners from article-less L1s with those from L1s that include articles, such as

Spanish, Turkish, and Japanese (Snape et al., 2013), as well as Spanish and Korean (Ionin & Montrul, 2010). These studies demonstrate that L1 transfer can impact the acquisition of genericity. The review provides evidence that the presence or absence of articles in the L1 can contribute to difficulties in acquiring generic form-meaning mappings, depending on L1-L2 differences and similarities. Additionally, it underscores the importance of input exposure in helping advanced Spanish learners overcome L1 transfer and adapt to L2 form-meaning mappings. Furthermore, the review shows that L2 learners' performance varies across different tasks. Consequently, it can be inferred that L1 Arabic-speaking learners would exhibit performance patterns consistent with their L1 due to language transfer. The following section reviews empirical evidence on the acquisition of genericity and the role of the L1 among Arabic-speaking learners of English.

3.3.3 The acquisition of genericity by Arabic-speaking learners of English

This section examines significant studies on the acquisition of genericity by learners whose L1 is Arabic. To provide a foundation for discussing the literature on the acquisition of genericity, I first review research on how Arabic-speaking learners acquire basic knowledge of articles. This overview aims to establish a baseline understanding of their knowledge of articles before delving into the specifics of genericity.

3.3.3.1 Baseline Knowledge of English articles by Arabic-speaking learners of English

It has been suggested that Arabic-speaking learners of English encounter difficulties in acquiring English articles, despite the presence of an article system in both languages. Early investigations on article acquisition were concerned with error analysis (Willcot, 1978; Bataineh, 2005; Crompton, 2011; Alhaysony, 2012). Later studies move beyond error analysis and consider cross-linguistic differences and L1 transfer as well as L2 complexity and proficiency levels in explaining English article acquisition by Arabic-speaking learners (Sarko, 2009; Abudalbuh, 2016; Hassan &Eng, 2018; Aboras, 2020a). A review of key findings on the difficulty faced by Arabic-speaking learners in these studies follows.

An early study conducted by Willcot (1978) focused on the written English of Arabic-speaking students. It provided an error analysis of issues in using definite and indefinite articles in a corpus of sixteen American history final exams written by Arabic-speaking learners. The results revealed that errors fell into different categories: deleting the definite article, adding 'the' when no article is needed commonly in generic contexts and deleting the indefinite article. Willcot found that the addition of the definite article when no article was needed was related to generic contexts and attributed this to L1 influence. The differences between the articles in the L1 and L2 caused the omission of the indefinite article.

In a similar vein, with the focus on errors in the use of the indefinite article ‘a’, Batanieh (2005) examined the type of errors in English composition writing of freshmen, sophomore, junior and senior college students in a Jordanian university. According to Batanieh, errors in the use of indefinite singular NPs included deletion of ‘a’, attaching ‘a’ to the following lexical item, substitution errors, using the indefinite article with plurals and using the indefinite article with uncountable nouns. The observed errors predominantly stem from common learning processes, such as overgeneralisation and simplification of the English article system. Crompton (2011) considered errors in articles by advanced Arabic-speaking learners of English by conducting an error analysis for a corpus of written English. The results revealed that using ‘the’ instead of no article in generic non-count and plural contexts was the most common error. This error was attributed to L1 transfer. Considering the Saudi context, error analysis studies revealed similar error types and some researchers attributed these errors to L1 transfer the complexity of the English article system as a contributing source (Alhaysony, 2012; Al-Qadi, 2017).

In a GenSLA framework, Abudalbuh (2016) studied how Arabic-speaking learners of English acquire articles, focusing on definiteness and specificity. Conducted at a Jordanian university with 30 adult learners, the study aimed to identify common errors and their causes, whether from L1 transfer or Article Choice Parameter fluctuation. The participants were divided into beginner, intermediate and advanced levels and completed a forced-choice task. The results indicate that Arabic-speakers’ article use is more accurate in definite contexts than indefinite contexts, regardless of specificity. He claimed that article omission is the most frequent error for beginners and intermediate learners, followed by the overuse of ‘the’ in indefinite contexts and the overuse of ‘a’ in definite contexts. The researcher proposes a model of the developmental stages of article acquisition by Arabic-speaking learners of English, arguing that both fluctuation and L1 transfer function concurrently in the early stage, causing two kinds of errors by beginners: ‘*a/the’* overuse and article omission, respectively. Under input guidance, learners stop fluctuating, gradually making fewer overuse errors. However, they continue omitting articles in compulsory contexts. The final stage shows how learners reach target-like performance under input processing. He highlights proficiency as a factor that influences L2 article acquisition.

Taken together, this review shows that Arabic-speaking learners of English find articles challenging and face difficulties due to multiple sources. Namely: the effect of L1 transfer, the

L1 and L2 differences, the complexity of the English article's semantics, and proficiency. Advanced learners show the ability to acquire basic article semantics, such as definiteness. The following section reviews key studies that focused on the acquisition of generic form-meaning mappings by Arabic-speaking learners of English.

3.3.3.2 Studies on the acquisition of genericity by Arabic-speaking learners of English

This section focuses on reviewing empirical evidence on the acquisition of genericity and the role of the L1 among Arabic-speaking learners of English. Few studies in the literature have focused on the acquisition of genericity by Arabic-speaking learners of English. Among these studies are Alzamil (2019), Hermes (2020) and Aboras (2020) in addition to two intervention studies conducted by Sabir (2015) and Abumelha (2017). All of these studies agree that Arabic-speaking learners of English find acquiring the generic form-meaning mapping challenging, highlighting the cross-linguistic differences, the L1 transfer and reassembly requirements as possible explanations of the difficulty. Studies on the acquisition status are presented here and intervention studies will be discussed in Section 4.4.2 within the discussion of the role of instruction.

Alzamil (2019) investigated whether Arabic–speaking learners of English are sensitive to the two types of generics in English and what difficulties Arabic-speaking learners may face. To do so, in addition to a proficiency level test, a Likert scale AJT with context was adopted from Ionin (2011b) and used with 36 Arabic-speaking learners of English (23 elementary and 13 low-intermediate) and 7 native speakers who set the baseline. The results revealed that both L2 learners' groups highly rated definite plurals as generics in both types of generics, unlike the native controls, who rated bare plurals as acceptable. Moreover, both elementary and low–intermediate groups rated definite singulars, definite plurals, and bare singulars (unacceptable forms) higher than the native control group. This study also found that bare plurals were rated higher with sentence-level generics and that definite singulars were rated higher with NP–level generics, suggesting that Arabic–speaking learners are sensitive to the distinction between the two types of genericity despite the acquisition difficulty. Sentence-level genericity was more difficult than NP-level genericity. The acceptance of the unacceptable definite plural was attributed to reliance on the L1 at an early stage of acquisition. This reliance on the L1 was noticed in the elementary group's rating of definite singulars higher than the lower-intermediate group with NP-level genericity. The study suggested that learners may overcome the difficulty once their proficiency improves. This study highlights the L1 effect on mapping articles to generic meanings for L1 Arabic speakers. However, the data comes from AJT only and the conclusions may not hold if different measures and parametric statistics were considered.

Hermes (2020) investigated the acquisition of genericity in English as an additional language by L1Arabic-L2 French adults. This study examined if a learner's understanding of genericity at an advanced level in English shows signs of influence from their L1 Arabic or their L2 French. Data was collected from 27 Moroccan-Arabic speaking learners and 12 English native speakers using a Likert scale AJT. The results revealed that the sample has no difficulty in accepting the definite singular and bare plural, significantly higher than unacceptable forms in NP-level generics. The low ratings of the bare singular indicated the successful acquisition as indicated the potential learning ability.

The results highlight the successful role of input in supporting the acceptance of bare plurals in both types of generic, stating that it supports the learners' mastery of the generic reading of bare plurals, which is different from their L1. Although the sample shows successful acquisition of the acceptable forms, they still face learning difficulty with definite plurals. Unlike the native speakers, the Arabic-speaking learners' ratings were around the mid-scale in the two levels of genericity. The researcher suggested that this fluctuation was due to transfer from L1 Arabic (and L2 French), where definite plurals can express genericity. Moreover, the results revealed that the indefinite singular was restricted to the existential reading and was not accepted in generic reading. The difficulty caused by definite plurals acceptance in generic meaning is aligned with Alzamil's (2019) results. Although this study is concerned with L3 acquisition, it provides informative results on the impact of Arabic on the acquisition of English generics and is therefore relevant. It highlights areas of difficulty, and it is the only study that included advanced-level Arabic-speaking learners of English; therefore, it discusses difficulties at a near-complete acquisition stage. Such difficulties can inform the intervention of this study to focus on these difficulties in teaching.

In the same vein, Aboras (2020) investigated Saudi-Arabic learners' accuracy with generic and anaphoric references in English as an L2 and whether they showed sensitivity to the morphological distinction between characterising generics and kind generics. In particular, the study tested the effect of the L1, vocabulary knowledge and proficiency levels on the learners' accuracy. The study was conducted using a Likert scale AJT with contexts and a forced choice task to measure accuracy in generic form-meaning mappings in addition to two tests to measure receptive and productive vocabulary knowledge and a proficiency measure. The sample included 160 undergraduate Saudi Arabic-speaking learners who were enrolled in the English language and literature department (80 low-intermediate and 80 high-intermediate) in addition to 10 native controls.

The tasks included control and test categories; the anaphoric use of articles composed the control category, and the generic contexts composed the test category. The results of the AJT revealed that both low- and high-intermediate learners showed high accuracy in the control categories as they accurately accepted the correct NPs in anaphoric singular and plural contexts. For the test categories (generic contexts), both groups showed accuracy in accepting the definite singular and bare plurals in NP-level generics. However, they faced difficulty in accepting the indefinite singular NP in sentence-level generics as both groups overused the definite article in this context. Despite this, they were accurate with the sentence-level bare plural. The results of the forced-choice task showed that both low- and high-intermediate learners were highly accurate in selecting the definite singular in NP-level generics, the zero article (bare plurals) in NP-level generics and the indefinite singular in sentence-level generics. However, both groups showed less accuracy selecting bare plural in sentence-level generics, indicating difficulty.

The influence of vocabulary knowledge was assessed by investigating the relation between productive and receptive vocabulary test results and accuracy in both tasks. The results of multivariate linear regression revealed that productive vocabulary knowledge was significantly related to the anaphoric plural for low-level learners and to NP-level singular generic and anaphoric for high-level learners. Receptive vocabulary knowledge was also significantly associated with NP-level plural generics for low-level learners and only with sentence-level plural generics for high-level learners. The proficiency effect was significant for low-level learners only, as a significant relationship was found between proficiency level and NP-level generic mapping to bare plurals.

This study discusses the results in relation to L1 transfer, proficiency, and vocabulary knowledge, stating that all three factors affected accuracy and contributed to the difficulty in genericity acquisition. Proficiency level affected accuracy in using bare plurals with NP-level and sentence-level generics and receptive vocabulary affected accuracy in using the indefinite article with sentence-level generics. A limitation that Aboras (2020b) acknowledges is that NP-level generic plural forms in the FCT were limited to having the definite article as the target as in ‘the Germans’. Therefore, we are not told about the acquisition of bare plurals in NP-level generics in the FCT and whether different tasks could lead to differences in accuracy in mapping this form to kind generics. This study suggested that the acquisition of genericity in the classroom needs to be addressed in further research as there is a lack of research in this area. Aboras's (2020) study investigated genericity acquisition in an EFL context and by an Arabic-speaking sample similar to the present study; hence, it is highly relevant. It informs this study to test the acquisition of genericity using different measures that all share similar generic contexts to enable the comparison between the L2 learners' responses in different tasks. Moreover, the classroom aspect of genericity acquisition is considered in this study as suggested by Aboras (2020).

3.4 Summary

This chapter discusses the conceptualisation of L2 learning, presents theoretical proposals that account for the difficulty in the L2 acquisition of form-meaning mappings and reviews empirical evidence on the L1 and L2 acquisition of genericity, the linguistic property in this study. The studies reviewed in Section 3.3 highlight that the acquisition of generic form−meaning mappings is challenging for L2 learners from different proficiency levels. These studies show that L1 transfer can affect the acquisition of genericity and provide evidence that the presence or absence of articles in the L1 can contribute to difficulties in acquiring generic form-meaning mappings. Few studies focused on genericity acquisition by Arabic-speaking learners of L2 English. These studies agree that Arabic-speaking learners of English find acquiring the generic form-meaning mapping challenging, highlighting cross-linguistic differences, L1 transfer and reassembly requirements as possible explanations of the difficulty. However, the picture of acquisition difficulty, the role of L1 and cross-linguistic differences are not yet complete as these studies were limited to a specific variety, a specific proficiency level or limited in scope to cover a specific form-meaning mapping. Therefore, more investigations are required to increase the evidence in order to have a better understanding of the acquisition difficulty and its possible explanations. Therefore, this study adopts the FRH and the BH to establish predictions about the learnability of generic interpretations in English as an L2 by Arabic-speaking learners.

The reviewed studies highlight some key points for this research. Firstly, it's important to have an L1 baseline group to set a target-like standard, validate tasks, and show possible variability in L1 form-meaning mappings, highlighting L2 complexity. Secondly, many studies use only one measure of acquisition, but using multiple measures, as seen in Aboras (2020) and Snape (2018), can provide a clearer picture of acquisition difficulty and its sources. These considerations are included in this study's design.

The studies also have implications for learning genericity in EFL contexts. Ionin & Montrul (2010) found that advanced Spanish-speaking learners of English struggle with mapping bare plurals to generic meanings in EFL settings. The quality and quantity of input are crucial, as learners with more naturalistic exposure show better recovery from L1 influences. Aboras (2020) suggests that teachers should emphasize specificity and genericity when teaching articles, not just definiteness. This study aims to bridge the gap between GenSLA research and practical application, especially in EFL L2 acquisition. The next chapter will detail this bridging process and the theoretical foundations of the instructional intervention used in this study.

Chapter 4 Instruction in SLA and Genericity Acquisition

4.1 Introduction

The previous chapter highlighted cross-linguistic differences, L1 transfer, and L2 input as influential factors that impact the acquisition of genericity. Ionin and Montrul (2023) shed light on the context of L2 acquisition as a factor that relates to the acquisition difficulty by discussing the differences in the quality and quantity of linguistic input received in different acquisition contexts; adult L2 learners in an instructed L2 learning context received less input compared to L2 learners in the L2 speaking context. Although instruction as a context of L2 learning is not directly related to GenSLA research, the latter's findings could have implications for the language classroom that contribute to accomplishing the aim of successfully communicating in the L2 (Marsden & Slabakova, 2018).

The aims of this chapter are to i) elaborate on the concepts related to instruction in SLA including the input types, knowledge types and instruction types, ii) give an overview of teaching grammar in SLA, iii) discuss the impact of instruction in L2 learning through reviewing the literature concerning the role of instruction in the acquisition of genericity and iv) operationalise the instruction in this project. The chapter begins by discussing instruction as a concept and clarifies what instruction means in this thesis. Then, it examines the context of L2 learning and the type of input it provides. Section 4.3 gives a note on grammar teaching approaches. Section 4.2 introduces important conceptual distinctions: explicit vs. implicit knowledge, learning and instruction. Finally, Section 4.4 reviews intervention studies that focus on genericity.

4.2 Instruction in SLA

L2 Instruction is an umbrella term incorporating different fields that cooperate to enable L2 instruction to achieve its broad goals, enabling successful communication in L2. Its focus broadly covers cognitive, social, and methodological elements that cooperate during instruction. This investigation assumes a GenSLA perspective; therefore, it is necessary to clarify precisely what instruction means for this study. Instruction in SLA can be broadly defined as “understand[ing] how the systematic manipulation of the mechanisms of learning and/or the conditions under which they occur enable or facilitate the development and acquisition of an additional language”, as stated by Loewen (2015, p. 2). In other words, instruction in L2 acquisition implies systematically manipulating the L2 learning conditions and mechanisms to develop the L2 learner's grammar. This study adopts Marsden and Slabakova's (2018) conception of language teaching in their introduction of the special issue that assembles empirical work on second language teaching and learning from a generative linguistic perspective. They state that since linguistic properties are the focus of generative linguistics, which findings can offer pertains to language structure, instruction can be viewed as engaging with the grammar element of language teaching. They clarify the importance of all other language-teaching elements and that this view is not meant to promote teaching grammar over all other elements of language instruction.

4.2.1 Instruction and evidence in the input

The nature of the input provided in L2 instruction as a context of L2 acquisition plays a role in L2 acquisition. In particular, the nature of the linguistic evidence provided for L2 learners in an L2 classroom context differs from the nature of the evidence received by L1 learners in a naturalistic setting. The L1 child receives positive evidence in the input through the naturalistic speech they heard. The "utterances in the input" constitute the positive evidence defined by White (1991, p. 134), which informs the child about the possible structures and lexical words. In contrast, the L2 classroom input includes negative and positive evidence (VanPatten, 2009). The negative evidence includes information about ungrammaticality (White, 2003). Positive input comprises the naturalistic input in speech and written texts, whereas instruction and error correction and feedback provide the negative evidence in the L2 classroom, according to Ionin and Montrul (2023), who situate the negative evidence provided by instruction (the role for instruction) in Schwartz and Sprouse's (1996) Full Transfer/ Full Access model of universal grammar in L2 acquisition, stating that error correction, feedback and grammar instruction, along with written language and natural speech, constitute the L2 input, as Figure 4.1 shows.

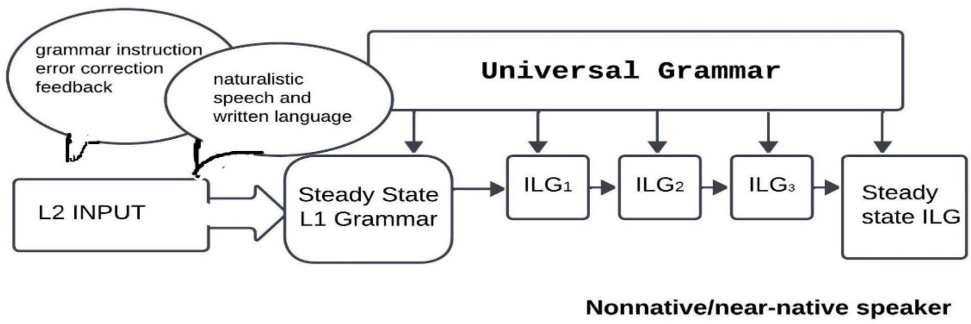


Figure 4.1 The Full Transfer/Full Access model of Universal Grammar in L2 acquisition (Schwartz & Sprouse, 1996), adopted from Ionin and Montrul (2023, p. 5).

Some researchers see that negative evidence is not essential for L2 acquisition as L1 interlanguage structure is only affected by positive evidence (Schwartz, 1993). However, several lines of evidence show that negative evidence plays a role in L2 acquisition. White (1991) suggests that providing form-focused instruction with the inclusion of negative evidence helped the French learners in her study to understand adverb placement principles in English. However, this instruction did not show a long-term effect of the negative evidence on the learners' knowledge as they did not retain the knowledge when tested five weeks later.

Hirakawa et al. (2019) claim that positive evidence alone does not guarantee the acquisition of adjective order by Japanese learners of English. In this study, Hirakawa et al. (2019) investigated the acquisition of adjective order in three settings: explicit instruction, naturalistic acquisition by study abroad and input flooding by conducting two studies. The first study focused on explicit instruction and naturalistic acquisition through study abroad. The results revealed that explicit instruction significantly improved understanding and use of the correct adjective order in English. In contrast, learners who studied abroad without explicit instruction showed a different level of improvement. The second study compared the effects of input flooding—where learners were exposed to numerous examples of correct adjective order for 15 weeks—to naturalistic acquisition. The results indicated that input flooding increased awareness of adjective ordering, but it was not as effective as explicit instruction in helping learners internalise and use the rule accurately. Overall, Hirakawa et al.'s (2019) findings suggest that while exposure to positive evidence (correct examples) is beneficial, it is not sufficient to acquire specific grammatical rules. Explicit instruction, which often includes explanations, rules, and structured practice, appears to be necessary for Japanese English learners to fully acquire the order of adjectives.

4.2.2 The Explicit and Implicit Distinction

4.2.2.1 Explicit vs. Implicit Knowledge

The main aim of instruction is to develop the learner's implicit linguistic knowledge. However, a significant concern is what constitutes this knowledge. Krashen (1981, 1982) epistemologically distinguished between knowledge gained through acquisition and learning. In particular, he contended that exposure to natural input leads to acquisition and that language knowledge is acquired automatically and implicitly. By contrast, teaching about a language's grammar leads to learning, which includes explicit knowledge about a language, derived through instruction. In a refinement of Krashen's view, Paradis (2009) proposed that two types of memory play a role in second language acquisition (SLA): procedural memory, which supports the development of automatic, unconscious skills, and declarative memory, which underlies conscious, factual learning. He argues that while procedural memory is essential for implicit knowledge — the type that enables fluent, rule-governed language use — adult learners rely more on declarative memory because procedural access for language learning declines with age.

Theories of how linguistic knowledge is acquired differ in their assumptions about its source. One view, grounded in Chomsky's Universal Grammar (UG), holds that language knowledge stems from an innate biological capacity shaped by exposure to input. An opposing, connectionist perspective suggests that linguistic knowledge — particularly implicit knowledge — emerges from gradual pattern recognition through input (Ellis, 2005). Despite these theoretical differences, both perspectives agree on the distinction between explicit and implicit knowledge (Esteki, 2014). Explicit knowledge refers to metalinguistic awareness — for instance, knowing that third-person singular verbs in English require an -s — whereas implicit knowledge allows for fluent, automatic production without conscious rule recall (Mitchell et al., 2019). A learner who knows the rule but says "Jack eat an apple" exemplifies explicit knowledge that has not yet been internalised for automatic use.

Ionin and Montrul (2023), drawing on Rod Ellis's (2005) framework, emphasise this distinction: implicit knowledge is intuitive, consistent, and typically used in fluent language tasks, while explicit knowledge is conscious, more variable, and often accessed during rule explanation or form-focused tasks. Their review also highlights that these knowledge types are stored differently in the brain — implicit knowledge in procedural memory and explicit knowledge in declarative memory. For example, native speakers often apply complex grammatical rules accurately without being able to explain them — as seen in a Spanish speaker's use of the subjunctive — while L2 learners may be able to describe a rule (like past tense marking) but still produce non-target-like output, such as "I walk to the store yesterday."

The relationship between these two types of knowledge and whether explicit knowledge becomes implicit is debated in the literature. Different positions exist in the literature regarding this relationship: the no interface position, the strong interface position, and the weak interface position. The no-interface position (Krashen, 1981, 1985; Schwartz, 1993) claims that the two types of language knowledge—implicit and explicit—are stored in different areas of the brain. Implicit knowledge, which is automatic and unconscious, is generally associated with the basal ganglia, while explicit knowledge, which involves conscious effort and awareness, is linked to the prefrontal cortex. Therefore, the possibility of an interface between explicit and implicit knowledge is impossible, leading to the view that explicit knowledge can never be transferred to implicit knowledge. This position dismissed the role of instruction, stating that instruction may impact explicit knowledge, “the learned linguistic knowledge,” but have no impact on implicit knowledge, “linguistic competence” (Schwartz, 1993, p.150).

Sharwood Smith (1981) proposed the strong interface position, arguing that explicit knowledge can be turned into implicit knowledge and vice-versa and that the move from declarative to procedural knowledge is possible through practising and rehearsal for declarative knowledge. DeKeyser (2003,2017) argues that grammar learning begins as declarative and explicit knowledge in most second-language instruction. Turning declarative knowledge (knowledge about something) to procedural knowledge (know-how) is essential for accurate language use. Automation involves using declarative knowledge to develop procedural knowledge and then refining it later through practice to be automatised knowledge. He highlights the role of practice in this transfer.

Arguing against these extreme views, the weak interface position proponents (Ellis, 2005; N.C. Ellis, 1994, 2005) argued that although moving from explicit to implicit knowledge is difficult, it is indirectly possible under suitable circumstances, including the learners’ developmental readiness. This means that learners must be in the right developmental stage, which allows them to acquire new explicit knowledge and convert it to automatic knowledge through explicit input with good quality and practice. N.C. Ellis (1994) stated that focusing on linguistic features in the input and explicit metalinguistic knowledge can help learners develop greater control over explicit and implicit linguistic knowledge. Ionin and Montul (2023) highlighted the implication made by the strong and weak transfer positions for language teaching, stating that, according to these positions, it is reasonable for teachers to teach grammar explicitly and expect the learnt knowledge to turn implicit. These interface positions allow for the understanding of positions towards how instruction can impact the L2 knowledge. Nevertheless, this study has no direct agenda for testing which knowledge type improved after the intervention. However, this understanding informs the tasks used to attempt tapping on both kinds.

Related to the implicit and explicit knowledge distinction is the account of the measurement methods used for each type. The abstract concept of explicit and implicit knowledge needs to be operationalised to make it a measurable construct (Ionin & Montrul, 2023). Measuring knowledge types has received attention in the literature, and several efforts have been made to develop and validate measures for each type. One of the most influential efforts is that of Ellis (2005), who operationalised the two types of knowledge by setting criteria to distinguish between implicit and explicit knowledge constructs. These characteristics include age, awareness, time pressure, focus of attention, systematicity, certainty and metalinguistic knowledge. Ellis (2005) created five English language examinations using these characteristics to evaluate explicit and implicit knowledge. These examinations included a metalinguistic knowledge test, an oral narration test, a timed grammaticality judgement test (GJT), an untimed GJT with identical grammatical structures and an oral imitation test. The tasks were administered to twenty English native speakers and 91 L2 learners at various proficiency levels. The findings showed that while the scores from the metalinguistic knowledge exam and the untimed grammatical judgement tasks tap into explicit knowledge, the scores from the oral imitation test, oral narration test and timed grammatical judgement task tap into implicit knowledge. He clarifies that the task that seeks the L2 learner's intuitions is timed, focuses on meaning, does not involve metalinguistic knowledge, and taps into learners' implicit knowledge. In contrast, tasks that tap into explicit knowledge include awareness of the rule, untimed, focus on form and metalinguistic knowledge.

Bowles (2011) replicated the tasks in Ellis's (2005) study by testing L2 learners and heritage speakers of Spanish, validating Ellis's findings. Moreover, Godfroid et al. (2015) validated Ellis' grammatical judgment tasks using eye-tracking (an online measure). They provided support that timed and untimed tasks tap into different constructs and types of knowledge. Similarly, Spada et al. (2015) validated the elicited imitation task in Ellis's study to measure implicit knowledge by comparing it to other measures, including timed and untimed GJTs. Their finding validated that elicited imitation tasks are associated with measuring implicit knowledge as they correlate with other measures of implicit knowledge. Whong et al. (2014) pointed out that measuring implicit knowledge should gain more focus in classroom research, suggesting that timed GJTs and elicited imitation tasks are reasonable measures of implicit knowledge. Other efforts claimed that Ellis' elicited imitation measured a third type of knowledge, automatised explicit knowledge (Suzuki et al., 2023). According to these efforts, implicit knowledge differed from automatised explicit knowledge in that the former is spontaneously used without awareness. In contrast, the latter started as explicit knowledge but is polished with practice to be recalled quickly but still consciously. Suzuki et al. (2023) argued that timed form-focused tasks measure automatic explicit knowledge while real-time comprehension tasks (tasks requiring real-time processing) measure implicit knowledge. Ellis and Rover (2021) argued against this view, stating that online processing tasks are artificial and may not tap into linguistic knowledge. Understanding the debate regarding explicit/implicit knowledge is essential to this study, although there is no agenda for directly comparing the two knowledge types. This understanding informed this study by encouraging the inclusion of explicit and implicit knowledge measures to validate the findings of the impact of instruction, if any. The operationalisation of the task selection is discussed at the end of this review.

4.2.2.2 Implicit vs. Explicit learning and instruction

Moving on to discuss the explicit and implicit distinction regarding learning and instruction. Hulstijin (2005) defined the implicit and explicit learning, stating that:

> Explicit learning is input processing with the conscious intention to determine whether the input information contains regularities and, if so, to work out the concepts and rules with which these regularities can be captured. Implicit learning is input processing without such an intention, taking place unconsciously. (p. 131)

Ionin and Montrul (2023) detailed the explicit and implicit distinctions regarding learning, instruction, and experimental tasks. They associate explicit learning with awareness and explicit instruction with attracting learners' attention to forms. In contrast, implicit learning is achieved without awareness, and implicit instruction avoids attracting attention to the linguistic form. As for the experimental tasks, they mentioned that tasks are not implicit or explicit strictly, but it is how they are designed to tap into one of the knowledge types. According to them, a task is implicit if it focuses on meaning and includes time pressure to tap implicit knowledge, and it is explicit if the focus is on form and is untimed.

Ionin and Montul (2023) summarised the differences between explicit and implicit instruction, as shown in Table 4.1.

Table 4-1 Differences between explicit and implicit form-focused instruction (adapted from Ionin & Montrul, 2023, p. 21).

Implicit form-focused instruction	**Explicit form-focused instruction**
Attention attracted to the target form	Attention directed to the target form
Spontaneous instructional delivery	Planned instructional delivery
Does not interrupt (or minimally interrupts) communication	Does interrupt communication
Forms presented in context	Forms presented in isolation
No metalanguage	Use of metalanguage
Free use of target form is encouraged	Target form is subject to controlled practice

Considering grammar teaching, implicit grammar teaching refers to teaching with no agenda to provide explanations of any grammar rules. It merely focuses on meaning, leaving the students to notice the forms incidentally without being aware that they are being taught a grammatical form (Ellis, 2008). In contrast, explicit teaching directly explains the grammatical rules and presents these forms in isolation, making the students aware that they are being taught a grammatical form (Sharwood Smith, 1981). The existing literature on intervention studies suggests that an L2 intervention may contain explicit and implicit elements, as Section 4.4 below will show. The ongoing question pertains to the impact of both types of instruction on L2 knowledge.

4.2.3 The impact of instruction on L2 acquisition

What we know about the impact of instruction cannot draw a particular conclusion about the impact of each instruction type on L2 knowledge. Rothman (2008, p. 99) argued that "only natural positive evidence leads to grammatical competence" and added that explicit instruction cannot improve underlying grammatical competence. In contrast, Norris and Ortega's (2000) meta-analysis of 49 instructed second language acquisition (ISLA) studies reported a positive effect on instruction depending on calculating the effect sizes of each selected study and aggregating the values into an average overall effect size, which was a large effect size (d = 0.96), suggesting a positive impact of instruction, with explicit instruction being more effective than implicit instruction. They acknowledged that their findings on the impact of explicit over implicit instruction should be considered cautiously. They highlighted that it was impacted by some limitations, including that selected studies varied in the quality of the data collection methods, the explicit teaching studies were more than implicit teaching studies, and the features of explicit and implicit instruction were inconsistently operationalised in the studies. Finally, the studies differ in many ways, including learners' proficiency and the length and intensity of instruction. Despite these limitations, this meta-analysis study's findings were suggestive as it encouraged researchers to investigate the impact of instruction empirically. Spada and Tomita (2010) investigated the effect of explicit and implicit instruction on acquiring simple and complex grammatical structures in English by conducting a meta-analysis of 41 studies. This study supports the previous findings on the positive effect of instruction, revealing that explicit instruction showed a larger effect size than implicit instruction and that instruction has a positive effect on both types of knowledge.

A recent meta-analysis study covering 35 years of research on the impact of instruction revealed a positive impact (Kang et al., 2019). This meta-analysis of 54 ISLA studies (ranging between 1980 and 2015) identified factors that can cause variation in the effect of instruction in SLA. While they did find a slight distinction between explicit and implicit instruction, they also found statistically significant effects for modes of outcome measures, learners' initial levels of

L2 proficiency, research settings and instructional intensity. Although these meta-analysis studies have the limitations mentioned above, they provide informative implications of where further research on the impact of instruction in L2 acquisition is needed, such as considering the learners' L2 proficiency, the instruction intensity, the need for longitudinal studies, as well as more studies on the impact of each instruction type and what knowledge type is impacted.

Ionin and Montul's (2023, p. 303) book synthesises the relative effectiveness of instruction in intervention studies in SLA with a focus on various simple and complex linguistic structures. They conclude that their book confirmed the findings of the previously mentioned meta-analysis, stating that "explicit grammatical information led to improvement, at least in the short term" and that explicit instruction led to a more significant improvement than implicit instruction. They add that no conclusions can be drawn about retaining the knowledge gained from explicit instruction over time or the contribution of explicit instruction to implicit knowledge.

Understanding the distinction between explicit and implicit knowledge, the interface positions, and the literature views about the impact of instruction on SLA is essential for understanding what instructional input and tasks are needed and how each type can affect implicit and explicit linguistic knowledge (Loewen & Sato, 2017). Therefore, the next section reviews the literature on intervention studies into the acquisition of genericity. Before that, I give a short note on different approaches to grammar instruction.

4.3 A Note on Grammar Teaching Approaches

The literature on grammar teaching and L2 acquisition reveals two contrasting views on teaching grammar in L2 acquisition; Krashen (1981) argues that teaching grammar has no impact on implicit knowledge. In contrast, Ellis (2006) argues that teaching grammar does impact L2 acquisition and highlights the importance of the choices made in grammar teaching about timing, focus and approach to grammar teaching. It has been stated above that the L2 acquisition's primary goal is to develop the L2 learner's implicit knowledge and that there are different positions on transforming explicit knowledge into implicit knowledge. Furthermore, the discussion on the role of explicit grammar teaching showed that explicit instruction could impact both kinds of knowledge and that explicit teaching might be valuable in terms of internalising explicit knowledge. Continuing from this, I now consider approaches that teach grammar with the intent of gaining implicit knowledge. This note focuses on the differentiation between traditional grammar teaching (focus on formS), focus on form and processing instruction approaches.

The main aim of teaching grammar in traditional focus-on-forms approaches is to teach grammar rules and forms in isolation from communication (Richard & Rogers, 2001). The primary instruction in this approach is L1-L2 translations (as in the grammar-translation method), drills, repetition, and memorisation (as in the audiolingual method). These methods focus on forms in isolation from communication.

In contrast, the focus on form approaches cover teaching methods that teach the form in the context of meaningful communication (Long, 1988, 1990). These approaches consider teaching grammatical forms by directing the learners' attention to the form in meaningful contexts, encouraging interaction and providing feedback when required. Moreover, it includes input flooding and input enhancement methods. In the input flooding method, learners are exposed to the target forms more frequently but more implicitly by increasing the frequency of the target form in the input. On the other hand, input enhancement is defined as "the process by which language input becomes salient to learners" (Sharwood Smith , 1991, p.118), and it involves drawing learners' attention explicitly to the target forms, making them more noticeable by underlining or colouring the target forms. Ellis (2001) differentiates between planned and incidental focus on form. According to him, planned focus on form includes using focused tasks and repetitive attention to a specific linguistic form during communicative activities.

Processing Instruction (PI) is a teaching method that emphasizes learners' ability to make form-meaning connections when interpreting language input. Developed by VanPatten (2002), this approach is grounded in the idea that grammar instruction should help learners process input meaningfully, rather than merely practicing forms. In this context, "teaching to process" means guiding learners to attend to grammatical forms in a way that allows them to derive correct meanings during comprehension. PI consists of three core components: (i) providing explicit grammatical explanations situated in communicative contexts, (ii) drawing learners' attention to processing problems caused by cross-linguistic differences between the first and second language, and (iii) engaging learners in practice activities that reinforce the link between grammatical forms and their meanings (VanPatten, 2020). Ionin & Montrul's (2023, p.307) synthesis found that processing instruction "to be at least as effective as traditional grammar teaching in production and often more effective for comprehension." Although this study is not designed to answer questions about the effectiveness of one method over the other, such understanding can bring logic to the formation of the instruction, as will be explained in Section 6.6. The following section reviews the literature on intervention studies that examined the acquisition of genericity.

4.4 Intervention studies on L2 genericity acquisition

As explored in 3.3, the literature has revealed that acquiring genericity in an L2 is a complex challenge for L2 learners. The literature on the acquisition of genericity has highlighted some factors that contribute to this challenge, including the learner’s L1 and the cross-linguistic differences between the learner’s L1 and L2, transfer and L2-related factors such as the complexity of form-meaning mapping and the low frequency of a form-meaning mapping in the input. Until now, few studies have attempted to use the theoretical linguistic analysis of articles as the starting point for conducting interventions that involve teaching L2 learners about genericity. A review of the main findings of these studies is provided in what follows.

4.4.1 A historical overview of intervention on articles acquisition

Early investigations tended to simplify English articles into definiteness and specificity. Master (1990) suggested a pedagogical order to teach articles to intermediate-level language learners. He based his design on the binary division between classification (*a/an* and *Ø*) and identification (*the)*. He suggested that article pedagogy should explain different notions to the language learner, including countability, first mention, subsequent mention, post-modification, proper nouns, and idiomatic phrases using several examples.

Master (1994, 1997) applied this suggestion in intervention studies that taught articles using these notions in nine writing sessions using explicit rule presentation. Six hours of teaching were devoted to article pedagogy in these sessions. The participants were learners in English as a second language classes. The study included an experimental group taught about the rules and a control group enrolled in the writing class without direct rule presentation. The study used fill-in-the-blank tasks in the pre-test and post-test. Both studies’ results suggested that systematic teaching helped the experimental group outperform the control group in the post-test. Although the sample included a mixture of various L1 speakers, no attention was given to the role of the learners' L1. Moreover, conducting these studies in the United States suggests consideration of the role of exposure outside the classroom, which is neglected. Sabir and Lopeze (2019) disagree with this simplification of articles as it ignores contexts in which definiteness and specificity mismatch.

In 2002, Master attempted to apply the previously discussed notions using instruction based on a systematic informational structure that focused on the logical sequence of presentation, direct explanation, practice, providing feedback and regular assessment that lasted for three hours during a 3-week intervention. The sample was divided into three groups: the structured instructional teaching group, the traditional teaching group, and a control group that received no teaching. The participants' performance in writing and oral tasks, as well as grammar exercises and reading comprehension, were analysed, and the results revealed that the subjects who were taught using the information structure system improved over both conventional teaching and the control group. Although Master's studies compared the impact of different teaching methods and indicated a positive effect of instruction on article use, the long-term effect of instruction and the impact of other factors, such as the L1 and exposure to input, were not accounted for. More importantly, the study did not consider what specific elements of article use showed improvement. Few studies attempt to build on a linguistic analysis of the articles in L1 and L2, as the next section shows.

4.4.2 Theoretically informed intervention studies on the acquisition of genericity

Applying the framework of Krifka et al. (1995), Snape and Yusa (2013) initiated an intervention investigation on the acquisition and perceptions of articles considering the findings of GenSLA. Their pilot study aimed to investigate the impact of instruction on improving article choice and perception. The participants included 14 university students who were high-intermediate Japanese learners of English. The participants were divided into experimental and control groups. This pilot study used a pre-test, intervention, post-test, and delayed post-test design. The participants completed three tasks for each testing session, including two versions of forced-choice elicitation and AJT tasks and a single version of a transcription task. The treatment lasted for 70 minutes once a week for three weeks; the experimental group was taught about definiteness and specificity distinctions in the first week, article perception in the second week, and genericity in the last week. The participants were tested immediately after the intervention (post-test1) with a delayed post-test two weeks later.

The pre-test results revealed that using definite articles in NP-level generics was the most challenging for Japanese learners of English. The post-test and the delayed post-test results revealed that the forced choice task showed no contrast between the experimental and control groups in both contexts. The AJT showed no difference between the groups except for mapping bare plurals to NP-level generics as the control group performed better in this condition. The transcription task showed no difference in perceiving the definite singular in the post-tests compared to the pre-tests for both groups. In addition, the experimental group showed better perception of the indefinite article in the post-test compared to the pretest. There was no difference between the groups in the perception of bare plurals, on which the experimental group's post-test performance was similar to that of the pre-test. Hence, Japanese learners of English did not greatly benefit from explicit instruction. Therefore, Snape and Yusa conclude by stating that the instruction in their study was ineffective in article acquisition, although it has some positive effect on perception.

The researchers account for the ineffectiveness of their teaching intervention using different factors, which is worth considering when conducting future research. First, article instruction is very complex if compared to the grammatical structures investigated in studies that found a positive impact on instruction. Second, the instruction content and presenting it in English may confuse the L2 learners. Third, the teaching time was short. This highlighted the need for a longer teaching time provided in the L2 learners' native language to clarify that confusion. Although this study provides valuable implications for teaching articles in terms of content and length, only the AJT was used to test genericity; hence, the study does not compare the Japanese learners' improvement using different measures that tapped into different types of knowledge about genericity. Moreover, the study could benefit from considering a native control group to validate the tasks and set the baseline.

Following up on Snape and Yusa's (2013) pilot study, Umeda et al. (2019) tested the explanation provided by Snape and Yusa (2013) for the ineffectiveness of their teaching intervention. They conducted a longer intervention study with the teaching done in the L2 learner's native language, Japanese, to test the impact of explicit instruction that relies on linguistic analysis of article semantics, including NP-level and sentence-level genericity, as well as testing whether any explicit knowledge is retained for a longer time. The sample included an instruction group (*n* = 21), a control group (*n* = 16) and native controls (*n* = 9). The proficiency level of the sample was high-intermediate to advanced. The experimental group received a 60-minute lesson once a week over nine weeks, and no instruction was provided to the control group. Both groups' performance levels were tested before instruction using a pre-test. The instruction included metalinguistic explanations about genericity for three weeks, then definiteness and specificity for four weeks and finally, a review for two weeks. Sessions included exercises (production and drills) in English and group and pair work creating short dialogues. During and after instruction, the authors administered four post-tests to both groups. The first post-test tested for generic acquisition after three weeks of intervention. The second tested article semantics in week 10. Two delayed post-tests were conducted in week 12 and a year later. All tests used an AJT.

Results from the pre-test showed that both groups were accurate in their ratings of bare plurals but not with definite singulars in NP-level generics. The same pattern of high ratings of bare plurals in the pre-test was found in sentence-level generics. The first post-test results showed that the instruction group's rating of the definite singular for NP-level generics improved between pre-test and post-test 1 and maintained the improvement in post-test two at week 10. The results of the delayed post-test 3 at week 12, showed that significant improvements were made in selecting definite singulars for NP-level generics and bare plurals for both generic meanings. The researchers suggested that the longer intervention and the provision of this instruction in the participants L1 contributed to this improvement. However, one year later, this improvement was not retained in the fourth post-test, as the sample reverted to their pre-test level rather than maintaining the post-test 1 level of accuracy. The researchers acknowledged that short-term improvement cannot address the issue of implicit knowledge development. They suggested that the move from explicit knowledge to implicit knowledge is unlikely to happen for this complex property over an extended period. Moreover, they highlighted that a linguistic description of generics might be more beneficial with advanced learners and that less evidence in the input, which overlaps inside and outside the class contributed to the difficulty. This study provided an innovative example for applying linguistic theory in teaching. It provided supportive implications for genericity instruction, highlighting the use of mapping features to forms in teaching. The researchers acknowledged that using various measures would give better findings about the participants' knowledge. However, genericity was only taught for over three weeks, which might not be enough to account for the complexity of form-meaning mappings. Moreover, we are not told whether negative evidence in the intervention input was used during the teaching or the practice. A more comprehensive study may benefit from including both types of evidence and splitting the teaching over a period longer than three weeks to allow more practice to support the acquisition of this complex property.

The previous studies considered teaching genericity to learners from an article-less L1 background. I turn now to present intervention studies on the acquisition of generics by learners whose L1 has articles. Sabir (2015) conducted an intervention study to explore the acquisition of the English article system by Saudi Arabic-speaking learners of English and to discover whether explicit instruction on article semantics (definiteness, specificity, and genericity) and translation activities that target article use can contribute to accurate article choice. The researcher predicted that L2 learners would be able to differentiate between the two types of generic meaning and show different ratings for the indefinite singular and definite singular in sentence-level and NP-level generics, respectively, while giving similar ratings for bare plurals in both contexts. Sabir further predicted that there would be an effect of L1 transfer and a positive impact of explicit instruction compared to implicit instruction, and translation activities compared to gap-filling activities.

The study followed a pre-post-test design with an intervention between the two tests and a one-month delayed post-test. The sample included 67 Saudi (Hejazi) Arabic-speaking learners of English and 23 native English speakers. For each testing session, the learners completed an elicited written production task, an AJT to test genericity and forced-choice tasks to test definiteness and specificity. The participants were divided into four experimental groups: the translation explicit, the translation implicit, the gap-fill explicit and the gap-fill implicit groups. The instruction lasted for three weeks, with two 60-minute sessions per week. Each group received a different type of instruction, as indicated by the group's name, to see which method was the best for teaching articles. The first three lessons presented teaching on definiteness and specificity, and the last three sessions focused on the two types of generics. The explicit translation group engaged in translation tasks focusing on article usage in generic and non-generic contexts. They received explicit instruction on article semantics, including definiteness, specificity, and genericity. The instruction was linguistically informed, using the terms specific, non-specific, generic, and non-generic to clarify these concepts. The implicit translation group was taught using the same translation activity with the lack of explicit teaching on articles and instead listening comprehension activity using real-life conversations. The other two groups received the same explicit and implicit teaching but used gap-filling activities instead of the translation ones. Because of the word count limits, only the results for genericity are reviewed here.

Concerning genericity, the pretest results revealed that all groups were target-like in accepting definite singulars and rejecting indefinite singulars in NP-level generics before the intervention. In addition, the two translation groups were native-like in accepting bare plurals, rejecting definite plurals and bare singulars in this meaning. In contrast, the two gap-filling groups faced difficulty accepting bare plurals and rejecting the ungrammatical L1 definite plurals and the L2 bare singulars in the NP-level generics pre-test. The post-test results showed that the intervention was not effective in reducing the high ratings of bare singulars. The intervention succeeded in increasing the rejection of definite plurals, although it was not supportive in increasing the ratings of bare plurals nor decreasing the ratings of bare singulars in NP-level generics. This improvement for the rejection of definite plurals was maintained after a month. In sentence-level generics, the pre-test results revealed that the explicit translation group was target-like, and the implicit translation group was target-like in all NPs except the bare singulars which they accepted. The two gap-filling groups faced difficulty in rating all NPs but not bare plurals, rating it as acceptable. The post-test results showed that the intervention did not change their ratings except for rejecting the definite plurals. However, this improvement was not maintained a month later. The results did not reflect a clear impact of instruction on article accuracy. This study is limited in a variety of ways. First, the researchers acknowledge the lack of a control group that received no intervention. Second, the production data was only analysed in terms of error rate, which can give a conclusion about the total accuracy, but does not allow us to draw potential conclusions about the nature of the errors and their possible sources.

With a focus on investigating the role of input in the acquisition of characterising generic meaning, Abumelha (2019) examined the acquisition of English articles to anticipate acquisition difficulties and investigate the role of two input types, namely, implicit and explicit, in accelerating generic feature reassembly in the L2. To do so, 45 Najdi Arabic–speaking learners of English were divided into two experimental groups: the explicit teaching group and the implicit teaching group and a third uninstructed control group. The researcher then administered a pre-test followed by 12 hours of instruction, an immediate post-test, and an eight-week–delayed post-test. The tests included a forced elicitation task and a sentence repetition task. The intervention included explicit grammar teaching and text presentation, and the implicit group intervention included reinforced input in genre analysis. The pre-test results showed that L2 learners faced difficulty in indefinite generic contexts in the forced elicitation task. The post-test revealed a significant improvement in two categories: [+generic, –plural] and [+generic, +plural] by the explicit instruction group in this task. However, only the [+generic, +plural] results maintained the improvement on the delayed post-test. In the repetition task pre-test, the explicit input teaching group performed better on both post-tests than the implicit input group. The results of the repetition task showed that the explicit input positively affected the [+generic] [+plural] immediate and delayed post-tests. Overall, the results revealed that explicit input has a positive effect on accelerating the feature-reassembly process and suggested the need for further investigation to refine the type and amount of input.

A strength of this study is its use of native controls to validate the tasks and set the baseline of comparison. It contributes to the literature on explicit and implicit input comparison, yet its focus is limited to characterising generic sentences only. Moreover, the long-term effect of instruction on the acquisition of bare plural relied on a two-month delayed post-test, which might disappear over a more extended period, as found in Umeda et al. (2019). Further research is needed to account for the acquisition difficulty and impact of instruction on the acquisition of kind generics.

The studies reviewed here have investigated instruction from different views; some studies investigated the role of explicit instruction, such as Snape and Yusa (2013), or the role of linguistically informed explicit instruction (Umeda et al., 2019), whilst others compared explicit vs. implicit instruction (Sabir, 2015) and explicit vs. implicit input (Abumelha, 2019). Overall, the research findings reported here consistently point toward the difficulty of acquiring generic form-meaning mappings as L2 learners from L1 backgrounds both with articles and without articles faced difficulty acquiring generics. The evidence reported here seems to suggest no clear conclusions about the role of instruction in facilitating the form-meaning mappings in L2 generics. Furthermore, the results on the short-term effect of explicit instruction are not consistent; some studies reported no impact (Snape & Yusa, 2013; Sabir, 2015), while others reported the positive short-term effect of explicit instruction (Umeda et al., 2019) and a longer effect on plural generic contexts (Abumelha, 2019). This review highlights a paucity in the literature of linguistically informed intervention studies on the acquisition of genericity. It points out the need for adhering to the call for GenSLA-informed interventions that consider different elements of instruction, including the type of evidence, improvement in implicit knowledge, the role of the L1 and input in intervention design to bridge the gap between theory and practice in SLA. In addition, it revealed many methodological and measurement challenges that require thoughtful consideration in future research. The literature reported in this chapter was used to inform the decisions in the design of this study, as elaborated in the following section.

4.5 Interim summary and operationalisation of instruction in this study

This chapter reviewed the literature on instruction in relation to L2 acquisition by discussing the following concepts: instruction; the evidence in input; the different types of linguistic knowledge, including explicit and implicit; the positions on the interface between the two types; and the measures of both explicit and implicit knowledge. This chapter concludes by summarising the empirical evidence found in the literature of linguistically-informed intervention research to set the gap that this study is contributing to fill. What follows is a detailed summary that shows the implications of the literature for this study.

Taken together, instruction in this study refers to teaching functional grammar and its semantic features and the form-meaning mapping rules using a GenSLA linguistically-informed intervention. The term intervention refers to designing input that accounts for findings on what is easy and challenging to acquire in SLA research, considers the role of L1 in this acquisition, and then presents this input in grammar teaching to facilitate overcoming difficulties in L2.

In Section 4.2.1, it was suggested that the input of the L2 classroom is different from the input received by a child acquiring their L1. L2 learners receive information about the ungrammatical forms in the form of input or feedback, negative evidence, in addition to positive evidence (VanPatten, 2009). The literature reveals different positions regarding the role of negative evidence in L2 acquisition; some argue that it is not essential for L2 acquisition

(Schwartz, 1993), while others suggest that it is beneficial (White,1991; Hirakawa et al.,2019). Following the lines of Hirakawa et al.'s research, I take the position that instruction that clarifies the proper and ungrammatical form-meaning mappings regarding genericity may facilitate learning (detailed discussion on the input used in instruction is provided in Chapter 6).

In Section 4.2.2, the differences between explicit and implicit knowledge were discussed. The debate on the interface between the two types of knowledge was discussed. Scholars who support the no interface position claim that the move from explicit to implicit knowledge is impossible (Krashen, 1981; Schwartz, 1993). In contrast, advocates of the strong interface position claim that explicit knowledge can become implicit and vice versa (Sharwood Smith, 1981; DeKeyser, 2003, 2017). A third position advocates for a weak interface position which sees the move from explicit to implicit knowledge as difficult but possible under suitable circumstances such as the learners' developmental readiness and the existence of an explicit input with good quality and practice (Ellis, 2005; N.C. Ellis, 1994). This section discusses the suitable measurements of each type. The present study has no direct agenda for testing which knowledge type improved after the intervention, as the necessary online measures are not available in the classroom. However, this understanding informs the tasks used to attempt tapping into both knowledge types. Considering the efforts provided in defining the measures of each knowledge type, relying on Ellis' (2005) criteria, this study selected three types of tasks that are supposed to tap into both explicit and implicit knowledge: an elicited written production task, an acceptability judgment task, and a forced-choice task.

After differentiating between explicit and implicit learning and instruction in Section 4.2.2.2, Section 4.2.3 presents the literature on the impact of instruction. The reviewed studies conclude that the impact of instruction is effective in the short term, with explicit instruction as more effective than implicit (Norris & Ortega, 2000; Spada & Tomita, 2010; Kang et al., 2019). The reported impact of explicit instruction informed the decision to teach form-meaning mapping explicitly in the present study, with a focus on the inclusion of negative evidence and increasing practice with delayed testing for the long-term effect of the instruction, if any, to contribute to the literature on the impact of instruction in SLA.

In Section 0, the literature reveals a paucity of linguistically informed intervention studies that are designed to account for acquisition facilitation rather than testing a specific pedagogical variable. Only two studies (Sabir, 2015; Abumelha, 2019) considered Arabic-speaking learners, and the findings of both studies on the impact of instruction on L2 generic form-meaning mappings are contradictory. This justifies the need for the present study, whose findings could contribute to the knowledge of the impact of teaching on L2 acquisition and implications for the intervention studies methodology and classroom input. In addition, the empirical studies reported in this literature review showed that the focus on genericity did not exceed 3-week sessions, and all the studies except Abumelha’s (2019) used a single measure for the acquisition of genericity. This informed the design of this intervention to split teaching on genericity over a longer time, including 16 teaching sessions over eight weeks, to present the complexity of genericity in a way that allows for better processing and more practice on the learners' side.

Chapter 5 Generic Form-Meaning Mapping in English and MSA: An Empirical Account

5.1 Introduction

This Chapter presents an empirical study that compares how native speakers of Arabic and native speakers of English express genericity. Cross-linguistic differences and the impact of such differences are central to the understanding of the acquisition of generic form-meaning mappings. As shown in 2.4, several theoretical attempts have been made to discuss how generic form-meaning mappings differ cross-linguistically (e.g., Chierchia, 1998; Dayal, 2004). Some researchers validate these theoretical accounts by conducting empirical investigations of the native speakers' use of generic form-meaning mapping in different languages, such as Ionin et al. (2011a) who compare English, Spanish and Brazilian Portuguese. Such experimental investigations account for actual use and variability among native speakers and contribute to the robustness of the predictions on acquisition difficulty. Considering MSA, Fassi Fehri (2012) enriches the theoretical discussion of nominals' interpretations, including genericity. However, investigations into how native speakers of Arabic express genericity do not exist, yet. Therefore, this thesis broadens the empirical base to include Arabic native speakers' generic form-meaning mappings compared to English to validate theoretical predictions and explain the learning task investigated in this thesis on robust grounds, inspired by Ionin et al. (2011a). This chapter begins by reviewing Ionin et al. (2011a). Then it outlines the methods, procedures and results of an experimental study. It concludes by defining the learning task investigated in this thesis.

5.2 A note on Ionin et al. (2011a)

Ionin et al.'s (2011a) study was the first step in the experimental examination of the expression of genericity cross-linguistically, in which they conducted a study to examine nominals in various generic contexts in different languages. The main aim was to test the availability of generic and kind meanings in English, Spanish, and Brazilian Portuguese (BrP), and the distribution of genericity among different NP types in these languages. In particular, they tested Dayal's (2004) proposal about the distribution of singular and plural generics. The main predictions included the obligatoriness of definiteness marking with singular generics, its optionality with plural generics, and the application of the well-defined kind restriction to definite singular generics. Using a 40-item AJT with contexts in English, with Spanish and BrP copies, they tested native speakers of English (n = 22) in Illinois, native speakers of Spanish (n = 16), and BrP native speakers (n = 19) in Brazil. The participants were college-educated.

They found that the English-speaking group rated definite singular and bare plural NPs higher with kind generics, as predicted, and that singular kind NPs must be definite in all three languages. The Spanish-speaking group accepted definite singulars and plurals, as predicted. The BrP group's ratings showed a significantly high rating for definite singulars, definite plurals and bare plurals, as predicted by Dayal's (2004) account. Considering characterising generics, the groups significantly rated the predicted nominals at sentence level higher: the indefinite singular and bare plural in English, indefinite singulars and definite plurals in Spanish and all nominal types except the definite singular in BrP. The results provide support for the following predictions of Dayal's account. This study highlights the importance of experimentally testing predictions given by semantic theories. It informs the investigation of how Arabic native speakers express genericity, setting the L1 baseline and validating theoretical predictions before discussing the learning task.

5.3 Hypotheses

This investigation aimed to test the similarities and differences highlighted in Section 2.3, and Section 2.5. To maintain the narrative, they are repeated here in the form of hypotheses:

H1: Dayal's (2004) first prediction argues that definiteness marking is obligatory for singular definite kind NPs in languages with definite articles. English and MSA have definite articles and are expected to have similarities in mapping the definite singulars onto kind meaning. Both groups are predicted to give similarly high ratings for the definite singular and low ratings for bare singulars in [+kind, −plural, +definite] condition.

H2: Dayal's (2004) second prediction argues that English lexicalises the iota operator by the definite article but not the down operator. The definite article blocks definite plurals from having the kind reading. Bare plurals have the kind reading. In contrast, MSA lexicalises both iota and down operators. The blocking principle applies to both, allowing definite plurals to have definite and kind readings and blocking both readings from being assigned to bare plurals. Therefore, it is predicted that the groups' ratings of definite and bare plurals will differ; English native speakers are expected to rate bare plurals higher and MSA native speakers are expected to rate definite plurals higher in both kind and characterising generics.

H3: In light of the first prediction, the blocking principle applies for the definite article, therefore, it is predicted that MSA does not allow bare singulars, which are indefinite in [−plural, −definite] while English allows indefinite singulars in characterising generics. MSA native speakers are expected to rate indefinite singulars lower than definite singulars, and English native speakers are expected to rate indefinite singulars higher than all other singular forms.

H4: If Dayal's account is accurate, predicted forms are expected to be rated higher than unpredicted ones by native speakers in each language.

What follows discusses the methodology used in testing these predictions.

5.4 Methodology

5.4.1 Study Sample

Native speakers of English and MSA were recruited for this study. In particular, the sample included twenty native speakers of English who were invited to participate in the study in the United Kingdom. English native-speaker participants' mean age was 34, ranging from 19 to 67 (*SD* = 13). In addition, twenty native speakers of MSA were invited to participate in Saudi Arabia. These participants' mean age was 36, ranging from 24 to 49 (*SD*= 5.7). All participants were tested in their native language. All the participants were college-educated, and some were college students at the time of data collection. Both groups were selected using convenience sampling based on local contacts in the UK and Saudi Arabia. A major drawback of this sampling method is that it includes a possible bias and, therefore, the results are unsuitable for generalisation (Gozlar et al., 2022). This empirical investigation aims to provide evidence for generic form-meaning mappings in both languages to support this thesis without any generalisation intentions.

5.4.2 Design, tasks, and procedures

This study is a cross-sectional study to test the predictions regarding how English and MSA express genericity by comparing native speakers' judgements of how five NP types (definite singular NP, definite plural NP, bare plural NP, bare singular NP, and indefinite singular NP) were accepted in two types of generic; kind denoting NPs and characterising generic sentences. AJTs are one of the most common methods used in the linguistic literature to measure intuitions on grammaticality. A major advantage of AJTs with a Likert scale is its reliability and robustness with small sample sizes. Langsford et al. (2018) compared different linguistic measures and found that the AJT with a Likert scale proved to be the most stable and reliable measure; it was robust even with a small sample size. Therefore, this study used a Likert scale AJT with contexts as the main data collection tool, adapted from Ionin et al. (2011a). This study used two copies of the AJT, an English copy and an MSA copy. To control for translation bias, the Arabic copy of the AJT was verified by two English-Arabic translators. The two AJTs were identical in content, order, format, and number of test items.

Each AJT included twenty items: six items of characterising generics, six items of kind generics, and eight distractors that focus on tense and aspectual interpretation. Of these twenty items, four items from each generic category and eight distractors (n = 16) were adopted from Ionin et al. (2011.a). Because Ionin et al. 's (2011a) AJT included only four items in each generic category, the researcher added two extra test items to increase the reliability and validity of the results. The participants were asked to rate their acceptability on a scale from 1 to 4 (where 4 is "highly accepted") for each of the five sentences (each sentence refers to a condition) in light of the context provided for each test item. Only one rating for each sentence is accepted. Examples of each category are (A) for kind-denoting generics and (B) for characterising generics. Expected values are in bold. Full MSA and English copies of the AJT are available in Appendix A.

(A) characterising generics (n = 6)

(English)

Food experts report that root vegetables are perfect for your health. It has been said that.

a) The carrot is a good source of vitamin K1 and antioxidants	1	2	3	**4**
b) Carrots are good sources of vitamin K1 and antioxidants	1	2	3	**4**
c) A carrot is a good source of vitamin K1 and antioxidants.	1	2	3	**4**
c) The carrots are good sources of vitamin K1 and antioxidants	**1**	2	3	4
d) Carrot is a good source of vitamin K1 and antioxidants	**1**	2	3	4

(Arabic)

قدم خبراء ا لتغذية تقارير بأن الخضروات الجذرية مثالية للصحة, لقد قيل أن...

أ. الجزرة مصدر جيد لفيتامين ك ومضادات الأكسدة.	1	2	3	**4**
ب. جزر مصدر جيد لفيتامين ك ومضادات الأكسدة.	**1**	2	3	4
ج. جزرة مصدر جيد لفيتامين ك ومضادات الأكسدة.	**1**	2	3	4
د. الجزر مصدر جيد لفيتامين ك ومضادات الأكسدة.	1	2	3	**4**
ه. جزر مصدر جيد لفيتامين ك ومضادات الأكسدة.	**1**	2	3	4

(B) kind generics (n = 6)

(English) The Netherlands is a great country to visit. It has wonderful museums, great food, and excellent public transportation. And, of course, it's a great place to buy flowers. As you probably know…

a)	Tulips are very popular in the Netherlands	1	2	3	**4**
b)	The tulip is very popular in the Netherlands	1	2	3	**4**
c)	Tulip is very popular in the Netherlands	**1**	2	3	4
d)	A tulip is very popular in the Netherlands	**1**	2	3	4
e)	The tulips are very popular in the Netherlands	**1**	2	3	4

(Arabic)

هولندا بلدٌ رائع يتوجب زيارته, ففيها متاحف رائعة و طعام لذيذ ووسائل نقل عام ممتازة, وهي تمتاز بالزهور فكما تعرف...

4	3	2	**1**	أ. زهور توليب تحظى بشعبية كبيرة في هولندا
4	3	2	1	ب. زهرة التوليب تحظى بشعبية كبيرة في هولندا
4	3	2	**1**	ج. زهرة توليب يحظى بشعبية كبيرة في هولندا
4	3	2	**1**	د. زهرة توليب يحظى بشعبية كبيرة في هولندا
4	3	2	1	ه. زهور التوليب تحظى بشعبية كبيرة في هولندا

Considering the study procedures, both copies were uploaded to Microsoft Forms, and a link was created for each copy of the AJT. The link for the English task was sent via email once the participant was willing to participate following a face-to-face discussion in the United Kingdom (ERGO 71758, attached to Appendix E). For the MSA-speaking group, the AJT link was sent via email to participants who expressed their willingness to participate in response to the researcher's request, which was distributed via WhatsApp to the researcher's colleagues and relatives. An electronic copy of the AJT was used because it has been found in the literature that the test format does not have a critical effect on the results; there is no difference between paper-based and computer-based AJT (Ionin & Montrul, 2009).

5.4.3 Data analysis

In terms of coding, the dataset comprises the ratings, the dependent variable, and conditions of generic form-meaning mapping, which is the predictor variable. The dependent variable, ratings, is an ordinal variable with four levels. The predictor variable, conditions of generic interpretation, was a five-level categorical variable. These levels describe the various form–meaning mapping conditions and include the five NP types. The purpose of using these different form–meaning mapping conditions is to examine how each NP type affects the acceptability of the test items. The data was analysed using R (R Core Team, 2023). Cumulative Link Mixed-effects Models (CLMMs) were appropriate for this type of ordinal rating data (Christensen, 2023). Therefore, the AJT data was analysed using the *clmm* function from the *ordinal* package (Christensen, 2023. Version: 2023.12–4.1) and the package *emmeans* (Lenth, 2023. Version: 1.8.7) for Tukey-corrected pairwise comparisons in. The model used the maximum possible structure for each model and was compared to a null model using the *lrtest* function to assess the goodness of fit. Since comparing the groups directly to one another could be confounded by the fact that both groups completed different tests, making it difficult to draw conclusions on how languages differ, each group's data was modelled separately. Whilst one AJT is a direct translation of the other and identical in organisation, length, and number of test items, there are some differences in lexical choices and agreements. Two within-group differences CLMMs were modelled. A p value $< .05$ was considered significant in this analysis. The following section presents the outcomes of both AJTs.

5.5 Results

This section presents the results of the native speakers' investigation. In terms of description, Figure 5.1 provides an overview of the density and means of each group's ratings in characterising and kind generic meanings.

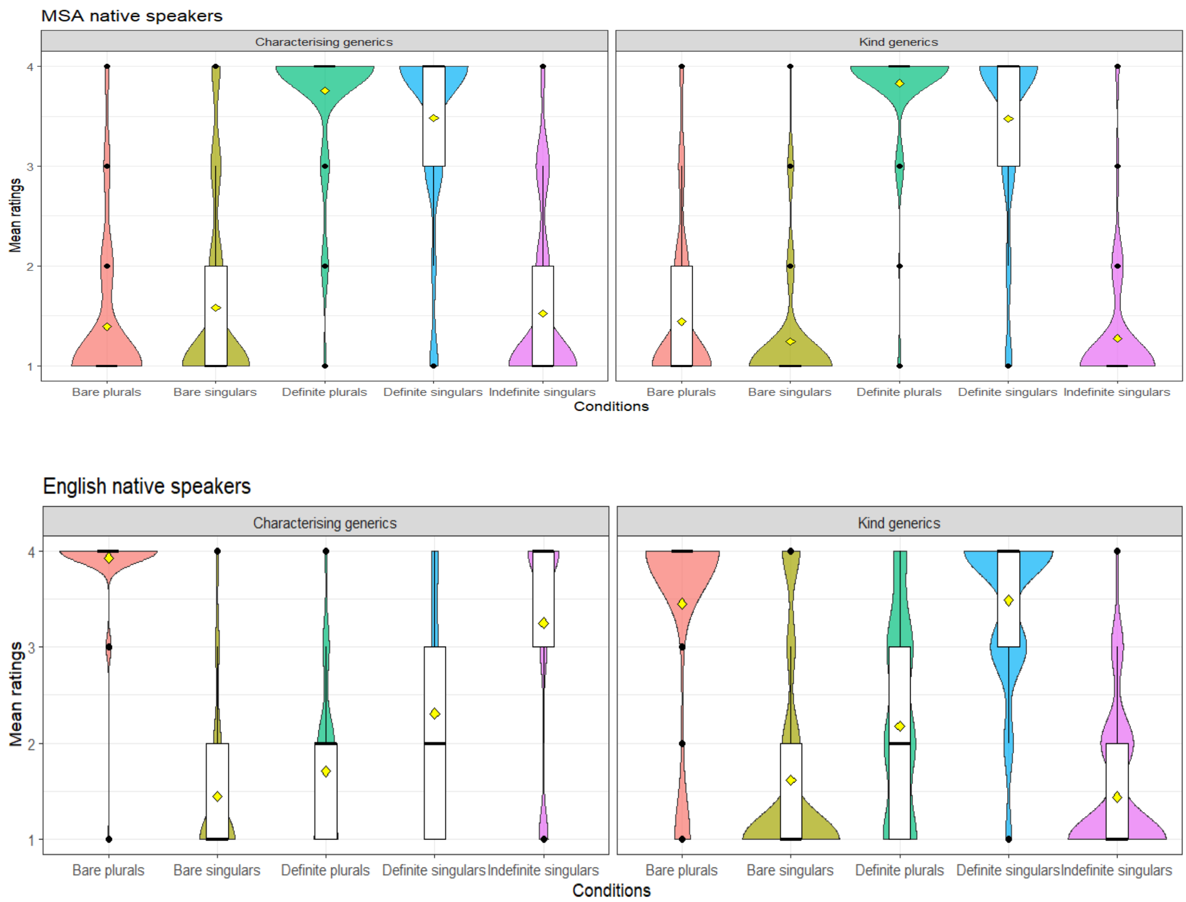


Figure 5.1 Violin plots showing raw rating densities, mean ratings, and standard deviation in AJT for the English native speakers and the MSA native speakers. The yellow diamond indicates the mean rating for each condition.

Figure 5.1 compares the group ratings for each NP in characterising and kind generics. The top half shows the MSA group's ratings for both meanings. The figure shows that the MSA group rated definite plurals and singulars higher than other NPs in both meanings. The bottom half of Figure 5.1 shows the ratings of the English native speakers. It shows that the English native speakers rated bare plurals and indefinite singulars higher than all other NPs in characterising generics. For kind generic meaning, English native speakers rated bare plurals and definite singulars higher than all other NPs.

Comparing the groups' ratings shows that, in characterising generics, the groups' ratings differ in almost all conditions except in rating bare singulars, as indicated by the mean ratings and densities. For all conditions in characterising generics, while the English native groups' mean ratings of bare plurals and indefinite singulars were higher than three, MSA native speakers rated definite plurals and definite singulars higher. With kind generics, the English native groups' mean ratings of bare plurals were higher than three while MSA native speakers rated bare plurals lower. Both groups rated definite singulars highly and gave low ratings for bare and indefinite singulars. In sum, the descriptive statistics show that native speakers agreed on accepting definite singulars with kind predicates, and they differ in their ratings of definite and bare plurals across both meanings. It is necessary to establish whether these differences are statistically significant.

To test the difference in ratings for each language, two CLMMs were constructed to test each group's differences separately. Starting with the English native speakers' data, a complete model was implemented using the ratings as the dependent variable, with the generic meaning and conditions (NPs) and their interaction as the predictor variable. Bare plurals and characterising generic meaning were the reference levels. A likelihood ratio test of the null model and the model with the fixed effects was significant (LR (9) = 785.4, $p < .001$); hence, the model with the fixed effect provided a significantly better fit for the data. The model results showed that the English native speakers' mean rating of bare plurals in characterising generics was significantly higher than rating all other conditions in both meanings ($p < .001$, for all comparisons as the model in Appendix A, A.3 shows). Tukey-corrected post hoc pairwise comparisons showed that in characterising generic meaning, the native speakers rated indefinite singulars significantly higher than all other unacceptable NP, $p < .001$ in all comparisons, as shown in Table 5.1.

Table 5-1 Tukey- corrected pairwise comparisons for characterising generic meaning.

Contrast	**Estimate**	**St. Error**	**Z ratio**	**P value**
Indefinite singulars to bare singulars	3.70	0.30	12.16	< .001*
Indefinite singulars to definite singulars	1.89	0.27	7.10	<.001*
Indefinite singulars to definite plurals	2.11	0.33	6.43	< .001*

Note: * significant at .05 level.

The results of the pairwise comparisons showed that the mean rating of definite singulars in characterising generics was higher than the mean ratings of bare singulars (*ES* = 1.81, *SE* = 0.27, z = 6.75, p <.001) and the definite plurals mean rating (*ES* = 0.96, *SE* = 0.27, z = 4.09, p = .002). The pairwise comparisons between bare singulars and definite plurals were not statistically significant (*ES*= −0.85, *SE* =0.27, z = −3.16, p <.06). These results showed that English native speakers rated bare plurals and indefinite singulars higher than all other NPs with characterising generic meaning. Moreover, their mean rating of the definite singular was higher than their rating of the unacceptable bare singulars and definite plurals.

Turning to the kind generic meaning, Table 5.2 summarises the pairwise comparisons of the English native speakers' ratings.

Table 5-2 kind generic meaning Tukey-corrected pairwise comparisons for the English native speakers' data

Contrast	**Estimate**	**St.Error**	**Z ratio**	**P value**
Bare plurals to definite singulars	0.52	0.31	1.70	0.79
Bare plurals to definite plurals	2.88	0.30	9.66	<.001*
Bare plurals to indefinite singulars	4.31	0.33	13.16	<.001*
Bare plurals to bare singulars	4.14	0.33	12.55	<.001*
Definite singulars to bare singulars	3.62	0.30	12.39	<.001*
Definite singulars to indefinite singulars	3.79	0.29	13.07	<.001*
Definite singulars to definite plurals	2.36	0.26	9.19	<.001*
Bare singulars to indefinite singulars	0.17	0.28	0.59	0.99
Bare singulars to definite plurals	1.26	0.26	4.82	.001*
definite plurals to Indefinite singulars	1.42	0.26	5.57	<.001*

Note: * significant at 0.05 level of significance.

Table 5.2 summarises the pairwise comparisons for the ratings of each NP type in kind generics by the English native speakers. It shows that this group's ratings of bare plurals and definite singulars were not statistically different, and these ratings were higher than the ratings of all other conditions. Moreover, this group's mean rating of definite plurals was lower than the midscale but significantly higher than bare and indefinite singulars. In short, English native speakers rated bare plurals and definite singulars significantly higher than all other conditions in kind generics.

Turning to the MSA group's data, the CLMM was constructed following the same procedures described above for the English native speaker group's model. The likelihood ratio test showed that the model with fixed effect provides significant better fit for the data (*LR* (9) = 105, $p < .001$) than a null model. The model results reveals that the reference level bare plurals in characterising generics was significantly lower than the mean ratings of definite plurals and definite singulars in both characterising and kind generics ($p < .001$ for the four comparisons).

The reference level rating was not significantly different from all other NPs in both meanings (see the model in Appendix A, A.3). This model was followed by Tukey-corrected post hoc pairwise comparisons to account for rating each NP in each generic meaning.

Considering characterising generics, it was predicted that MSA speakers would rate definite plurals and definite singulars higher than all other NPs. Table 5.3 summarises the pairwise comparison results.

Table 5-3 Characterising generic meaning Tukey-corrected pairwise comparisons for the MSA native speakers' data

Contrast	**Estimate**	**St.Error**	**Z ratio**	***P* value**
Definite plurals to definite singulars	1.03	0.34	3.07	.07
Definite plurals to indefinite singulars	5.07	0.36	13.96	< .001*
Definite plurals to bare singulars	4.95	0.36	13.705	< .001*
Definite singulars to bare singulars	3.92	0.31	12.64	< .001*
Definite singulars to indefinite singulars	4.04	0.31	12.93	< .001*
Bare singulars to indefinite singulars	0.12	0.28	0.42	.99

Note: * significant at 0.05 level of significance.

Table 5.3 summarises the post-hoc pairwise comparisons between the MSA native speakers' ratings for NPs in characterising generics. What stands out in this table is that this group rated definite singulars and definite plurals similarly as the comparison was not significant, as shown in the table. They rated these two conditions significantly higher than all other conditions as indicated by $p < .001$ for all comparisons. No significant difference was shown between the ratings of bare and indefinite singulars, as shown by the table above. This indicates that MSA native speakers rated the definite singulars and plurals similarly higher than all other NPs in characterising generic meaning.

Considering the MSA native speakers' ratings for each NP in kind generics, Table 5.4 presents the results of the post-hoc pairwise comparisons for the NP interpretations in this meaning.

Table 5-4 kind generic meaning Tukey-corrected pairwise comparisons for the MSA native speakers' data

Contrast	**Estimate**	**St.Error**	**Z ratio**	**P value**
Definite plurals to bare plurals	5.62	0.40	14.03	< .001*
Definite plurals to bare singulars	6.48	0.44	14.80	< .001*
Definite plurals to Indefinite singulars	6.28	0.43	14.77	< .001*
Definite singulars to bare plurals	4.19	0.32	13.05	< .001*
Definite singulars to indefinite singulars	4.85	0.35	13.83	< .001*
Definite singulars to bare singulars	5.06	0.37	13.81	< .001*
Definite singulars to definite plurals	1.43	0.37	3.87	0.04*
Bare plurals to bare singulars	0.87	0.34	2.55	0.24
Bare singulars to indefinite singulars	-0.2	0.362	-0.57	0.99
Bare plurals to indefinite singulars	0.6	0.32	2.04	0.57

Note: * significant at 0.05 level of significance.

The post hoc pairwise comparison showed that the ratings of the definite plurals in kind reference were significantly different from the ratings of all other NPs. These results indicate that the Arabic native speakers rated definite plurals significantly higher than all non-target NPs, and higher than the acceptable definite singulars. At the same time, their ratings of definite singulars were significantly higher than their ratings for bare plurals, indefinite singulars and bare singulars. All other comparisons of unacceptable NPs were not significant, suggesting that this group rated bare plurals, indefinite singulars, and bare singulars similarly low in kind generic meaning. In short, MSA native speakers rated definite plurals and definite singulars higher than all other NPs in kind generics. Table 5.5 summarises the differences between the languages based on the results.

Table 5-5 Summary of the accepted forms by both groups

Property	MSA NP types and feature bundle	English NP types and feature bundle
Kind generics singular forms	Definite singulars [+kind, +definite, −plural]	Definite singulars [+kind, +definite, −plural]
Kind generics plural forms	Definite plurals [+kind, +definite, +plural]	Bare plurals [+kind, −definite, + plural]
Characterising generics singular forms	Definite singulars [+generic, +definite, −plural]	Indefinite singulars [+generic, −definite, −plural]
Characterising generics plural forms	Definite plurals [+generic, +definite, +plural]	Bare plurals [+generic, −definite, + plural]

5.6 Discussion

The main goal of this investigation was to empirically investigate cross-linguistic differences between English and MSA in expressing genericity by testing native speakers' evaluations in both languages. In this section, I discuss the results in light of the study hypotheses and the different accounts for NP generic interpretations in the semantic literature. The results showed that the English native speakers rated bare plurals and indefinite singulars high with characterising generics and they gave low ratings for all other NPs. With kind generic meaning, they rated bare plurals and definite singulars significantly higher than other NPs. This suggests that English native speakers' results are in line with hypothesis 4. Turning to the results of the native MSA speakers, they showed high ratings for definite plurals and singulars compared to other NPs in both meanings. This suggests that MSA native speakers accepted definite singulars and plurals similarly in both meanings. These results were in line with the hypotheses.

These results suggest that English and MSA differ in mapping kind and characterising generic meanings to plural forms. While English maps this form to bare plurals, MSA maps these meanings to definite plurals. The high ratings of bare plurals by the English speakers is consistent with what Ionin et al. (2011a) found in their investigation of English native speakers' evaluation of NPs in generic meanings. In ionin et al. (2011a), the English-speaking group rated bare plurals highly across both categories. These results are compatible with Dayal's (2004) proposal and support the claim that plural kinds are derived by the down operator. Dayal (2004) argues that languages that lexicalised iota and down operators map both kind and definite meanings to definite plurals and block them for bare plurals. MSA lexicalises ι and ∩, using the definite article ‘al’; and the results provide empirical evidence that bare plurals lack both readings in MSA. Hence, the results could provide empirical support for Dayal’s (2004) proposal. Therefore, Dayal's account could provide an explanation of the difference between English and the MSA in expressing genericity using bare and definite plurals. One possible implication of this result for L2 acquisition is that MSA speakers learning English as an L2 may overgeneralise the use of definite plurals in English generic meanings due to L1 transfer.

An interesting result is reported in the English-speaking group’s ratings of definite plurals in kind generics. The results showed that although the group’s ratings of definite plurals were significantly lower than accepted bare plurals and definite singulars in kind generics, it was significantly higher than other unacceptable singular forms, with a wider density of ratings, compared to other unacceptable NPs. The English-speaking group's acceptance of the definite plural in kind generics could have several explanations, including education level and exposure to complex properties (Slabakova, 2013), personal preferences, or irresponsible performance in the AJT. Another explanation could be provided in light of Lyon’s (1999) observation of the flexibility in using definite plurals in generic expressions, particularly when referring to groups larger than a single species. Lyons (1999, p.182) states that definite plurals can be appropriately employed in generic expressions when referring to certain “names of animals and plants representing groups larger than the species”, such as in “The dinosaurs dominated the earth for a very long time”. Ionin et al. (2011b, p. 251) explained the English-speaking group ratings of the definite plural by the fact that definite plurals can express genericity if it includes a collection of taxonomic entities as in "[t]he lions [=all subspecies of lions] are dangerous." However, since this group's acceptance was not significantly high, investigating this case is beyond the scope of this study. However, it reflects that the L2 input received by the L2 learners may be confusing because of accepting definite plurals in some cases, especially if such cases are not highlighted through instruction when available.

Turning to the mapping of singular forms to generic meanings in English and MSA, in this investigation it was predicted that English and MSA would have similarities in mapping definite singulars onto kind generics. Both groups are predicted to give similarly high ratings for the definite singular and low ratings for bare singulars in the [+kind, −plural, +definite] condition. The results of both groups showed high ratings for definite singulars in this meaning. This suggests that MSA and English are similar in mapping definite singulars to this meaning, supporting the investigation prediction (H1). It also supports Dayal’s views on the obligatory nature of definite articles in both languages. An important result to discuss is that the Arabic-speaking group accepted definite singulars in characterising generics in the same way they accepted them in kind generics, which was predicted (H3). Moreover, the English-speaking group showed a significantly higher rating for definite singulars, compared to bare singulars in characterising generics, reflecting variability in the native speakers' data. While the Arabic-speaking ratings can be accounted for by the blocking principle which applies to the definite singulars, English native speakers' ratings can be explained by the lexical items in the task. Some items used a well-defined kind in the subject position of a characterising generics (the carrot). English restricts using definite singulars in generic sentences to only a well-defined kind. Due to the similarity between the two languages in mapping the definite singular to kind generics, it is possible that acquiring the definite singular mapping onto kind generics is easier for Arabic-speaking learners of English, as the similarity between the languages decreases the reassembly required in L2 learning (Lardiere, 2009). However, Arabic-speaking learners of English may need to learn the well-defined kind, [+taxonomic] feature, to avoid overgeneralising the L1 forms in L2 characterising generics.

As for mapping singular forms onto characterising generics, it was predicted that the indefinite singular would express [+generic] in characterising English generics but not in MSA. Therefore, the English-speaking group's ratings of the indefinite singular were expected to be high and the Arabic-speaking group's ratings were expected to be low. Considering the significantly high rating of the indefinite singular in generic sentences by the English-speaking group and its low rating by MSA-speaking participants, the results suggest that both groups rated as predicted. Therefore, this hypothesis is supported. This result, as well as those previously discussed, coincide with what Ionin et al. (2011a) found for English as well as the discussions of MSA by Fassi Fehri (2012). The MSA ratings for the indefinite singular and bare singular were statistically similar. Recall that MSA does not have an indefinite article and bare singulars in MSA behave like indefinites. An important implication for L2 acquisition is that acquiring mapping [+generic, − plural, −definite] onto the indefinite singulars in English as an L2 is predicted to be the most difficult generic form-meaning mapping. That is because it requires acquiring the indefinite article before reconfiguring the L2 features onto this form. As discussed in Section 2.4.2, (footnote 5), Fassi Fehri's (2004) analysis shows that in Arabic, referentiality and indefiniteness are not strictly tied to the presence of an overt determiner as in English. Arabic allows for referential DPs via internal structure (e.g., N-to-D movement) or with empty D heads, and lacks an overt indefinite article altogether. This means that when Arab learners acquire English, they cannot rely on simply externalising an already-present [-definite] feature in their L1. Instead, they must reassemble how these features are structurally encoded and interpreted, learning that English *requires* an overt article ("a") for both referential and quantificational uses—a reconfiguration of the feature-DP mapping.

To offer a more complete picture of what acquiring the English indefinite article entails, I discussed its use in episodic sentences, where the article signals a discourse-new, existential interpretation. Consider the English sentence:

- *A dog barked last night.*

Here, the indefinite article *a* introduces a new referent into the discourse — the speaker assumes the listener is not yet familiar with the dog. This use of *a* reflects a particular configuration of features: the noun phrase is [–definite], [–specific], and [+existential/quantificational]. The indefinite article is thus not optional; it plays a key role in encoding both discourse status and semantic scope.

In contrast, Arabic does not have an overt indefinite article. The equivalent sentence in Arabic would be:

- (2) *naḥaḥa kalb(un) ʾams*
  barked dog(indef) yesterday
  "A dog barked yesterday."

In Arabic, indefiniteness is typically inferred through bare nouns. Crucially, there is no overt morpheme that signals the same complex bundle of features associated with English *a(n)*. In Arabic, referentiality and quantification are more distributed — across syntax, morphology, and discourse — whereas in English, they are more tightly encoded within the DP via overt determiners. Thus, when Arab learners acquire the English indefinite article, they must reconfigure the way their grammar encodes and interprets these semantic features.

This is why the acquisition process cannot be described as simply learning to pronounce an existing abstract feature in L1 (a PF externalisation). Instead, learners must reassemble how features like [±specific], [±definite], and [±quantificational] are syntactically expressed and semantically interpreted in English DP — a deeper grammatical shift that affects both form and meaning.

In other words, Arab learners acquiring the English indefinite article aren't just learning to pronounce a new word ("a")—they're learning a new way to build meaning around determiner phrases. Arabic speaking learners must reassemble how features like indefiniteness and genericity are encoded in a determiner phrase, making it a deeper syntactic-semantic shift, not just a pronunciation or surface-level one.

Overall, the findings of this study of native speakers' judgments were crucial for this thesis as it helped to build this study on an empirical base that might account for the variability in the L1 and L2 input. These findings showed how MSA and English differed in expressing genericity; MSA maps generic meaning to definite singular and plural NPs similarly in both characterising and kind generic meanings, while English maps characterising generics onto bare plurals and the indefinite singulars and kind generics onto bare plurals and definite singulars. The principal theoretical implication of this part of the investigation is that it could inform linguistic theory by extending the understanding of how genericity works in MSA through empirically validating existing theoretical linguistic analysis. This study is the first to present evidence for generic form-meaning mappings in MSA. Its findings could help further investigate the role of L1–L2 cross-linguistic differences in the acquisition of genericity by Arabic–speaking learners by providing an empirical account of how L1 and English differed. Several limitations to this supporting study need to be acknowledged. The empirical evidence is limited to the selected structures, characterising generic and kind predicate generics with count NPs, and by the use of only an AJT. In addition, this study was limited in the scope of its sampling and discussion as it was not the main study. This study should be replicated using a broader scope to account for genericity form-meaning mappings in MSA, considering all NP types and all generic contexts and using both production and interpretation tasks. Such an account could provide a broader contribution to the semantic literature by considering languages other than the Romance languages. Moreover, this native speakers' investigation has potential implications for the L2 learning task by highlighting differences between the native speakers of both languages on an empirical ground.

5.7 The learning task

Recall that the theoretical motivation for this thesis stems from the discrepancy between MSA and English form-meaning mappings for genericity and the implications for difficulty of L2 acquisition arising from these differences as well as the impact of L2 instruction on this acquisition. Based on the L1-L2 differences, as shown in Table 5.5, and assuming the FRH, the learning task faced by Arabic-speaking English learners for each generic condition is discussed in what follows providing predictions of its difficulty:

It is posited in this thesis that L1-L2 differences contribute to difficulty in L2 acquisition and that the L2 learning task, as proposed by Lardiere (2009), includes acquiring new language-specific configurations of features. Lardiere (2009) adds that the acquisition challenge varies depending on the context, clarifying that if a feature of the L2 is available in the L1, but the context of that feature is not related to its L2 context, acquiring it is more challenging, though possible. L2 learners may need to disassemble a set of features that are bundled in the L1 and reassemble them into a different bundle of features in the L2 based on the L2 conditioning environment. In this study, the L2 learners have the task of reconfiguring the following L2 features [± generic], [± plural] and [± definite] correctly when mapping NPs to genericity.

Prediction 1: The results of the empirical native speaker study showed that MSA and English similarly map definite singulars to [+kind, −plural, +definite], and that the English definite article is a free morpheme whereas the MSA definite article is a prefix 'al'. According to FRH, The learning task requires the reassembly of [+definite, +kind, −plural] from the overt L1 prefix (al) to the overt L2 free morpheme (the). Cho and Slabakova (2014) added that if reassembly is from an overt morpheme in the L1 to an overt morpheme in the L2, the acquisition of the L2 morpheme is argued to be less challenging when the input provides enough evidence. Therefore, Arabic–speaking learners of English may face no significant difficulty in mapping the definite-singular NP onto the reference to kind generic meaning due to the similarity between the L1 and L2 form-meaning mappings.

Prediction 2: In [+kind ,−definite, +plural], and [+generic ,−definite, +plural] contexts, English and MSA differ. While English maps both meanings to bare plurals, MSA maps these contexts to definite plurals (the L1 feature bundle is [+definite, +plural, +generic]). According to the FRH, the learning task includes disassembling the [+generic] feature from the L1 feature bundle, reassembling it to the feature bundle in L2, and remapping it to L2 bare plural NPs, a form mapped to indefinite meaning in the learners' L1. At first glance, the difference between plural generics in Arabic and English might seem to be just a matter of form—specifically, whether or not there is an overt definiteness marker like *al-* in Arabic. However, as discussed in footnote 5, Fassi Fehri (2004, 2012) shows, the real difference lies deeper, in how each language handles syntax and meaning. In Arabic, bare plurals cannot express general or kind-level meanings unless they are marked as definite, while in English, bare plurals can naturally express these meanings without any special marking. This means the contrast is not just about how the words sound or look (i.e., not purely a PF issue), but about how sentence structure and meaning interact. Still, for Arabic-speaking learners of English, the difference may be misunderstood as one of surface form. These learners might assume that, just like in Arabic, a noun without a definite article in English cannot have a generic meaning. As a result, they may incorrectly apply rules from their first language. So, while the core difference is not about PF from a theoretical standpoint, it can seem that way to learners. Therefore, Arabic–speaking learners of English are expected to face difficulty acquiring this form-meaning mapping in English and overgeneralise the L1 use of the definite plural NP with a generic interpretation as a result of L1 transfer. The fact that bare plurals in English are ambiguous, and can express generic and existential readings, as discussed in 2.4.1, makes the bare plural the most difficult acquisition task as the L2 learners may encounter both meanings in the L2 input and have to rely on the context to disambiguate each use.

Prediction 3: In characterising generics [−definite, −plural, +generic], English maps this meaning to indefinite singulars and MSA maps the characterising generic to the singular definite NP. The learning task includes reassembling the [−definite, −plural] features from the bare singular in the L1 (Ø covert form) to the indefinite article ‘a’ in English. Then, they need to reconfigure the [+generic] feature from the L1 feature bundle [+definite,−plural, +generic], which is mapped onto the definite singular (al) in MSA and allow it with the L2 configuration [−definite, −plural, +generic] and remap the new feature bundle to the L2 indefinite article ‘a’. Due to these reassembly requirements, it is anticipated that Arabic-speaking learners would exhibit lower levels of accuracy in mapping the English indefinite singular article to the characterising generic meaning. In this generic condition, they may tend to rely on the L1 form.

Building on Slabakova’s (2009) and Cho and Slabakova’s (2014) cline of difficulty proposals, a detailed cline of difficulty in acquiring English generic form-meaning mappings by Arabic-speaking learners is presented in Figure 5.3.

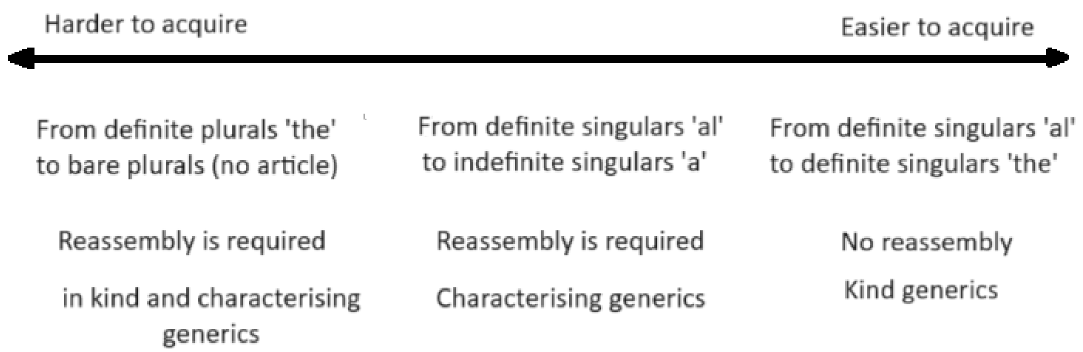


Figure 5.2 Difficulty in learning English generic form-meaning mappings by Arabic-speaking learners.

Prediction 4: Success in mastering the learning tasks mentioned above is possible; therefore, providing the learners with instruction on generic properties, how the forms are mapped to the meaning in English, and how it differs from the L1 is predicted to improve the L2 learners’ form–meaning mapping accuracy in L2 characterising and kind generic meanings.

5.8 Summary

The main goal of this chapter was to investigate how genericity is expressed in MSA and English by exploring native speakers' acceptance of five NP conditions in two generic categories, namely kind and characterising generics. This chapter detailed the methods, procedures and results of an experimental study. The general results reveal that, although the definite singular condition denotes kind interpretation at kind generics similarly in both languages, they differ under other conditions. In particular, MSA uses both definite singular and plural morphemes in kind and characterising generics. In English, bare plurals were found to denote genericity in both categories, whereas the indefinite singular can express generic meaning in characterising generics only and the definite singular can denote kind reading. Although this study was based on a small sample of participants, it was a critical element of this thesis. This study served as a source of MSA native baseline. This baseline's generic form-meaning mappings were compared to the English baselines to validate the predictions of acquisition difficulty in the study.

Chapter 6 Methodology

6.1 Introduction

This chapter presents the experimental design, participants, data collection methods, procedures for data collection and analysis, and the instruction intervention of the present study. The chapter begins by restating the research goal and questions and provides an overview and justification for the research design. In Section 6.3, the study participants are presented. The experimental methods and procedures of data collection are outlined in Section 6.4 and 6.5. Details on the instruction intervention are presented in Section 6.6. Specifically, the rationale for the intervention, the context, the content and procedures for implementing instruction sessions are outlined. In Section 6.7, procedures for coding and analysing the data are described.

6.2 Research Design

To recall, the overarching goal of this thesis was to determine the relative ease with which Arabic-speaking learning of English as an L2 can acquire English generic form-meaning mappings, as well as whether or not an intervention based on findings from SLA research impacts this acquisition. This thesis seeks to answer two research questions, which are repeated here:

RQ1: Do intermediate Arabic-speaking English learners demonstrate target-like mapping of English articles to kind and characterising generic meanings? If not, what is the difficulty and how can it be accounted for according to FRH and BH?

RQ2: After being exposed to an 8-week acquisition-informed explicit instruction intervention on English genericity form–meaning mappings, can the experimental group improve their form-meaning mappings in kind and characterising generic meanings compared to the comparison group? Is any such improvement retained three months after instruction?

To investigate the impact of theoretically informed classroom instruction on the acquisition of English genericity by Arabic-speaking L2 learners, this study utilised a quasi-experimental design. According to Rogers and Révész (2020), the core of experimental and quasi-experimental research designs is examining the possibility of a cause-and-effect relationship between independent and dependent variables. The independent variable is the variable that is manipulated and anticipated to cause a change in the dependent variable

(Mackey and Gass, 2021). The difference between true- and quasi-experimental designs is the lack of random assignment for participants in the latter. This study examines how L2 instruction (the independent variable) affects the acquisition of L2 genericity (the dependent variable). This cause-and-effect relationship between the variables rationalised the use of the quasi-experimental design. I opted to use a quasi-experimental design because it was impossible to use a true experimental design in light of conducting the research with existing L2 classes where the learners were already assigned by the institution, as detailed in Section 6.3. This study design consisted of quantitative research that follows four stages: pre-test, treatment, post-test, and delayed post-test, as summarised in Table 6.1. Mackey and Gass (2021) discuss some advantages and drawbacks of the pre-test/post-test design stating that it has the advantage of determining the immediate effect of the treatment through post-test and that using delayed post-tests can give a more comprehensive picture of the effect of a treatment. However, losing participants, and increased extra-experimental exposure are some downsides.

Table 6-1 Experimental Design

L2 Groups	**Week 1**	**Week 2 to Week 9**	**Week 9**	**Week 21**
Control group	Pre-test	Traditional grammar classes without genericity-focused treatment.	Post-test	Delayed post-test
Experimental group		Genericity-focused treatment		

It has been stated that the non-random assignment of participants in a quasi-experimental design can limit external validity, that is, the generalisability of the results (Rogers & Révész, 2020). However, this thesis accounts for the external and internal validity limitation resulting from the non-random assignment of L2 learners by controlling sampling, data collection and treatment procedures, as illustrated in the subsequent sections.

6.3 Participants

This study had two language groups: the native control group and the Arabic-speaking L2 learners. The L2 learners were divided into two further groups after the pre-test: the experimental and comparison group. I opted to label the L2 learners' control group as a comparison group to maintain clarity and avoid the confusion that may result from having native and L2 control groups with similar labels when reporting results.

6.3.1 Native control group

The use of a native control group is a key methodological consideration in this investigation. This group served different purposes. First, it served as the source of the baseline data to which L2 learners' data is compared. Domínguez and Arche (2021) argue that having such a baseline group is necessary for providing meaningful explanations of L2 acquisition. According to them, interlanguage grammar describes the stages that second-language learners go through as they work towards a target and the native control provides the baseline to which interlanguages are compared. For L2 acquisition, the appropriate native controls are people who have a fully developed (target) grammar. It makes sense to consider native speakers' grammars as a potential destination along the road of L2 acquisition. Second, evidence from native control behaviour is crucial for verifying the suitability of the instruments and the validity of the theoretical assumptions (Domínguez and Arche, 2021).

In total, twenty undergraduate native speakers of English from the University of Southampton comprised the native controls. They were aged between 20 and 27 (mean 23) and were recruited by local contacts at the University of Southampton. All were living in the United Kingdom at the time of data collection. The decision to have a control group that was different from the native English speakers tested in Chapter 5 is that this group matched the L2 learners in age and level of education. Moreover, the native English speakers in the previous study only completed an AJT. Therefore, there was a need to have another native control group for the other tasks in this study to set the baseline and validate the tasks. This group completed all three tasks—the elicited production task, the acceptability judgement task, and the forced choice task—before the intervention had started.

6.3.2 L2 learners experimental and comparison groups

The study started with seventy female undergraduate learners of English at King Khalid University, Saudi Arabia (total *n* = 70). However, some participants couldn't attend all 8-week sessions or were absent in the post-test and, therefore, were not included in the study. Based on this exclusion of absentees, the study's sample consists of sixty-four participants. All L2 speakers were recruited and tested in their home country, where English is a foreign language (Saudi Arabia). They were selected from two intact level-four classes in the English language and translation department. One of the classes acted as the experimental group (*n* = 35), while the other was the control group (*n* = 29). To increase the validity, these two intact groups were randomly assigned to be the experimental and comparison groups, according to guidelines from Mackey and Gass (2021).

The decision to conduct the intervention on level-4 students is taken with consideration of the curriculum in the department. In this level, L2 learners have completed three prior levels in the department. In the first two levels, the focus was on teaching skill-based courses—reading, writing, listening, and speaking. In level 3, the participants received explicit instruction on core grammatical structures, including verb tenses and the English article system. Therefore, the learners in level 4 are assumed to have a foundational knowledge of the basic rules governing the use of English articles, particularly with respect to definiteness (i.e., the distinction between definite and indefinite noun phrases), as they have already completed level 3 and passed level 3 grammar exams that tested their understanding of these concepts.

The selection of lower-intermediate (level-4) students is also based on pedagogical and practical considerations. At this stage, learners have moved beyond elementary proficiency but have not yet reached advanced levels, which makes them ideal for introducing more nuanced and conceptually challenging uses of articles, such as those involving genericity. It is assumed that these learners are at a developmental stage where they can benefit from focused instruction on the subtleties of article use, particularly in conveying generic meaning, while still being flexible in their language acquisition process. Moreover, working with this group allows for identifying and addressing persistent gaps or misconceptions about articles—especially where definiteness and specificity overlap or diverge.

As reviewed in Section 3.3.3.1, a substantial body of research has shown that the English article system remains one of the most challenging aspects of grammar for L2 learners, particularly for those at the lower-intermediate level (Ionin, Ko, & Wexler, 2004; Master, 1997). This difficulty is even more pronounced among Arabic-speaking learners of English, whose native language lacks an equivalent article system. According to Crompton (2011) and Abdullah (2016), Arabic-speaking learners frequently misuse English articles, especially in contexts that involve generic reference, as they often transfer rules from their L1 that do not align with English article usage. These learners tend to overuse the definite article "the" or omit articles entirely, even after receiving formal instruction. The challenge is not only grammatical but also conceptual, as the notions of definiteness, specificity, and genericity do not map directly between Arabic and English. Therefore, learners at this level—though assumed to be familiar with the article forms—require further instruction to grasp the underlying semantic and pragmatic rules that govern their appropriate use. This makes teaching generics more feasible for students who have been introduced to the basic functions of English articles but have not yet been exposed to the more abstract semantic and pragmatic dimensions of generic reference.

According to Mackey and Gass (2021), using intact classes can limit internal validity and matching the intact classes as much as possible is essential to increase study validity. To this end, it was important to ensure that the learners in both groups were homogenous in terms of age, language background, and proficiency level. Before the intervention, both L2 groups completed a language background questionnaire to ensure they had similar previous language experiences. Participants' ages, years of English language education, L1s, possible additional languages, and whether they have lived in an English-speaking country before, were collected through this questionnaire. The questionnaire revealed that none of the participants grew up in an English-speaking country, and Arabic is their native language. No one speaks any additional language besides English. This eliminates the possibility of confounding factors, such as exposure to L2 in a natural context and having a third language.

To match the L2 learners' groups in proficiency, the L2 groups' proficiency was assessed using a cloze test adapted from the Standardized Oxford Proficiency Test (details of which are provided in Section 6.4.1). Table 6.2 summarises proficiency levels and background information of L2 learner groups.

Table 6-2 Participant information of the L2 learner groups

Group	**Age in years**	**Years of Studying English in school**	**Proficiency levels**
Experimental group (n = 35)	M = 20, SD=0.88	M = 9	Beginners (n= 2) Intermediate (n = 33)
Comparison group (n = 29)	M = 19, SD= 0.77	M = 8.5	Beginners (n = 3) Intermediate (n = 26)

It was important to ensure that both groups were homogeneous before the intervention. To do so, the groups' scores in the proficiency cloze test were analysed. The experimental group participants' average language proficiency score (M = 20.2, SD = 0.67) was higher than the control group participants' average proficiency score (M = 19.72, SD = 0.79). The data had no outliers, as assessed by inspection of the violin plot in Figure 6.1. The language proficiency scores were not normally distributed, as assessed by Shapiro-Wilk's test (w =0.97, p < .001), and variances were not homogeneous, as assessed by Levene's test for equality of variances (F =4.80,p = .02). Therefore, a Wilcoxon rank sum test was run, which showed that the scores of language proficiency between the groups were not statistically different (W = 280.87, p =.15). Therefore, proficiency level was ruled out as a confounding factor.

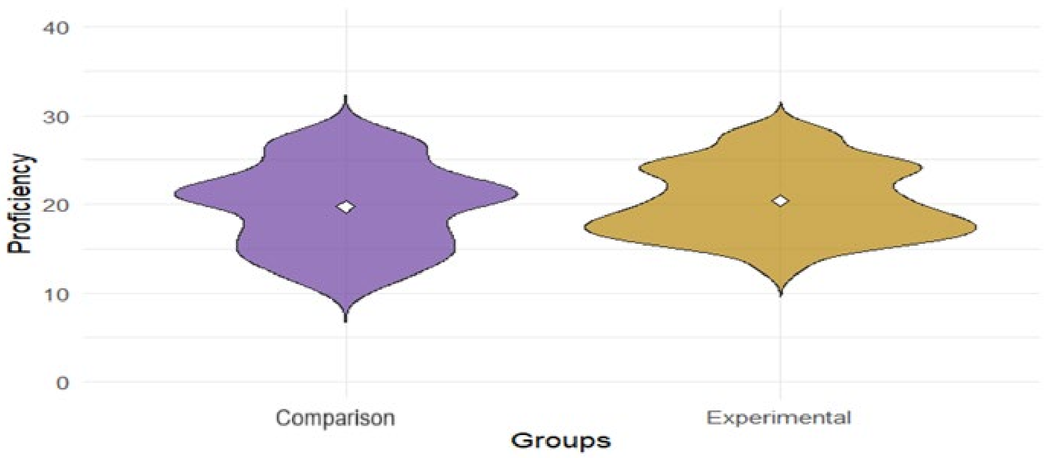


Figure 6.1 L2 learner groups' proficiency

6.4 Data collection methods

Norris and Ortega (2003) stated that "SLA researchers must acknowledge that a single measure will not provide a sufficient evidence for informing the range of interpretations typically sought in most SLA studies" (p.748). Therefore, this thesis employed various tasks to measure the acquisition of generic form-meaning mappings by L2 learners and the impact of instruction on this acquisition. This study used a written elicited production task, an acceptability judgement task with contexts, and a forced-choice task to prevent the task type from influencing the outcome of the data, thereby increasing internal validity. All the tasks were in paper and pencil format. In this section, I describe the reasoning behind task selection, task content and structure, any noted weaknesses of each task, the descriptive results from piloting the tasks, and any adjustments that were made based on piloting.

6.4.1 Proficiency cloze test

In this study, the proficiency measure consisted of a cloze test adapted from the Standardized Oxford Proficiency Test. This test has the advantage of being user friendly and measuring general English language ability with a focus on grammar. The original test consisted of 100 multiple-choice items. Following Slabakova and Garcia Mayo (2015) and Aboras (2020), I used 40 grammar-oriented items as a subset of this test to test the L2 learners' proficiency. The first twenty items included individual sentences where a gap needed to be filled with a word or a phrase from three possible answers. The focus of these items is the form. The second twenty items had the same format but included a story in which correct answers were related to the context and grammar.

The results of the test were used to assign participants to three proficiency levels (beginner, intermediate, and advanced) based on the following criteria; Learners who scored more than 28 out of 40 (70% correct answers) were considered advanced. Those who scored less than 16 (40% correct answers) were placed in the beginner group. Those who scored between 16 and 28 were assigned to the intermediate group. None of the participants were advanced and only a few were beginners as shown in Table 6.2 above. These participants were not excluded as the results showed a homogeneous proficiency in both groups.

The use of this task is not without limitations. First, it was limited to using a subset of the original test. This is not a problem for this study as the main aim of measuring proficiency is to ensure that the L2 groups are homogenous before the intervention. Therefore, a 40-item subset is economical and suitable for the purposes of this study. If the study aimed to investigate acquisition by learners across various proficiency levels, it would be necessary to use a full standardized copy of the test. Second, this test is limited to grammar-based proficiency. Proficiency is a difficult area to control for as it consists of many skills and grammar components (Mackey & Gass, 2021). However, since the aim of this study is concerned with teaching grammar and measuring improvement related to generic form-meaning mappings, the use of grammar only as a measure of proficiency is suitable for the purposes of the study.

6.4.2 Elicited written production task (EWPT)

Research into SLA has a long history of relying on production tasks, despite their limitations. Selinker (1972) argued that interlanguage analyses should be based on what is visible, specifically students' attempts at production in "meaningful performance situations" (p. 214). Ellis (2005) categorised production as a measure of implicit knowledge. Ionin & Montrul (2023) highlight the importance of using different measures in intervention studies to tap into different knowledge types. This task is rationalised by the need to prevent the task type from affecting the study results.

White (2020) argues that if the task is set up to elicit production of a certain structure and learners avoid this structure, this indirectly suggests that the learners face difficulty in this structure. In light of this, the EWPT task in this study was set up to force the use of articles and NPs in both kind and characterising generics in production. In particular, this task includes eight items. For each item, the learners were given a short conversation as a context and two wh-questions to generate a generic use and an anaphoric article use in light of the picture cue and the context; the anaphoric NP production is used as a distracter for each item (see Appendix B for the full task). This task focuses on the following meanings: kind generic meaning which includes two conditions, namely [−plural] [+kind] and [+plural] [+kind], with two test items targeting each condition; and characterising generic meanings which includes two conditions, namely [−plural] [+generic] and [+plural] [+generic]. Two test items targeted each characterising generic meaning condition. The following are examples of each category:

Example 1: Kind generics:

Classmates talk in art class: **Jack:** What should we draw for the art group project? **Tim:** I guess animals, especially the one I am drawing right now. **Jack**: If you visit Australia, Tim, you will see lots of this. [pointing at the drawing]. **Q1: Which animal is Tim drawing?** **Q2: Which animal is common in Australia?**	

Example 2: Characterising sentences.

Mother/ daughter talk: **Mother**: Try to eat more vegetables. **Sara**: I hate it. [pointing at the picture]. **Mother:** They are colourful and decrease the risk of heart disease. **Q1: Which vegetables does Sara hate?** **Q2: Which food reduces heart disease risks?**	

This task allows the production of genericity without focusing the learner's attention on the articles in a generic context. This enables tapping into unconscious use of articles in generic sentences. To avoid learner boredom and comply with the study's time limitations, this task is kept short. The expected answers in the obligatory contexts of test items are summarised in the following Table 6.3:

Table 6-3 Expected Production

Items	Expected production
Characterising generics (singular)	Indefinite singulars
Characterising generics (Plural)	Bare plurals
Kind generics (singular)	Definite singulars
kind generics (plural)	Bare plurals

A disadvantage of production tasks is that the researcher does not have control over the production of the tested structure; production does not necessarily reflect the L2 learners' knowledge; production data may lead to underestimating or overestimating the L2 learners' competence (White, 2020). Moreover, in freer production tasks, learners may avoid producing some structures (Ionin &Monrul, 2023). To deal with this limitation, the decision was made to keep this production task controlled by the contexts, the questions and picture cues to elicit generic form-meaning mappings production. Moreover, this task is still used in conjunction with other tasks to get a better picture of the learners' mental grammars, as Ionin and Zyzik (2014) suggest.

It is important to note that this task is designed by the researcher of this study. To validate the task, two English native speakers checked the task wording, contexts, and predicted productions. Moreover, I piloted the task to ensure that it was suitable for eliciting the expected production summarised in Table 6.3 and to check for potential challenges before using the task for data collection. The descriptive results of piloting and amendments are discussed in the following section.

6.4.2.1 EWPT piloting

To test the appropriateness of the instrument described above, I piloted the method using a small sample (n = 6) of intermediate Saudi undergraduate learners of English who did not participate in the main study. The average percentages of both target-like and non-target-like productions were calculated and are summarised in Table 6.4.

Table 6-4 Results of EWPT piloting.

Item	Expected Production	Average % of Target-like production	Type of error	Error Average %
Characterising generics (2 test items)	Indefinite singulars	0%	Bare singulars Bare plurals Definite singulars Definite plurals	**41.7%** 25% 25% 8.3%
Characterising generics (2 test items)	Bare plurals	41.7%	Definite plurals Definite singulars Bare singulars	**33.3%** 8.4% 16.7%
Kind generics (3 test items)	Definite singulars	33.3%	Definite plurals Bare plurals Bare singulars	**33.3%** 16.7% 16.7%
Kind generics (1 test item)	Bare plurals	33.3%	Bare singulars Definite plurals Definite singulars	16.7% **33.3%** 16.7%

The results show that mapping indefinite singulars to singular characterising generic contexts in production is particularly challenging because there is 100% inaccuracy. For plural generics in characterising generics, 33.3% of the sample overgeneralised the L1 form, definite plurals.

For kind generics, 33% of the sample correctly mapped the definite singular to the generic interpretation, while 33% overgeneralised the use of the definite plural, regardless of number agreement. For plural kind generics in particular, 33.3% of the sample correctly expressed kind-generics with bare plurals, while 33.3% overgeneralised the use of definite plurals. Due to the small sample size, no inferential statistical findings are reported.

Based on feedback from pilot participants and native speakers, several adjustments were made. Firstly, visual cues were carefully selected to avoid unnecessary difficulty. In particular, in the old version a photo of a carrot was used to elicit the production of bare plurals in generic context such as ‘ root vegetables contain vitamins’. The pilot study results showed that the choice of this picture cue led to the use of a singular NP rather than bare plurals which caused unbalanced design and a difficulty in eliciting ‘root vegetables’ from the participants. The final version included a picture of a group of red vegetables to easily elicit bare plural NPs in generic context, as appendix B.1 showed. The pilot study revealed an imbalance with three test items focusing on singular forms and only one on plural forms with kind generics. In particular, three items led to the production of singular definite NP as in 'the dog is faithful'; the kangaroo is common in Australia, the owl is rare in Antarctica'. To avoid this, the final task was adjusted to ensure an equal number of questions per condition; i.e. two items for singular kind generics and two items for bare plurals, as shown in appendix b.1. The native speakers who validated the task highlighted the importance of context length and complexity as methodological considerations. Therefore, in the final task, all test item contexts were rewritten to maintain uniform length and format.

6.4.3 Acceptability Judgement Task with Context

The second task in this study is an AJT with contexts adopted from Ionin et al. (2011a). This task was chosen because it can 'inform us about the learners' linguistic competence in their target language' (Ionin, 2012, p. 31). Gross (2021) describes linguistic judgements as products of mental aspects from which theorists infer a theory. In this sense, an AJT is appropriate for this study as it enables inferring about L2 learners' mental grammar development based on their judgements. If the students have learned to map the semantic meaning to the morphosyntax, they would be able to activate a specific interpretation when they encounter the target morphosyntactic property. This will help them select appropriate forms in the generic contexts or reject non-matching ones. Moreover, an advantage of AJTs is that they provide more information about the L2 learners' grammar, i.e., knowledge of ungrammaticality (Mackey & Gass, 2021)

This task consisted of contextualised acceptability judgements, in which the learners are asked to read a context in the form of a story followed by five sentences. These five sentences are identical, apart from the subject NP. The sentences for each context include five different nominals: definite singulars, indefinite singulars, bare singulars, definite plurals, and bare plurals. This task design ensures that any differences in rating will be based on the interpretation of the subject NP in the generic context, thereby meeting the goal of the present study. The original task developed by Ionin et al. (2011a) included control and test categories. For this study, only kind generics, characterising generics, and distractor test items were kept from the original AJT as this study focused on kind and characterising generics only. For each generic meaning, the original task had 4 test contexts. However, in order to generate more evidence for the learners' knowledge, the researcher added two test items for each generic meaning. The total number of test items is six for each generic meaning, with eight distractors that focus on tenses, as shown in Appendix B.

For each test item, the participants were asked to rate its acceptability on a Likert scale from 1 to 4 (where 4 is "highly accepted", and 1 is “completely unacceptable”) for each of the five sentences (each sentence refers to a condition) in light of the context provided. Only one rating for each sentence is accepted and the participants were instructed that they may give the same rating to two or more of the sentences that follow a single story. Examples of each category are given in (1) for kind-denoting generics and (2) for characterising generics, where expected ratings are bolded:

(1) Kind generics

I know that you like birds. Well, if you ever visit California, you will see lots of different kinds of birds there. For example...

a) Pelican is widespread on the California coast.	**1**	2	3	4
b) The pelicans are widespread on the California coast.	**1**	2	3	4
c) The pelican is widespread on the California coast.	1	2	3	**4**
d) Pelicans are widespread on the California coast.	1	2	3	**4**
e) A pelican is widespread on the California coast.	**1**	2	3	4

(2) Characterising generic

My brother has been in a bad mood lately. And no wonder – his apartment is so uncomfortable; it must be very depressing to live there. And he has a very dim and unpleasant overhead light. I told him he should buy a new lamp – something pleasant. For example, I know that...

a) A green lamp is very relaxing.	1	2	3	**4**
b) Green lamp is very relaxing.	**1**	2	3	4
c) The green lamps are very relaxing.	**1**	2	3	4
d) The green lamp is very relaxing.	**1**	2	3	4
e) Green lamps are very relaxing.	1	2	3	**4**

According to Ellis (2005) and Ionin and Montrul (2023), timed AJTs tap into implicit knowledge. Therefore, a limitation of this AJT is that it was untimed. Untimed AJTs may tap into explicit grammatical knowledge and intuitions rather than learners' implicit mental grammars. The decision not to time this task was made because it was difficult to decide on a specific time limit that has not been established in the pilot or in the native speaker study. However, this task includes contexts and therefore taps into interpretations rather than just explicit grammaticality. Moreover, caution was made during the administration of the task as the learners were not allowed to go back and change their answers to ensure that they don't rely on the explicit grammatical knowledge, as suggested by Mackey and Gass (2021).

6.4.3.1 AJT Piloting

The main aim of piloting was to validate the task and to anticipate any challenges in the task instructions or content. The same pilot sample that piloted the production task also piloted the AJT (n = 6). Only descriptive statistics for the piloting are presented due to the small sample size. The descriptive statistics seem to support the predicted difficulty in Section 5.7. Figure 6.2 shows the mean ratings of kind generics. Although the sample rated bare plurals highly, the expected overuse of the L1 definite plurals is noticeable (the mean rating was 3, which is similar to or higher than the midscale score of 2.5). The sample rated the definite singular high, indicating acceptability.

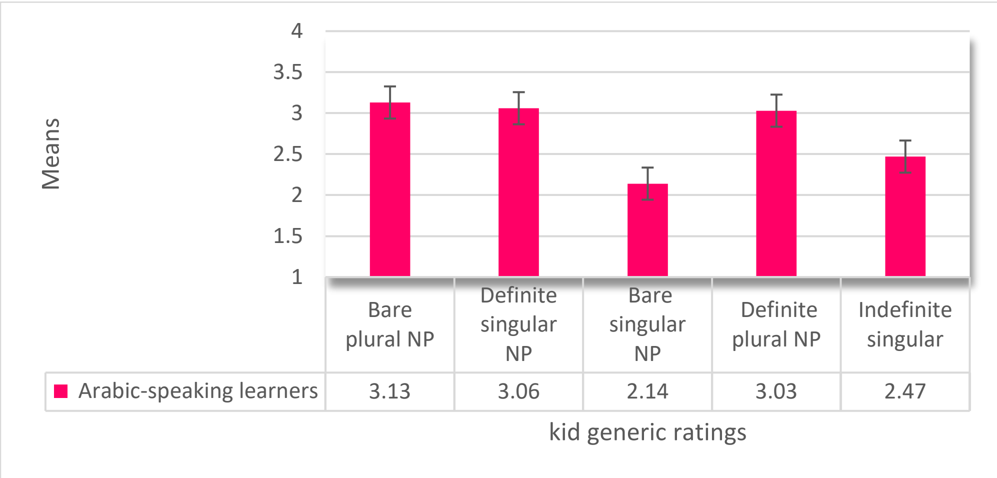


Figure 6.2. Descriptive statistics for kind generics

For the characterising generics, Figure 6.3 shows a similar pattern to the kind generic mean ratings, namely accepting both the definite plural and the bare plural. It is interesting that the indefinite singular NP is rated higher than the midscale point, and the high ratings of the bare singulars may further reflect the inaccuracy of the sample's ratings; the definite singular is rated as expected similar to the L2 singular form.

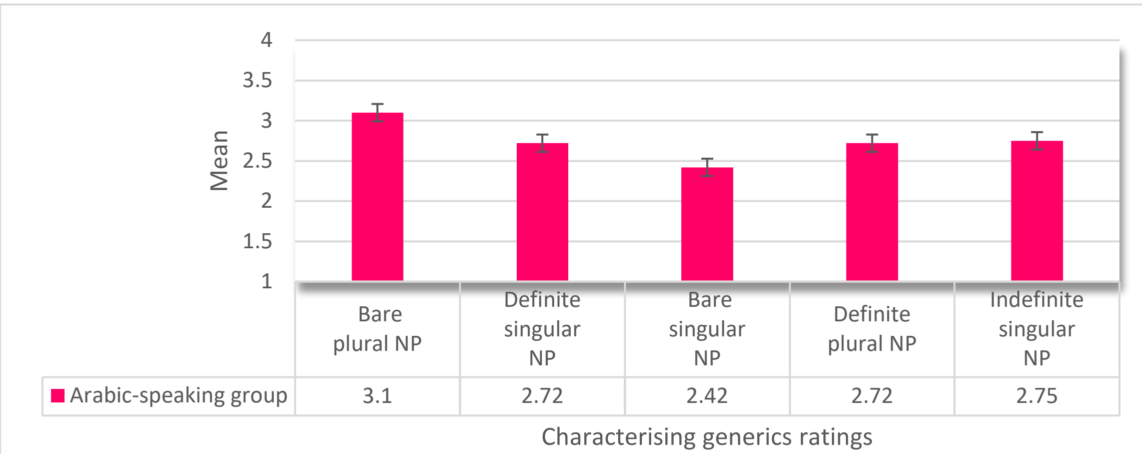


Figure 6.2 Descriptive statistics for characterising generics

In light of the piloting, some methodological decisions were considered to improve this task. First, the piloting revealed that some test items combined a kind denoting NP with a characterising statement. Such sentences were not included in the intervention and therefore were replaced with sentences that only express characterising generics or kind generics. Second, the task was revised in terms of vocabulary suitability and clarity of contexts based on the feedback gathered from the pilot sample.

6.4.4 Forced-Choice Task (FCT)

The third task in this study is a forced-choice task. Ionin and Zyzik (2014) classified this as a type of preference task, in which participants are required to select one of two (or more) related forms in the target language. According to them, preference tasks can be used to test both interpretation (i.e., one form expresses the target interpretation better than another) and grammaticality. In general, some context comes before the choice of form if interpretation is targeted. In light of this, I decided to use an FCT to measure the L2 learners' use of articles in generic interpretation for this thesis. This task allows the evaluation of learners' explicit knowledge of articles used in generic contexts, according to Ellis's (2005) classification of explicit tasks. This task facilitates the elicitation of more controlled data, yet, it also contributes to a fuller understanding of L2 learners' acquisition by providing information on their production of articles in particular in addition to the NP production in the first task and judgements of NPs in the AJT.

In this task, the participants were asked to read a context and fill in a blank by choosing only one option of the English articles (*a*, *an*, *the,* or Ø) based on the context. Ø refers to the no article option, as clarified in the task's instructions page. The context is provided in the form of conversations. This task consists of 24 items: eight items test kind generics, eight items test characterising generics, and eight are distractors. Ten items in this task were adopted from Snape et al.'s (2013) study, while the others were created by the researcher. The task is attached in Appendix B. This task focuses on the following conditions in each generic context:

Context 1: Kind generics

[+kind][+definite][−plural] (Four test items)

[+kind] [−definite] [+plural] (Four test items)

Context 2: Characterising Generics

[+generic] [−definite][−plural] (Four test items)

[+generic] [−definite] [+plural] (Four test items)

The distractors included eight items that focused on tenses. Example test items from this task are provided for each category below.

1. Examples of kind generics

Condition: [+kind] [−definite] [+plural] bare plurals are the targets

A: I am planning to visit North America this summer.

B: It is a great choice. You will enjoy discovering the history of some flowers.

A: For example, __ Ø __ Sunflowers were originally planted in North America.

the	a	an	Ø

Condition 2: [+kind][+definite][−plural] definite singulars are the target

A: I think that climate change is leading some species to leave their home.

B: Seriously!

A: _The_vulture is widespread in South America. In the last decade, they have been moving north as a result of warmer weather.

the	a	an	Ø

(2) Examples of characterising generics:

Condition : [+generic] [−definite] [−plural] indefinite singulars are the target

A: Why are they criticising Prof. Brown?

B: Well, __a__ scientist should be able to produce evidence supporting his theory.

A: I agree.

the	a	an	Ø

Condition : [+generic] [−definite] [+plural] bare plurals are the target.

A: Many food experts say that people should consider eating vegetables.

B: Do you think it is important?

A: I am not sure, but_ Ø _ green vegetables are highly digestible.

the	a	an	Ø

6.4.4.1 FCT Piloting

As this task included some items that were designed for this study, piloting the task to anticipate any potential challenges was essential. The same sample (n= 6) used in piloting the previous tasks also piloted this task. The frequencies of the sample's choices for each test item were calculated, and the average percentages of the students' choices in each category (kind and characterising generics) are reported in Table 6.5.

Table 6-5 Average percentage (%) of the participants' accurate and false choices in FCT for both generic categories.

Categories	**Kind**	**reference**	**generics**	**Characterising**	**generics**	
[−plural] [+generic]	**the**	*a	* Ø	*the	**a**	* Ø
	66.70	25	8.30	29.15	50	20.85
[+plural] [+generic]	*the	*a /an	**Ø**	*the	*a	**Ø**
	62.50	8.35	29.15	66.70	16.65	16.65

Note: The symbol Ø represents that no article is required. * means none-target choice.

The average percentage of correct choices of selecting '*the'* with singular NPs in kind generics is 66.67%; this matches the AJT results for definite singular NPs in generic meanings. For plural kind generics, definite plurals were incorrectly selected by a majority of the sample (62.5%), and bare plurals were correctly selected by 29.15%. This highlights the challenge of these two conditions.

The results for characterising generics showed that, while an average of 50% of the students correctly selected indefinite singulars in singular contexts, 29.2% of the sample incorrectly selected definite singulars. As for plural characterising generics, the results reveal that there is a high tendency for the L1 pattern of the definite plural NP in generic plural contexts (66.7%), with only 16.65% correctly producing bare plurals in this context. This means that both definite and bare plurals were challenging for the pilot study participants.

Based on the pilot study, several modifications were implemented to the FCT. The pilot task analysis highlighted an unbalanced design; the number of singular and plural test items for characterising generics was not balanced in the piloting task (4 test items for the singular and 2 test items for the plural generics). Moreover, the task included two test items that combine both kind-denoting subjects and characterising generics. The following examples showed these test items:

A: Many food experts say that people should consider eating potatoes.

B: Do you think it is important?

A: I am not sure but,____ potato is highly digestible.

the	a	an	Ø

(1) A: Animals have different diets. Some of them eat plants

B: Exactly. __panda feeds on bamboo leaves.

the	a	an	Ø

This forces the use of the definite singular (the) in characterising sentences, a context that was not included in the planned intervention treatment nor in this study aims. As a result, the task was adjusted so that there would be an equal number of test items in each condition as shown in appendix b.3, and generic statements that combined both generic categories (as in the examples above) were avoided in the adjusted version. The revised version was validated by three native speakers and the native control group.

6.5 Procedures

Before any data collection, ethical approval for this study was obtained from the Faculty of Arts and Humanities Ethics Committee at the University of Southampton (ERGO no. 71758). The approval email is found in Appendix E. The study started with baseline native speaker data collection in the United Kingdom. Then, the researcher moved to Saudi Arabia to conduct the intervention study. The researcher arranged using intact classes with the Department of English Language and Translation at King Khalid University, Saudi Arabia. This study began with a meeting with the experimental group. The researcher introduced the study's aim and procedures to the participants using the information sheet. The learners were given a week to decide. After that, the participants signed consent forms. The same introductory procedures were followed with the comparison group in a separate session. A day before the pre-test session, both groups filled out the background questionnaire and completed the proficiency test. The proficiency test results were analysed before conducting the pre-test to ensure that the groups were homogenous before the study started. A week before the teaching intervention started, both groups were given a pre-test to measure their knowledge of genericity, consisting of the EWPT, the AJT, and the FCT in paper-and-pencil format, in the presence of the researcher in the classroom. Before starting each task, the researcher explained the instructions in both the L1 and L2 and discussed the example presented on the task instruction page. To maintain clarity when explaining the task instruction, each task instruction sheet with the example was displayed on the classroom screen. During this explanation, the participants were informed that they were not allowed to go back and change their responses. Each task was distributed after the completion of the previous task by the whole group. Participants were encouraged to ask questions or ask for translations if they were faced with a difficult word during each task. All tasks were conducted offline with the presence of the researcher and the grammar instructor to ensure that all participants were acting as directed.

Each group was given the same sequence of tasks to complete: the EWPT, the AJT, and the FCT. The decision to start with the production task aimed to prevent their attention from being influenced by articles under the influence of the other tasks. Then, an eight-week-long explicit teaching intervention was conducted for 40 minutes twice a week for the experimental group in addition to the basic grammar course sessions. The comparison group received traditional instruction on grammar course content without reference to genericity. The intervention content and procedures are described in the following section. Then, a day after session 8, the first post-test was conducted with the same copy and procedures as the pre-test. Finally, a delayed post-test was held 12-weeks after the post-test. The researcher contacted the participants about the date and location for the delayed post-test via email and an announcement was posted by the department on the announcement board. Then, on the assigned date the researcher conducted the delayed post-test with the same copy and procedures used for both pre- and post-tests.

The study utilised the same test for each session to ensure comparability. However, this approach has a drawback: using the same test may increase the testing effect, i.e., learning by testing (Mackey & Gass, 2021). In this study, a testing effect did not occur because the extended intervals between sessions were sufficient to prevent it. Rogers and Révész (2020) highlight another potential threat to the validity of the study's pre-test and post-test procedures: participants in the control and experimental groups might communicate about the study outside the experiment, potentially affecting the findings. To limit this threat, the experimental and comparison groups were selected from two different campuses which provided the same English and translation program in order to minimise the possibility of between-group communication. Detailed explanations of the instructional content and procedures are provided below.

6.6 The Intervention Design

This intervention was developed based on a comprehensive review of literature on the acquisition of English generic form-meaning mappings, and empirically investigating how it is expressed in the learners' L1 and L2 and what possible difficulties may be predicted by the FRH and BH. This section presents the rationale behind the intervention, the context in which this intervention was conducted, the content of each session and the classroom procedures.

6.6.1 Rationale for the intervention

Recall that some instructional interventions on articles tend to oversimplify the article uses (Master, 1990, 1994, 1997, 2002) in a way that could support the learning of some functions of English articles. Yet this simplification leads to neglecting some complex situations in which articles are used, leading learners to misconceptions of some functions, such as indefiniteness and specificity. Recent studies tend to make use of generative language acquisition research on the role of L1 in L2 acquisition to inspire second language teaching, with special regard to genericity, especially in contexts that lack exposure to natural input (Snape and Yusa, 2013; Snape et al., 2016; Umada et al., 2019; Sabir, 2015, among others). Although the outcomes of such interventions on acquisition do not lead to long-term support of acquisition, these studies successfully shed light on important methodological and instructional content considerations.

In light of the reviewed studies in Section 4.4, the intervention is affected by various factors that need to be considered during the design of the intervention: the content of the intervention, the length of the intervention, the use of both positive and negative evidence, the effect of L1, and the teacher effect. This intervention was based mainly on Krifka et al. 's (1995) framework; therefore, teaching the main semantic characteristics of genericity and how they map to different forms was the focus of this intervention, considering the cross-linguistic differences between the sample's L1 and L2. The classroom teaching was presented using classroom discussion and guided communicative activities. As the previous intervention on genericity, in Section 4.4.2, showed that genericity did not receive longer than 3-week teaching sessions, this study was designed to be longer, including 16 teaching sessions over 8 weeks (80 minutes per week). After presenting the context in which this intervention was conducted, the content and procedures of each session are illustrated.

6.6.2 Context, intervention content and procedures

The intervention took place in the Department of English Language and Translation at King Khalid University. The curriculum includes teaching English language, literature and translation courses over four years. In the first year (levels 1 and 2), students learn reading, writing, speaking, and listening skills, along with Grammar 1 and 2 elective courses. In level 3, they take introductory courses in literature, linguistics, and translation, as well as Grammar 3. By level 4, students study Grammar 4 and courses in linguistics, literature, and translation. The intervention specifically targeted students enrolled in the Grammar 4 course.

In the Grammar 4 course, both the experimental and comparison groups attended a mandatory 80-minute session each week. For the intervention, the experimental group had two additional weekly sessions, and participants received two extra marks from the grammar course teacher for their time and effort. These extra sessions were scheduled by the department at times convenient for the participants. The comparison group did not receive any instruction on genericity but was awarded two extra marks for attending all pre- and post-test sessions. During the data collection period, both groups covered topics in Grammar 4, including coordinating conjunctions, adverb clauses, adverbial phrases, connectives that express cause and effect, contrast and condition, and conditional sentences.

As for the experiment group, in addition to receiving the same teaching content as the comparison group, the additional intervention sessions included the following content: The first week was assigned to revise definiteness and specificity as they were already presented to the sample in previous grammar courses, and the other 7 sessions were devoted to teaching characterising and kind-denoting generics. Each week includes two 40-minute sessions: the first was for instruction, and the second was a seminar session where the learners practised genericity. The researcher conducted the teaching for the experimental group during the additional intervention sessions. Both the experimental and comparison groups were taught by grammar teachers assigned by the institution. However, to eliminate the effect of having different teachers on the content of the intervention, the researcher attended with the teacher during the eight weeks to have a fuller understanding of what teaching was received by the comparison group and to ensure that genericity was not taught to the comparison group. For the intervention sessions, the teaching was provided in English, and the learners' L1 was used when comparison to L1 occurred using Arabic examples.

The content and order of information presented in this intervention are based on the cline of difficulty for the acquisition of English generics by Arabic-speaking learners, as discussed in Section 5.7. As for the what was taught and how as well as what activities were used in this intervention, the following details the procedures and activities used in each session. Generally speaking, the instruction was delivered over eight weeks, with two 40-minute sessions each week. The first session focused on teaching, while the second was dedicated to practice. Each teaching session was structured into four stages: the warm-up stage, the focus on meaning stage, the focus on linguistic forms stage, and the closing stage. Below is a description of the procedures followed in each stage.

In the warm-up stage, the objective was to introduce the session's aims and focus. To achieve this, the teacher employed various activities such as picture description, reading-based discussions, reviews of the previous session, or video-based discussions.

During the second stage, the emphasis was on linguistic meaning. Here, the teacher used example-based discussions to explain the characteristics of generic interpretation and how to differentiate between generic and non-generic meanings using diagnostic tests for each generic condition. This discussion typically involved the whole class, with learners examining examples and answering questions to present the meaning deductively.

In the third stage, the discussion shifted to the forms of noun phrases (NPs). Examples and form–meaning mappings were clearly and explicitly presented on the board, comparing English to Modern Standard Arabic (MSA) mappings for the condition discussed in the session. Learners were also asked to provide examples for discussion in terms of form–meaning mappings.

In the final stage, the teacher used wrap-up activities to summarise the session's content. Summaries were provided in various ways: in some sessions, students worked in groups to write a summary of the main points, followed by a whole-class discussion; in others, the teacher presented the summary on the board. Occasionally, students were asked to present their summaries on the board.

After a ten-minute break, the second session, primarily devoted to practice, commenced. In these practice sessions, learners worked in small groups or pairs on various activities that reinforced the form-meaning mappings covered in the first session. These activities included picture based production, reading with error correction and justification, differentiating between generic and non-generic sentences. The teacher monitored group work, provided feedback, and offered assistance as needed. At the end of each practice session, the teacher provided feedback to the whole class through discussion of groups' answer for each activity. The detailed descriptions of each the procedures of each session is provided in each week are summarised in Table 6.6 which offers a detailed description of the procedures, aims and the activities used in each session. Moreover, sample activities are exemplified in Appendix C.

Table 6-6 Intervention sessions' topics, aims, activities and procedures

Week 1 Title: Articles (Definiteness and specificity revision)	**Learning Aims:** Identify different types of articles, review definite and indefinite articles, practice giving and analysing examples. **Procedures and Content:** - Warm-up: Circle and underline articles in a given text, classroom discussion on identified noun phrases. - Focus on Linguistic Meaning: Explanation of the definite and indefinite/ specific and nonspecific meanings using examples-based discussions - Focus on Functional Forms: Form-meaning mappings on the board (articles for definite and indefinite meanings revision) - Wind-down: learners provide a summary of the lesson in small groups. - **Seminar Session:** Problem-solving activities focusing on definiteness, including story reading with article errors correction and article selection with justification.
Week 2 Title: Articles used in plural generics (Introduction to characterising generics)	**Learning Aims:** Identify noun phrases, characteristics of generic sentences, form of articles in plural generics, give examples. **Procedures and Content:** - Warm-up: Guided discussion with pictures of whales, analysis of example sentences based on the picture to introduce the concept of generic sentences. Example sentences: Whales breathe under the water; The whales are swimming Focus on Linguistic Meaning: The teacher discusses the truth and meaning of generic sentences, using examples like "Whales breathe under the water." Questions such as "Did whales breathe under the water in the past and will they in the future?" help students understand that generic sentences are timeless. Students provide their own examples of generalizations for further discussion. The teacher encourages critical thinking about the general truth of various statements despite exceptions. Focus on Functional Forms: Learners analyse sentence structures by underlining subjects and circling verbs in two groups: Group A (general statements) and Group B (specific instances). An awareness-raising

	discussion follows, focusing on subjects' plurality, definiteness, and specificity. Group A sentences denote general regularities, while Group B sentences describe specific situations. The use of the definite article "the" in Group B is examined, and learners learn that bare plurals with Ø articles are used for generalizations. Wind-down: Summary of generic statements and form-meaning mapping rules. - **Seminar Session:** Problem-solving activities focusing on properties of generic sentences, error correction, and picture-based sentence production.
Week 3 Title: Articles use in generics (Introduction to generics 2/Plural sentence-level generics)	**Learning Aims**: Identify characteristics of generic sentences, meaning of sentence-level generics, form of articles in plural generics, give examples. **Procedures and Content:** - Warm-up: Review of previous class examples, discussion on the difference in meaning. - Focus on Linguistic Meaning: Definition and characteristics of characterising generics as statements that express general truths about entities or situations. Using examples, the characteristic of being resistance to contextual restriction was discussed. For example, sentence like "Lions are dangerous" is analysed to understand their general, context-free nature. Summarize that generics allow exceptions, are true over time, and are not context-bound. Focus on Functional Forms: Sentence Analysis: Underline subjects and circle verbs in two groups: Group A (general statements) and Group B (specific instances). Group A denotes generalities; Group B describes specifics. Examine the use of "the" in Group B and bare plurals with Ø articles for generalizations. Discuss differences in form-meaning mapping between English and MSA. - Wind-down: Summary of generic sentences and form-meaning mapping rules. - **Seminar Session:** Problem-solving activities focusing on properties of generic sentences, error correction, and article selection with justification.

Week 4 Articles used in generics (characterising singular generics)	**Learning Aims:** Review characteristics of generic sentences, identify meaning of sentence-level generics, form of articles in singular generics, give examples. **Procedures and Content:** -Warm-up: Matching activity to introduce singular generics Group A (Meanings): a. Gifts in general. b. One gift, but not specific. c. A definite and specific gift. Group B (Sentences): I received a gift. The gift was very thoughtful. It is better to give and receive gifts. Using "Gifts make people happy" , In pairs, students test if this sentence is true regardless of time/context and if it allows exceptions to review previous session content. - Focus on Linguistic Meaning: The teacher reviews characterising generics, which express general truths. Students convert plural subjects to singular ones in given examples (e.g., "Koalas sleep up to 22 hours daily" becomes "A koala sleeps up to 22 hours daily"). The class discusses whether these sentences provide generalizations or specific episodes, confirming that generic sentences with singular NPs allow exceptions, are timeless, and context-free. - Focus on Functional Forms: Students receive a worksheet with sentences to change plural subjects to singular ones. The teacher guides a discussion on the differences between general (Group A) and specific (Group B) sentences, focusing on the use of articles ("a" for generics, "the" for specifics). The discussion includes comparisons with Modern Standard Arabic (MSA), reinforcing that singular NPs in generics use the article "a." - Wind-down: Summary of singular and plural generic sentences. - **Seminar Session:** Problem-solving activities focusing on picture and sentence matching, error correction, and article selection with justification.

Week 5 Revision	**Procedures and Content:** - Activities: Context followed by picture selection, guided linguistic production using short story retelling with peer feedback, error correction with justification.
Week 6 Title: Articles used in generics (kind plural generics)	**Learning Aims:** Discuss characteristics of kind-referring NP, identify meaning of NP-level generics, form of articles in plural kind-referring NP, give examples. **Procedures and Content:** - Warm-up: Video clip about Dinosaurs, analysis of example sentence. Focus on Linguistic Meaning: The teacher defines kind-referring noun phrases (NPs) and explores their characteristics. Using examples like "Lions are predatory cats" (general kind) and "The lions escaped from the zoo yesterday" (specific group), the class discusses the different references of "lions." This helps students understand that kind-referring NPs can denote either natural kinds or specific instances, depending on context. The teacher explains that kind-referring NPs are identified by kind-denoting predicates, such as "Lions live in Africa" or "Cell phones were invented in 1973 by Martin Cooper," which indicate general truths about the kind rather than specific objects. Focus on Functional Forms: Students analyse sentences in a worksheet, identifying subjects and verbs. They work in groups to determine grammatical characteristics and meanings, concluding that bare plurals denote kinds in English, while definite plurals refer to specific objects. In Arabic, the definite generic article "al" is used for plural kind-denoting NPs. - Wind-down: Summary of kind-denoting plural NPs. - **Seminar Session:** Problem-solving activities focusing on picture and sentence matching, error correction, and article selection with justification.
Week 7 Title: Articles used in generics (kind singular generics	**Learning Aims**: Discuss characteristics of kind-referring NP, identify meaning of kind generics, form of articles in singular kind-referring NP, give examples. **Procedures and Content:**

	- Warm-up: Review of kind-referring NPs, examples provided by students. - Focus on Linguistic Meaning: The teacher introduces kind generics using examples like "Elephants are large mammals" and "The elephants were seen at the watering hole." Students change plural NPs to singular ones, producing sentences such as "The elephant is a large mammal." The class discusses the references of "elephant" in each example, distinguishing between general kinds and specific instances. The teacher explains that kind-referring NPs denote general truths and can be identified by kind-denoting predicates, such as "Elephants live in Africa" or "Computers were invented in the 20th century." These predicates indicate general truths about the kind rather than specific objects. Focus on Functional Forms: Students work on a worksheet to analyze sentences, identifying subjects and verbs. They work in groups to determine grammatical characteristics and meanings, concluding that bare plurals and "the + singular NP" denote kinds in English, while definite plurals refer to specific objects. In Arabic, the definite generic article "al" is used for both singular and plural kind-denoting NPs. The teacher facilitates the rule deduction process and summarizes the findings on the board with the help of the learners. - Wind-down: Summary of kind-denoting singular NPs form-meaning mappings with examples. - **Seminar Session**: Problem-solving activities focusing on picture based production, error correction, and article selection with justification.
Week 8 **Revision**	**Session 1: Review of Generics** This session focuses on revising sentence-level generics (both singular and plural) and NP-level generics. Students will work in small groups to fill in meaning-form mapping trees with examples, summarizing the rules of form mapping. The teacher will monitor the process and facilitate group discussions. Each group will share their mappings with the whole class to receive feedback from both the teacher and peers. The mappings will include: Bare Plurals: Generic sentences and plural kind-denoting NPs.

	a + N: Singular NP in generic sentences/indefinite NP. The + N: Singular kind-denoting NP/singular definite NP. **Seminar Session:** The seminar will include the following activities: 1. Context and Picture Selection: Students will select pictures based on given contexts. 2. Error Correction with Justifications: Students will correct errors in sentences and provide justifications for their corrections.

6.7 Data analysis

Following data collection, the answers to the background questionnaire and the proficiency test scores were input into an Excel spreadsheet. Then, a separate tab on the same spreadsheet was assigned for each group. After that, the researcher manually inserted the data from the three tasks. The inserted data was validated using data validation methods in Excel to ensure that the insertion process was correct. In the case of outliers, the researcher rechecked that the value was the same in the paper format. The same procedures were followed in each testing session. The data was cleaned and sorted, and the participants were assigned numbers as an anonymizing procedure, and then were extracted into a CSV file and exported to the R statistical software package (R Core Team, 2023) for analysis. A detailed description of the data analysis procedure for each task is provided prior to presenting the results of each task in the following chapter.

6.8 Summary

In sum, I have elaborated on the experimental design of the current study in this chapter. I began with a description of the research design and participants. A thorough explanation of each of the tasks and the results from piloting them followed. Then, I detailed the data collection procedures. Next, the teaching context, content and procedures of the intervention were outlined. As indicated throughout this chapter, several appendices are related to the experimental design described in this chapter, which I briefly summarise here for convenience. Appendix B presents the EWPT, the AJT and the FCT task. Appendix E includes the ethical approval. Finally, Appendix C presents examples of the classroom activities. In the following chapter, I present a detailed data analysis.

Chapter 7 Results

7.1 Introduction

In this chapter, I present the results of the L2 learners intervention study, as outlined in the previous chapter. Before statistically describing and analysing the data from this study, I briefly review the aim and questions to maintain narrative coherence.

As a reminder, this thesis aims to investigate what is easy and difficult in acquiring the English generic form-meaning mapping by Arabic-speaking learners and the impact of intervention based on the results of acquisition research on the acquisition of this form-meaning mappings. It evaluated the accuracy of Arabic-speaking learners of English in mapping the forms (NP-types) to the generic meaning, considering the cross-linguistic differences and the predictions made under the Feature Reassembly Hypothesis (FRH) and Bottleneck Hypothesis (BH) (see Section 5.7). To this end, this thesis addressed two research questions:

RQ1: Do intermediate Arabic-speaking English learners demonstrate a target-like mapping of English articles to kind and characterising generic meanings? If not, what is the difficulty and how can it be accounted for in terms of FRH and BH?

RQ2: After being exposed to 8-week acquisition-informed explicit instruction on English genericity form-meaning mappings, can the experimental group show improvement in form-meaning mappings in kind and characterising generic meanings compared to the comparison group? If any, is this improvement retained three months after instruction?

To answer the research questions, this study used three tasks in a pre-test, post-test, and delayed post-test design, including an elicited written production task, an acceptability judgment task, and a forced-choice task. This chapter presents the results concerning the research questions. Section 7.2 presents the results of the main tasks in the pre-test, i.e., before receiving teaching about genericity (RQ1). Then, Section 7.3 fully displays the results of the post-test main tasks with a focus on comparing the L2 learners' comparison and experimental groups, i.e., the results concerning the impact of the intervention, and reports the results of the delayed post-test task by task (RQ2).

7.2 Pre-intervention results

In this section, the main goal is to understand the acquisition problems L2 learners have before teaching them about genericity (RQ1). For each pre-test task, descriptive and inferential statistical results are presented. For each task, I followed the Comparative Logic (Dominguez & Arche, 2021) method, which (1) describes, (2) analyses, and (3) explains interlanguage grammar by using controls and comparisons between learner grammars and a baseline on purely scientific grounds. I used the performance of the native controls as a standard for the analysis. Since the pre-test tasks aim to provide a snapshot of second language acquisition before the intervention for all L2 participants, the sample of L2 learners in this section included all L2 learners who participated in the study (*n* = 64).

7.2.1 Elicited written production task results.

The EWPT aimed to understand how L2 learners associate generic meaning with various NPs in elicited written production. To recap, the task involved presenting a context through a brief conversation and a picture cue, followed by a wh-question designed to elicit the production of generic form-meaning mappings. Table 7.1 summarises the task's generic contexts and conditions, presenting the appropriate forms based on empirical data from native speakers (as discussed in Chapter 5) and insights from the semantic literature on English genericity marking.

Table 7-1 Target-like articles usage in both generic contexts based on the native speakers' results and the semantic literature.

Meanings	**Test items/ correct production**
Characterising generic meaning	Plural forms: In items (1) and (5), the target use is 'bare plural NP.' Singular forms In items (3) and (7), the target use is 'indefinite singular NP.'
Kind generic meaning	Plural forms In items (2) and (6), the target use is 'bare plural NP.' Singular forms In items (4) and (8), the target use is 'definite singular NP.'

The participant's responses to this task included nominal production in the form of short sentences in which the subject is the NP that reflects generic form-meaning mapping. In terms of coding, the decision was made to keep the data as informative as possible by coding the participants' responses as a nominal categorical variable, which includes all the possible NPs produced by the sample. The response variable levels include five NP types: bare plurals, bare singulars, definite singulars, indefinite singulars, and definite plurals. The 'Other NP' level was added to code unrelated production, such as using demonstratives, quantification, and phrases referring to the depicted item as ‘the item in the picture.' These levels were coded as unordered discrete categories of the response variable. In case missing data occurred, it was recorded using the notation ‘NA’ and was excluded from the analysis. It is crucial to clarify that the task tried to use singular and plural picture cues to encourage the production of both forms. However, when participants produced either the singular or the plural form in both generic contexts, it was considered target-like regardless of the picture cue. The rationale behind this approach was that both the singular and plural forms can accurately convey the intended meaning, rendering them both appropriate for this task. The production rates of the native control group, presented in the following section, supported this decision. For each participant in both groups, a total production score of the target-like forms was calculated based on how often they hit the target in each context. Then, a percentage accuracy score was calculated (total score multiplied by one hundred and divided by the total number of test items). For descriptive statistics, violin plots were used to illustrate participants' diversity in the total percentages. Bar graphs were used to compare the proportion of each NP produced per each generic condition. Then, a multinomial logistic regression model was used to estimate the probability of the production of the different NP types (categorical nominal variable with six levels) in the light of the generic conditions as the predictor variable using the function *multinom* in the *nnet* package (Venables & Ripley, 2002). There are many ways to run this model, but the choice of the *multinom* function was justified by the lower complexity of this method, as it does not require data reshaping (Hua et al., 2021). In all models, the reference level was the category ‘bare plurals.' The modelling structure was to start with the maximal model structure and then reduce the model using the *lrtest* function from the *lmtest* package (Hua et al., 2021). What follows are the descriptive and inferential statistical results of the baseline native speakers’ and the L2 learners' data.

7.2.1.1 Native Controls' Results

In terms of the descriptive statistics, the variability between participants' performance in each generic context was described by plotting the densities and mean of each participant's target-like form production total percentage (see Figure 7.1).

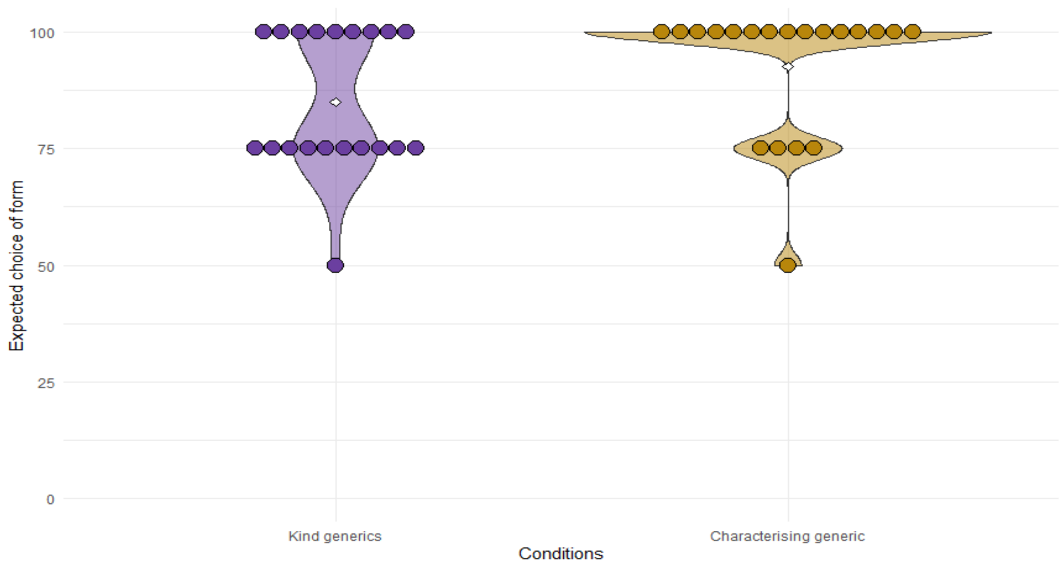


Figure 7.1 The percentage of target-like forms produced per participant by the native control group in the EWPT task. The diamond shape indicates the group's mean percentage.

This figure illustrates the percentage of the native control participants' production of expected forms based on Table 7.1. It demonstrates that the English native control group showed over 75% of expected production (bare plurals and definite singulars to refer to kind and used bare plurals and indefinite singulars in characterising generics). Only one out of twenty participants demonstrated significantly lower than 75% overall expected forms production in any given generic context. The native controls had a total target-like form production mean percentage of 93% (*SD*= 14.28) in characterising generics and (*M* = 84%, *SD* = 18.62) in kind generics, as indicated by the diamond shape in Figure 7.1, which revealed the elevated level of the native control participants in their production.

Moreover, the proportions of NP types produced by the native control group are visualised in Figure 7.2.

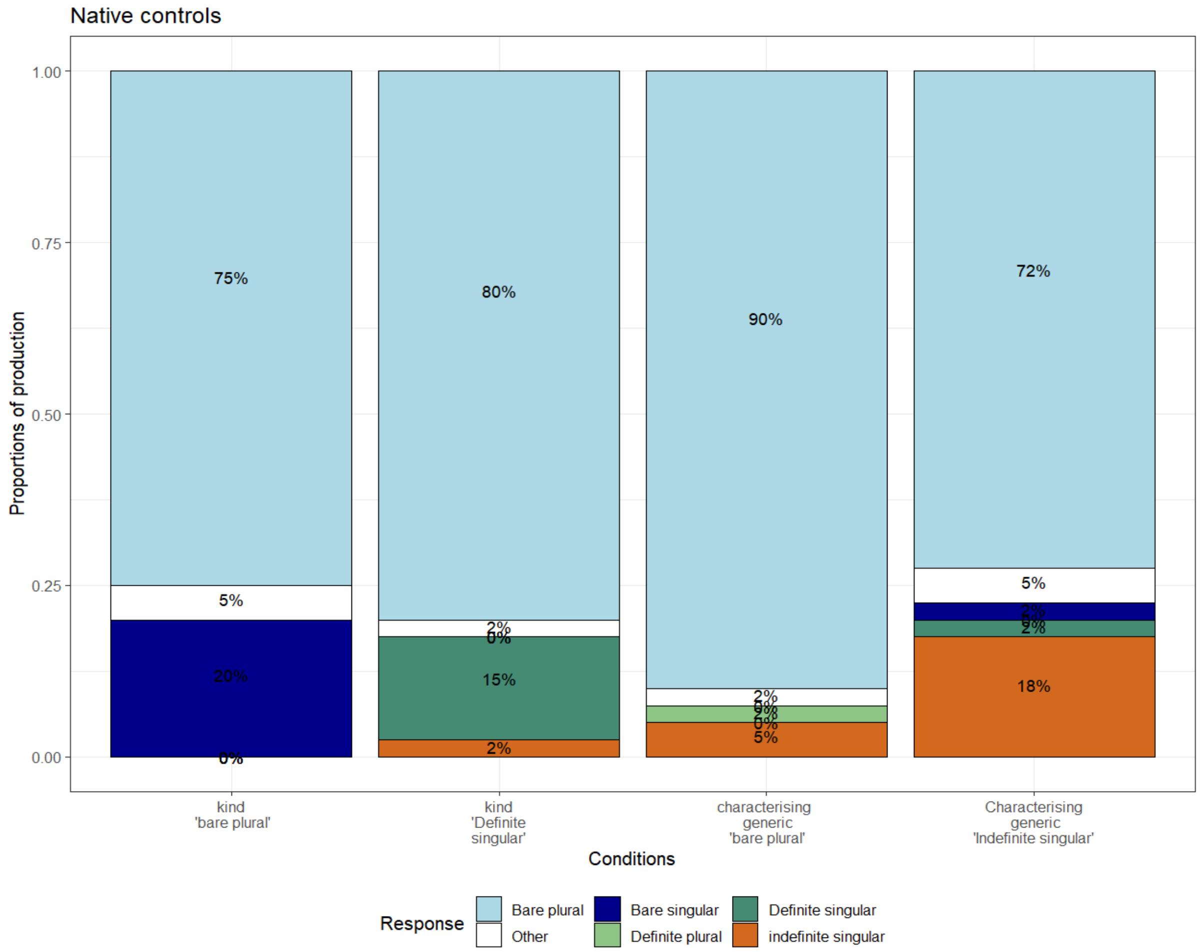


Figure 7.2 Proportions of NP-type production in each generic condition in the EWPT by the native controls.

Figure 7.2 displays the proportions of the native control group NP production in characterising and kind generic meanings (singular and plural conditions). It reveals that the native baseline produced predicted acceptable plural and singular forms in each context. The figure shows native speakers produced predicted plural and singular forms interchangeably across various conditions, with the highest proportions of bare plurals. Notably, bare plurals were particularly prominent in the native speakers' production, accounting for over 70% of their output in all conditions. In characterising generic test items that aim to generate singular forms through using singular items in the picture cues, 72% of the production consisted of bare plurals, while only 18 % featured indefinite singular. When characterising generic test items included plural forms picture cues, 90% of the production consisted of bare plurals, while only 5% used indefinite singulars.

Turning to kind generics, when kind generic test items had singular form picture cues, 80% of the production consisted of bare plurals, while only 15% used definite singulars. In the test items that used the plural form picture cues, 75% of the production used bare plurals, whereas 20% employed bare singulars. This striking result is explained by the test item (6) (lavender flowers\ lavenders are popular in the Netherlands, contributing 20% of the unexpected production of bare singulars in this condition. Because the word lavender can be a count and mass noun, some participants employed bare singulars, which is acceptable with mass nouns in kind generics. I reviewed whether removing this item would lead to better production. Doing so raised the production rate of bare plurals onto kind generic meaning from 75% to 95% by the native controls. Using mass NPs with generic meaning is beyond this study's aims, and including this item can affect the results. Therefore, it was removed from all subsequent analysis. Considering this decision, the proportions of the native benchmark group after removing this test item are presented in Figure 7.3.

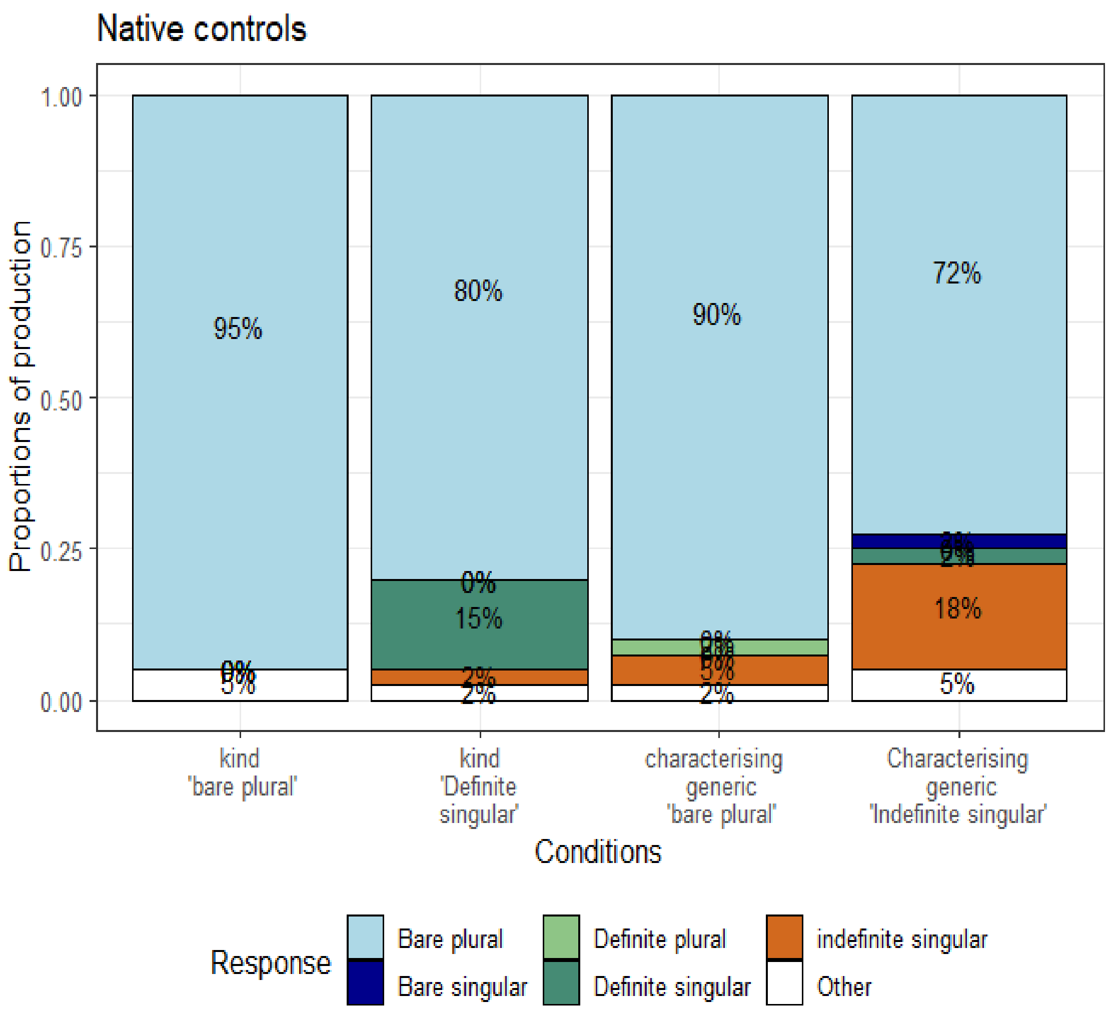


Figure 7.3 Production proportions per generic meaning by the native control group in the EWPT.

Overall, the descriptive statistics of the native controls revealed that bare plurals were the most frequent production choice in all generic conditions. Singular forms were not frequent in the native baseline data. However, these results are merely descriptives and require further statistical evidence.

A multinomial mixed effect model was constructed based on the native controls' production to assess the probability of producing the correct NP type in generic contexts. In the model, the dependent variable was the NP type of the participant's response with the six levels, including different NP types and the generic meaning contexts as the fixed effect. The model goodness of fit was measured by the likelihood ratio test using the *lrtest* function, which showed that this model was significantly better than the null model (*LR* (5) =27.58, *p* =.02), supporting the validity of this model in providing a better prediction of production than the null model. Table 7.2 summarises the model results.

Table 7-2 Native control group's response to the production task multinomial logistic regression model results.

predictor	**Bare singular (2)**	**Definite plural (3)**	**Definite singular (4)**	**Indefinite singular (5)**	**others (6)**
Condition [kind definite singular]	-2.436*** (0.00000)	-2.297 (383.632)	20.987*** (0.450)	22.552*** (0.790)	-0.521 (1.444)
Condition [Characterising bare plural]	-3.047	14.463 (1,411.862)	-4.905	23.127*** (0.610)	0.639 (1.442)
Condition [Characterising indefinite singular]	15.694*** (0.509)	-1.800 (658.887)	19.294*** (0.694)	24.596*** (0.444)	0.270 (1.260)
Constant	-19.061*** (0.509)	-18.047 (1,411.861)	-22.661*** (0.370)	-26.017*** (0.329)	-2.945*** (1.026)

Note:*p<0.1; **p<0.05; ***p<0.01, Akaike Inf. Crit. 203.859, Formula: multinom(Response~ Condition, data=ProN2no, Hess=T); Observations, 140; R2 / R2 adjusted0.144 / 0.134.

Table 7.2 summarises the model estimates and *p*-values. It shows that the estimated log odds of producing bare plurals (the reference level) was significantly higher than all NP types, as indicated by the negative estimate ratio (*p* <.001 for all comparisons in the constant). Additional analysis used Tukey-corrected pairwise comparisons to compare the probability of producing the target-like forms in Table 7.1 to all the other forms per condition.

Considering characterising generics with bare plurals as the target response, the pairwise comparisons showed that the probability of producing a bare plural response was significantly higher than producing a definite plural NP (F (20) = 192.15, p <.001), and bare singular as well (F (20) = 395.99, p <.001). It also was significantly higher than the probability of producing a predicted indefinite singular NP (F (20) = 89.70, p<.001). When an indefinite singular NP was the target form in characterising generics, the pairwise comparisons showed that the probability of producing a bare plurals response was significantly higher than producing the predicted from indefinite singular as (F (20) =20.25, p <.001). The native baseline production of indefinite singulars was nonetheless significantly higher than bare singulars and definite singulars production (F (20) = 5.07, p =.03 for both comparisons). This result indicates that the baseline group successfully produced both bare plural and indefinite singular significantly higher than all other NPs in characterising generics.

Turning to the pairwise comparison between 'bare plural' and all other responses in kind generic plural conditions, the native baseline only produced bare plurals and 'others' in this condition. The results showed that the probability of producing bare plurals was significantly higher than the probability of producing 'other' (F (20) = 85.27, p< 0.001). In singular-form production with kind generic meaning, the baseline production includes bare plural, definite singular, and indefinite singular NPs. The pairwise comparison showed that the production of bare plurals was significantly higher than the production of the correct definite singular NP (F (20) = 32.04, p<.001), and the indefinite singular NP (F (20) = 107.70, p<.001). Additionally, the probability of producing a predicted definite singular NP with kind generic meaning was significantly higher than that of producing a non-target-like indefinite singular NP (F (20) = 3.92, p =.006). This result indicates that the baseline group successfully produced both bare plural and definite singular significantly higher than all other NPs in kind generics.

Considering the predictor generic context, the pairwise comparisons showed that the conditions as a predictor did not have a notable effect on the production of bare plural (F (20) =.54, p =.47) for using bare plural form in kind generics as compared to characterising generics. as shown in the context effect plot in Figure 7.4.

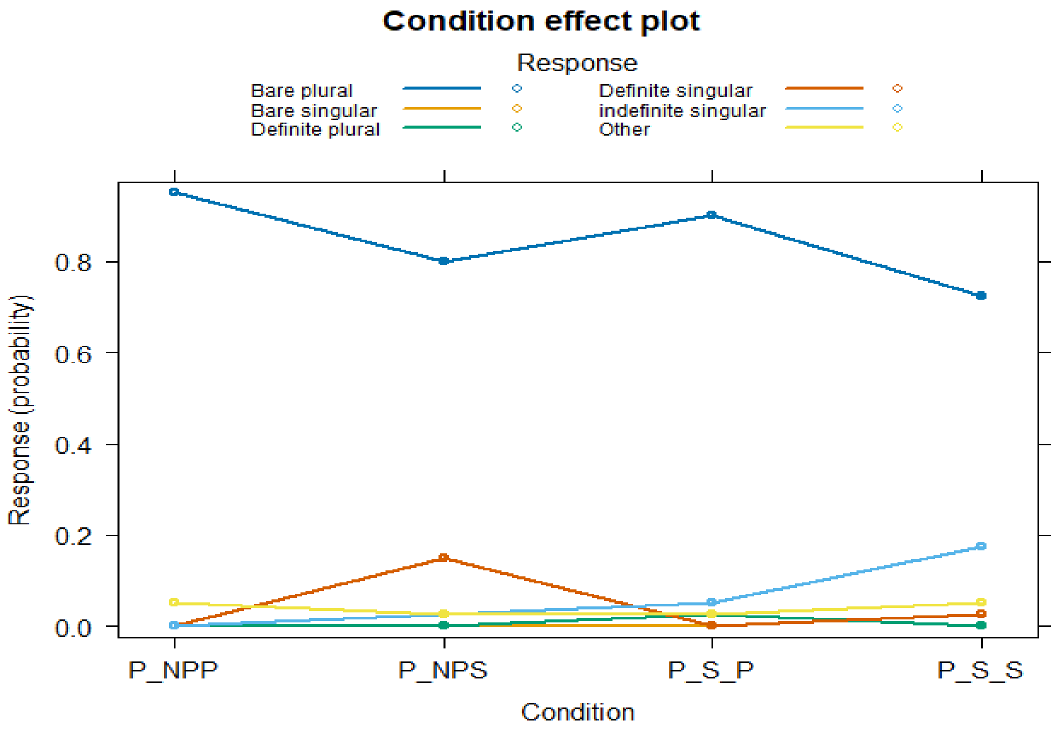


Figure 7.4 Native controls context effect plot. [9]

To recap, the native control group's response to the production task included producing bare plural NPs more significantly than the predicted singular form in all generic conditions. Nonetheless, they produced the expected singular form at a significantly higher rate than the unexpected alternatives in each singular condition. This production provides the baseline for comparing the L2 learners' production and validates the suitability of the production task to elicit generic form-meaning mappings.

7.2.1.2 L2 Learners' Results

Before instructing the L2 learners on genericity, it was necessary to analyse their responses to the pre-test EWPT to understand their current acquisition of this property, if any. In terms of description, the total score of accuracy in target-like form production was calculated for each participant, and its percentage was visualised using the densities to compare the range of L2

[9] P_NPP refers to plural forms in kind generics, P_NPS refers to the singular form in kind generics, P_S_P refers to the plural form in characterising generics, and P_S_S refers to the singular form in characterising generics.

learners' production accuracy in the two generic contexts, as Figure 7.5 shows.

Figure 7.5 L2 learners' total accuracy percentage in target-like form production per participant in both generic contexts in the pre-test production task. The diamond shape indicates the group's mean accuracy percentage.

In contrast to the native controls' production, L2 learners' production showed more variability in the percentage of their accuracy in target-like form production, as expected. Furthermore, more learners had an accuracy rate of 50% or less in kind generics (59/64)than in characterising generics (52/64). The number of learners whose production in the characterising generics was target-like was 12/64. In the kind generics, only 5/64 of the L2 learners were like the native benchmark in total percentages of the expected form production. It appears that the L2 learners' production reflects acquisition difficulties, as indicated by the low mean percentage of correct form production by L2 learners (*M* = 36, *SD* = 28.74 for characterising generics and *M* = 33, *SD* = 23.59 for kind generics) as compared to the native benchmark (*M* = 93, *SD* = 14.28 in characterising generics and *M* = 84, *SD* = 18.62 in the kind generics).

The proportion of NP-types per generic condition was calculated and visualised in Figure 7.6.

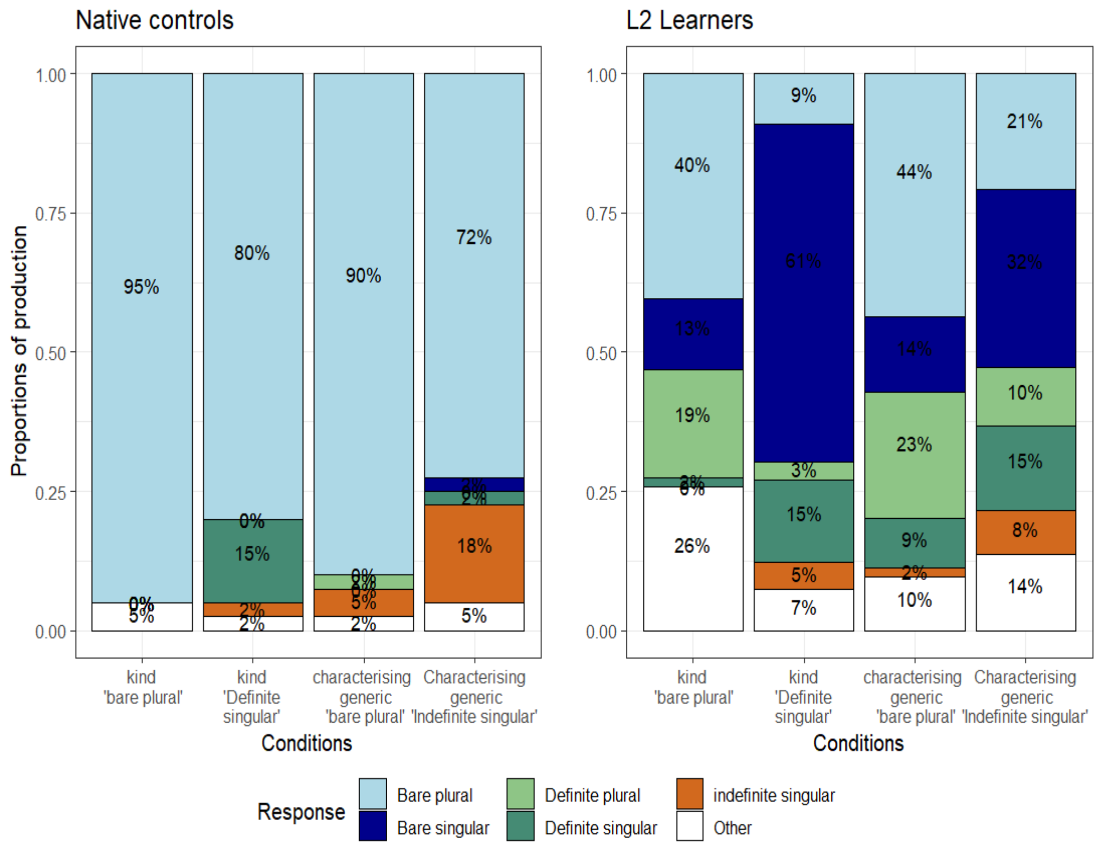


Figure 7.6 L2 learners produced NP forms compared to the native benchmark in each condition on the pre-test EWPT.

Figure 7.6 displays the proportion of the production of all NP forms in each generic meaning by L2 learners compared to the benchmark group. It demonstrates that L2 learners showed low proportions of the bare plurals production in kind generics and characterising generics, with 40% and 44%, respectively, unlike the native control group, which produced about 95% and 90% of this form. In characterising generics, the expected indefinite singulars were the least produced, accounting for 8% of the production. Surprisingly, the proportion of the expected definite singulars with kind generic meaning was low, accounting for 15% of the production in the relevant condition. This result shows that while the proportions of the native control group's choices of expected NP forms (plural and singular forms proportions are merged for each condition) in production were above 90% in all conditions, the L2 learners' proportions of expected NP forms in production did not exceed 45% in any condition.

Turning to the proportions of the unpredicted forms produced, a close inspection of Figure 7.6 showed that bare singulars were the dominant error in mapping the predicted singular form for each kind and characterising generics (61% of the production in kind generics with definite singular condition and 32% of NPs in characterising generics with indefinite singular). Conversely, the L1 form (definite plurals) was the most frequent erroneous production in the two conditions that targeted bare plural production (19% in kind generics and 23% in characterising generics). The descriptive results revealed that the L2 learners produced varied form-meaning mappings in both generic meanings. These findings, however, are merely descriptive and inferential statistics of the L2 learners' responses is worth.

To ensure that the L2 learner's low production of the predicted forms described above was significant when compared to the baseline controls, a multinomial logistic regression model was used, with response as the dependent variable and condition and group as predictor variables. The model goodness of fit was measured, and the results of a Chi-squared test showed that this model was a significantly better fit than the null model (X^2 =140.5, *df*= 5, *p* <.001), supporting the validity of this model for providing a better prediction of production than the null model. The results of this model showed that the native control baseline's production of bare plural was significantly higher than the production of all NPs by the L2 learners (*p* <.05; the entire model is in Appendix D). Tukey-corrected pairwise comparison showed that the L2 learners' production of bare plurals was significantly lower than the native baseline in kind generics (*F* (40) =92.5, *p* < .001) and characterising generics (*F* (40) = 155.6, *p* <.001). As for the singular forms in each generic meaning, only relevant pairwise comparisons are reported here. Considering kind generics with the definite singular as the accurate form, the L2 learners' production proportion was not significantly different from the baseline (*F* (40) =1.53, *p* =.22). However, the L2 learners produced the ungrammatical bare singulars at a significantly higher rate than the baseline did in this condition (*F* (40) =155.5, *p*<.001). Turning to the use of the indefinite singulars with characterising generics, L2 learners' production of the indefinite singular was not significantly different than the native baseline (*F* (40) = 1.16, *p* =.29). However, the L2 learners produced the unpredicted bare singular and definite singular significantly higher than the baseline in this condition (*F* (40) =60.20, *p* <.001 and *F* (40) =8.93, *p* <.001, respectively). Comparing L2 learners to the baseline showed that their production significantly differed from the target-like. To understand the L2 learners' production difficulty, within-group differences in the L2 learners' production were analysed further. The model dependent variable was the NP type of the response, and the predictors included the generic conditions. The model's goodness of fit showed that this model was significantly better than the null model (X^2 =141.94, *df*= 5, *p* <.001).

Table 7-3 The L2 learner group's response to the production task multinomial logistic regression model results

predictor	**Bare singular (2)**	**Definite plural (3)**	**Definite singular (4)**	**Indefinite singular (5)**	**others (6)**
Condition [kind definite singular]	3.046*** (0.519)	−0.278 (0.681)	3.711*** (1.089)	15.413*** (0.431)	0.246 (0.552)
Condition [Characterising bare plural]	−0.016 (0.492)	0.077 (0.421)	1.628 (1.072)	12.723*** (0.563)	−1.058** (0.452)
Condition [Characterising indefinite singular]	1.570*** (0.478)	0.041 (0.489)	2.905*** (1.064)	15.063*** (0.356)	0.021 (0.447)
Constant	−1.139*** (0.406)	−0.734** (0.351)	−3.219*** (1.020)	−16.019*** (0.239)	−0.446 (0.320)

Note: *p<0.1; **p<0.05; ***p<0.01, Akaike Inf. Crit. 1,303.532, Formula: multinom (Response~ Condition, data=L2, Hess=T); Observations, 433; R2 / R2 adjusted 0.101 / 0.100.

Table 7.3 displays the estimated log-odds and standard errors of the L2 learners' use of different NPs in the EWPT pre-test. The result suggests that the probability of producing bare plurals with kind generic plurals was not significantly different from that of producing definite plurals in all other conditions (*p* >.05 in the shaded cells). Further, Tukey-corrected pairwise comparisons were used to assess the significance of the probability of producing all other NPs in each generic condition.

Starting with mapping both kind and characterising generic meanings onto bare plurals, Table 7.1 states that the predicted acceptable forms include bare plurals in both conditions. Tukey-corrected pairwise comparisons showed that the production of bare plural with both conditions was significantly higher than the production of all other NPs in each condition, as shown in Table 7.4.

Table 7-4 Tukey-corrected pairwise comparisons between the correct form ‘bare plural’ and all NPs in both plural kind and characterising generics by the L2 learners

Contrast	**Kind generics**	**Characterising generics**
Bare plural - definite plural	F (20) = 4.93, p =.03	F (20) = 8.83, p =.007
Bare plural -definite singular	F (20) = 34.47, p <.001	F (20) =36.92. p <.001*
Bare plural - indefinite singular	F (20) = 41.89, p <*.001*	F (20) = 79.08, p <*.001**
Bare plural - bare singular	F (20) = 10.20, p= .004	F (20) = 22.83, p =.001*

Note: * Significant at .05 level.

Although L2 learners’ production of the correct bare plural form was significantly higher than all other forms, their production of definite plural NPs was significantly higher than the possible correct singular form, definite singulars in kind generics (F (20) = −10.95, p =.003) and the indefinite singulars form in characterising generics (F (20) = −27.53, p <.001) which reflects a high production of the L1 definite plural.

Turning to the L2 learners' use of the correct singular form, indefinite singulars, with characterising generic meaning, the pairwise comparisons showed that the probability of producing the indefinite singular was significantly lower than that of producing the ungrammatical bare singular in this condition (F (20) = 21.03, p <.001). Moreover, the production of indefinite singular NPs was not significantly different from that of L1 forms when compared to definite plural (F (20) = 0.40, p = .54) and definite singular (F (20) = 2.86, p = .12). When considering the ungrammatical production, use of bare singulars was the most frequent error. These results suggest that L2 learners found it challenging to produce indefinite singular NPs, as they produced high proportions of bare singulars and a considerable proportion of L1 forms.

Considering mapping the kind generic meaning onto the definite singular, the pairwise comparisons showed that the probability of producing bare singulars was significantly higher than the probability of producing the predicted correct NP (F (20) = 47.30, p < .001). The production of definite singulars in this condition was significantly higher than the production of the other ungrammatical forms, including definite plurals (F (20) = 9.611, p = .005) and indefinite singulars (F (20) = 6.31, p = .02). In sum, the probability of producing grammatical definite singulars was significantly lower than the probability of producing bare singulars and higher than the probability of producing indefinite singulars. These results suggested L2 learners faced challenges producing the expected singular form in kind generic meaning.

The predictor generic conditions showed how much the log odds changed when the generic condition existed. A significant effect of condition was observed in the production of bare singular NP responses, in which probability significantly increased in the singular kind generic context compared to the reference level plural kind generics, as shown in the context effect plot below. Also, the probability of producing bare plurals was significantly affected by being in both kind and characterising generic conditions that used plural picture cues.

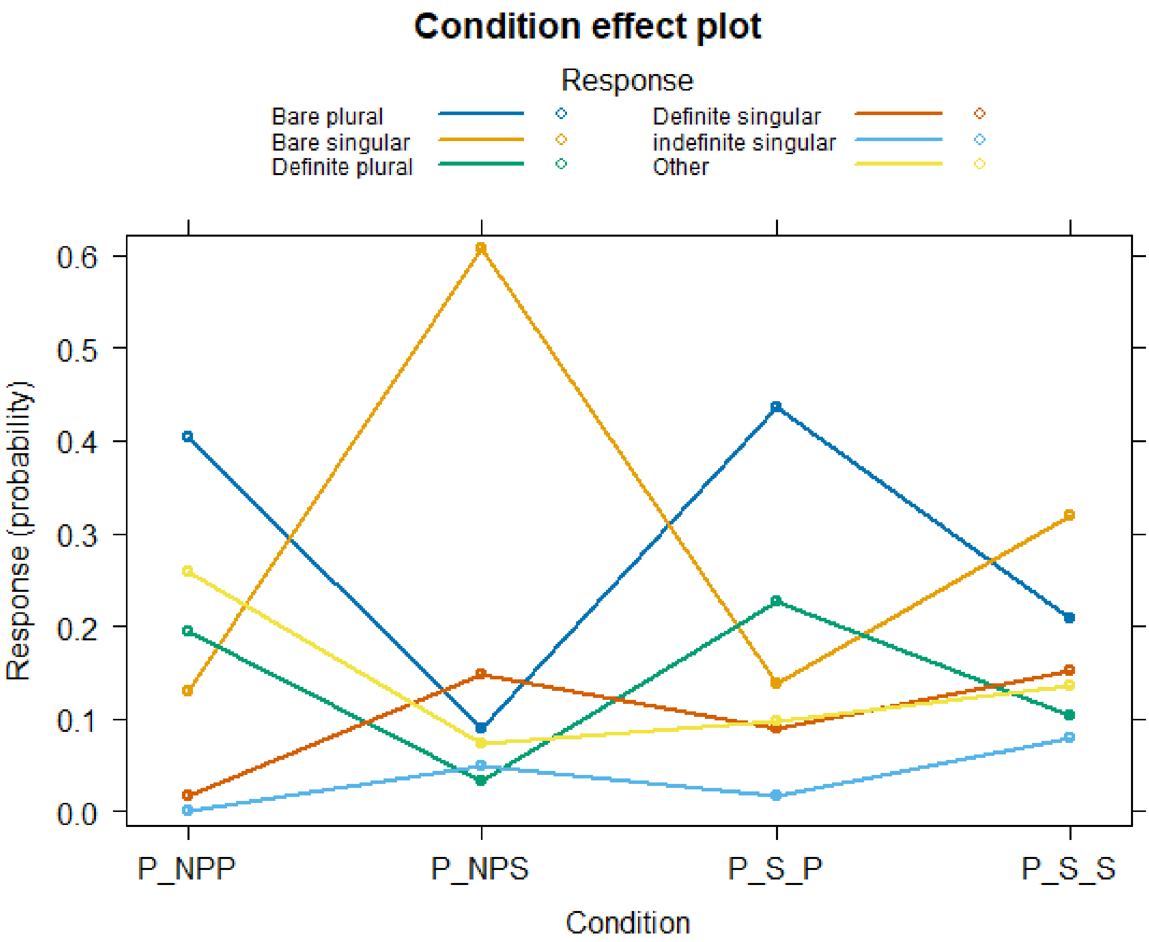


Figure 7.7 Context effect plot for L2 learners' production task pre-test. [10]

Having presented how the L2 learners performed in the production task and how they differed from the native benchmark, recall that after the pre-test, the L2 learners were first divided into two groups (experimental (*n*=35) and comparison (*n*=29)) prior to the intervention. It was essential to ensure that L2 learners' groups were not different in their production of form-meaning mappings before starting the intervention to limit the variance in accuracy before the intervention from being a confounding factor that affects the results (Mackey & Gass, 2021).

[10] P_NPP refers to plural forms in kind generics, P_NPS refers to the singular form in kind generics, P_S_P refers to the plural form in characterising generics, and P_S_S refers to the singular form in characterising generics.

Therefore, the homogenous production by the L2 learner groups before the intervention was investigated.

Regarding by-participant accuracy, individual participants' accuracy rates are summarised in Figure 7.8.

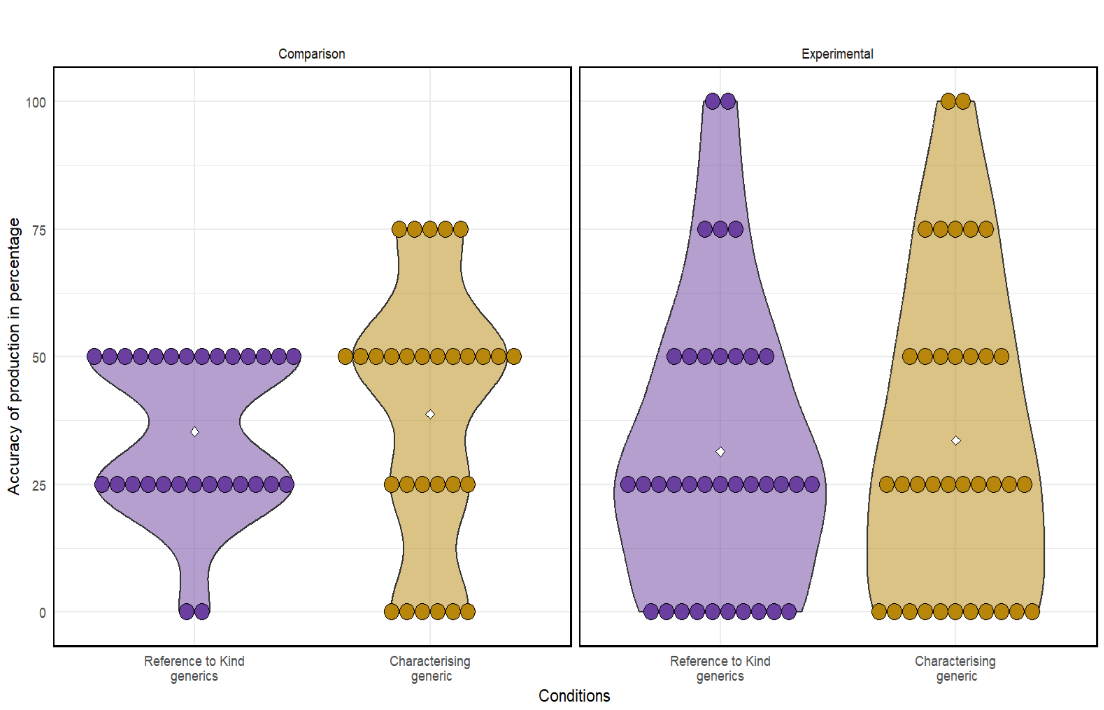


Figure 7.8 L2 learners' comparison and experimental groups' accuracy in the production task pre-test. The diamond shape illustrates the group mean percentage of accuracy.

Figure 7.8 shows the accuracy percentage for each participant and the mean accuracy percentage for the groups in each condition. This figure demonstrates that almost all the comparison group participants have 50% or less accuracy in both conditions, except for 5/29 participants, who had 75% accuracy in characterising generic conditions. As for the experimental group, few participants showed 75% or above accuracy, 5/35 in kind generics and 7/35 in the characterising generic condition. The rest of the experimental group showed either 50% or less accuracy. The group mean accuracy percentage for the comparison group was slightly higher than that of the experimental group in both conditions, as indicated by the diamond shape in Figure 7.8. The mean accuracy rates in the comparison group were 35.3% (*SD* = 15.69) for the kind generics, and 38.79% (*SD* = 25.5) for the characterising generics. In contrast, the experimental group's mean rates were 31.42% (*SD* = 28.66) for the kind generics and 33.57% (*SD* = 30.88) for the characterising generics.

Turning to the differences between production proportions, Figure 7.9 visualises the proportions of producing each NP type in each generic context for both groups. This descriptive analysis revealed that the two groups slightly varied in their proportions of producing expected forms in different conditions.

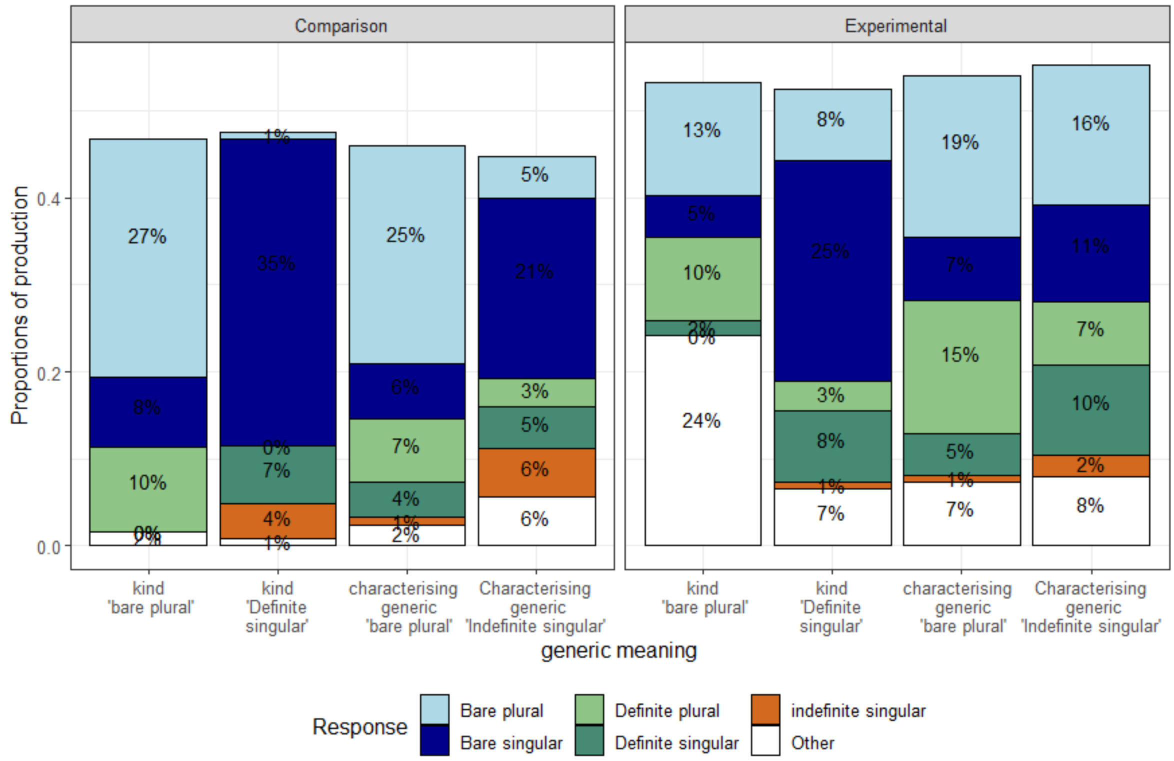


Figure 7.9 The proportion of different responses produced by the comparison and experimental groups of L2 learners in generic meanings in the production task pre-test.

A closer look at the proportions of accuracy in each generic context revealed that the proportions of correct production of bare plurals in characterising and kind generics by the comparison group were 25% and 27%, respectively, which were higher than the proportion of the production by the experimental group in the same conditions (16% in plural characterising and 13% in kind generics). As for singular NP forms in both generic contexts, the experimental group produced indefinite singulars in 2% of the characterising generic meaning condition tokens and definite singulars in 8% of the kind generic meaning condition tokens. In the comparison group, indefinite singulars were produced in 6% of the characterising generics and definite singulars in 7% of the kind generics. This result suggests that the two groups were somewhat alike in their proportions of producing the expected forms in each generic context.

Moving on now to ensure that the variance in the production of expected form proportions between the groups was not significant and that the groups were homogenous in producing the expected NP types in each generic condition before the intervention. A multinomial logistic model was applied to check the difference in the probability of producing each NP between the groups. The model structure included NP response type as the dependent variable and the interaction between contexts and groups as the fixed effect, using the same procedure to analyse the native controls' data. The fitted model showed a significantly better fit than the null model (X^2 = 67.42, *df* = 5, *p* <.001, *p* <.05). The model results are shown in Appendix D. Post hoc Tukey-corrected pairwise comparisons were used to compare each NP per each condition. Related comparisons are presented in Table 1.5, and complete pairwise comparisons are in Appendix D.

Table 7-5 Tukey-corrected pairwise comparisons of pre-test production by the experimental and comparison groups.

Contrast	**DF**	**F**	***P*.value**
Kind generics bare plural Comparison - Experimental	40	8.48	.006*
Characterising generics bare plural Comparison - Experimental	40	5.21	.03*
Kind generics Definite singular Comparison - Experimental	40	.08	.77
Characterising generics indefinite singular Comparison - Experimental	40	2.60	.11

Note: * significant at the .05 level of significance

Table 7.5 displays the correct form pairwise comparisons between the experimental and comparison groups in the pretest production task. The table showed that the difference between the groups was not statistically significant, and both groups were homogenous before the teaching intervention in using the definite singular and indefinite singular. However, the groups differed in using bare plurals, with the comparison group showing higher accuracy than the experimental group. This result will be taken into consideration when discussing post-instruction improvement. All other comparisons were insignificant in each condition except for using the bare singular with singular characterising and singular kind generics, as shown in the effect plot below.

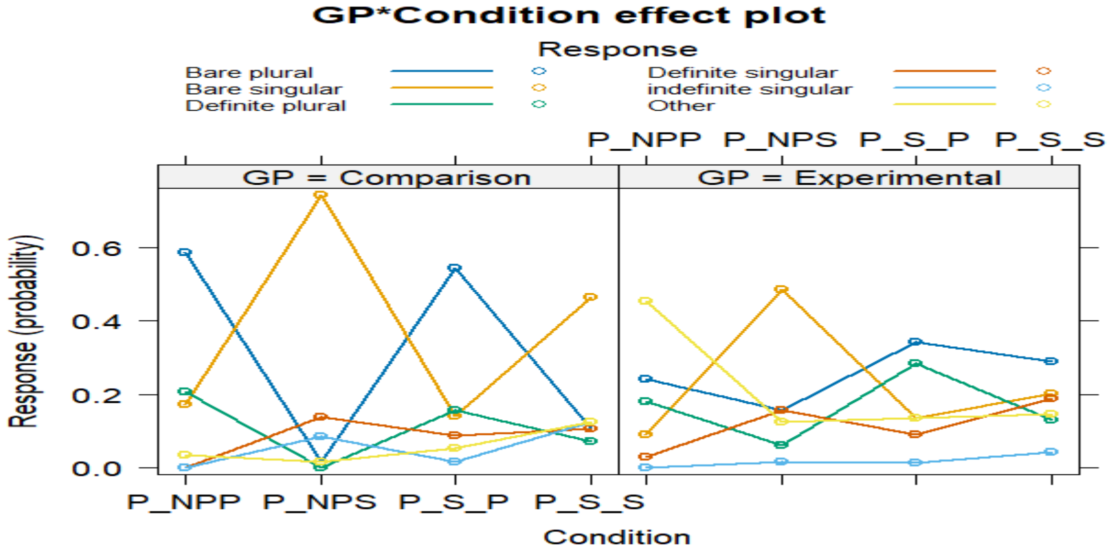


Figure 7.10 Effect plot of the interaction group*context between-groups multinomial logistic regression model[11].

At this point, I can conclude that the groups of L2 learners are not significantly different in their production proportions of NP-type production in singular kind and characterising generics before the instruction intervention. They were only significantly different in their production of bare plural in both generic contexts, yet both groups are significantly less accurate than the native baseline. This difference between the groups could make the pre-test differences a confounding variable that affects the post-test results on the impact of instruction on genericity acquisition if not considered in discussing the impact of the instruction.

7.2.2 Acceptability Judgement Task Results

The AJT tested the speakers' mental grammars by involving them in rating the mapping of five NPs to two generic contexts (characterising and kind generic meanings) on a 1 to 4 Likert scale, whereby one denoted complete confidence that the sentence was unacceptable, two denoted that it was likely unacceptable, three denoted that it was likely acceptable, and four denoted complete confidence that the sentence was acceptable. The following table summarises the predicted target-like ratings in each generic context based on the semantic literature on genericity and the results of the native speakers.

[11] P_NPP refers to plural forms in kind generics, P_NPS refers to the singular form in kind generics, P_S_P refers to the plural form in characterising generics, and P_S_S refers to the singular form in characterising generics.

Table 7-6 Predicted target-like ratings in the AJT based on the semantic literature on genericity and the native speakers' study results.

NP types/contexts	Characterising generics	Reference to kind
Bare plural NP	Acceptable (either 3 or 4)	Acceptable (either 3 or 4)
Bare singular NP	Unacceptable (either 1 or 2)	Unacceptable (either 1 or 2)
Definite singular NP	Unacceptable (either 1 or 2)	Acceptable (either 3 or 4)
Definite plural NP	Unacceptable (either 1 or 2)	Unacceptable (either 1 or 2)
Indefinite singular NP	Acceptable (either 3 or 4)	Unacceptable (either 1 or 2)

In terms of coding, the dataset comprises the ratings, the dependent variable, and conditions of generic form-meaning mapping, which is the predictor variable. The dependent variable, ratings, is an ordinal variable with four levels. The predictor variable, conditions of generic interpretation, was a five-level categorical variable. These levels describe the various form–meaning mapping conditions and include the five NPs, as explained in Table 7.6. By analysing the ratings under each level of the predictor variable, we can gain insights into nuances and variations in generic form-meaning mappings for these speakers. This coding method was more effective in capturing the results of the task than reducing ratings to the binary acceptable or unacceptable. As done in the analysis of native speakers study AJT in Section 5.4.3, the AJT data was analysed using the *clmm* function from the *ordinal* package. In each model, the maximum possible structure was used and the model was compared to a null model using the *lrtest* function to assess the goodness of fit. The significance of each fixed effect was tested using the single-term deletion test. The following section presents the outcomes of the pretest AJT. Descriptive statistics and statistical modelling results are showcased for the native controls initially, serving as a benchmark and validating the instrument. Subsequently, the results for the L2 learners are presented.

7.2.2.1 Native control AJT results

In terms of description, Figures 7.11 provide an overview of the density, means, and standard deviations of the native controls' ratings.

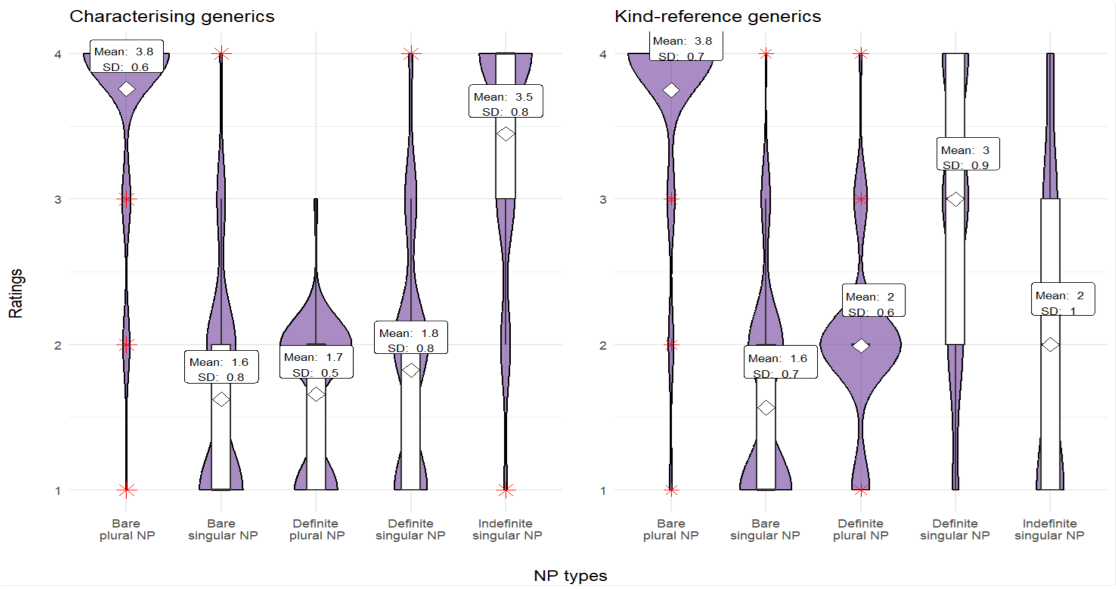


Figure 7.11 Distributions, mean, and SD of the native controls' judgements in the AJT pre-test. The diamond shape represents the mean, and the red stars reflect the outliers.

This figure shows that the average ratings of the bare plural NP were consistently the highest in both types of generic context (*M* =3.8), followed by a high rating average of the indefinite singular in the characterising generic context (*M* = 3.5) and the definite singular in kind-reference generics (*M* = 3). Also, the conditions marked as unacceptable in Table 7.6 had mean ratings of two or less, which is lower than the middle of the rating scale. This shows that the native control group, used as a baseline for comparison, did not accept these NP conditions to express generic meaning. The red stars stand for the outliers, which revealed that few native controls rated the acceptable NP types as unacceptable. In sum, the descriptive results showed that the native control ratings accepted the predicted from-meaning mappings in both generic sentences. However, it is necessary to check the significance of the different mean ratings in different conditions by constructing a CLMM and undertaking a pairwise comparison.

First, a null model was created with only the random effects (participants and items), revealing that the random effects did not explain any variance. However, it will be kept following the experimental protocol (Christophe, 2021). Next, a complete model was implemented using the ratings as the dependent variable, the conditions as the predictor variable, and bare plural NP with kind generic meaning as the reference level. Finally, using the *emmeans* package, Tukey-corrected post hoc pairwise comparisons were conducted. In what follows, the relevant pairwise comparisons for the native controls' rating model are reported. A likelihood ratio test of the null model and the model with the fixed effects was significant (LR (9) = 991.2, $p < .001$); hence, the model with the fixed effect provided a significantly better fit for the data. The odd ratios, estimate, standard error, and *z* and *p* values are summarised in Table 7.7.

Table 7-7 Native control's AJT model; odd ratios, estimate, standard error, and z and p values

Predictors	**Odds Ratios (CI)**	**Estimate**	**Std. Error**	***Z* value**	***p***
Kind reference bare singular	0.00 (0.00 – 0.00)	−6.07	0.36	−16.62	< .001*
Kind reference Definite plural	0.01 (0.00 – 0.02)	−4.75	0.34	−13.86	< .001*
Kind reference Definite singular	0.08 (0.04 – 0.15)	−2.57	0.32	−7.87	< .001*
Kind reference Indefinite singular	0.01 (0.00 – 0.01)	−4.97	0.36	−13.99	< .001*
Characterising sentence Bare plural	0.95 (0.44 – 2.05)	−0.05	0.39	−0.13	.90
Characterising sentence Bare singular	0.00 (0.00 – 0.01)	−5.97	0.38	−15.82	< .001*
Characterising sentence Definite plural	0.00 (0.00 – 0.01)	−5.65	0.37	−15.36	< .001*
Characterising sentence Definite singular	0.00 (0.00 – 0.01)	−5.36	0.37	−14.58	< .001*
Characterising sentence Indefinite singular	0.26 (0.13 – 0.51)	−1.35	0.35	−3.87	< .001*

Note: the model formula is Ratings ~ Conditions + (1 | id) + (1 | item); number of obs.:1200; participants, 20; items, 12; * significant at the .05 significance level.

What stands out in Table 7.7 is that the native controls rated bare plural similarly in both contexts, as there was no significant difference between the mean ratings of bare plural in both contexts (p =.90, p >.05). Considering kind generics, a comparison of the bare plural acceptance in kind generics to other conditions showed that there was a significant statistical difference between the ratings, as indicated by the *p value*s in the table, when comparing the reference level bare plural in kind generics to all the NP types in this generic meaning. Moreover, the post hoc pairwise comparison showed that the ratings of the definite singular in kind generics were significantly different from the ratings of the non-target-like NPs, including definite plurals (*ES* = 2.18, *SE* =.25, z = 8.68, p <.001), indefinite singulars (*ES* = 2.40, *SE* =.27, z = *8.95*, p <.001), and bare singulars (*ES* = 3.50, *SE* =.28, z = *12.65*, p <.001). These results indicate that the native controls rated definite singular NPs significantly higher than all non-target NPs.

Turning now to the native controls' ratings in the characterising generics, pairwise comparisons revealed that the native controls' mean ratings of the bare plural were significantly different from their mean ratings of all other non-target-like NPs, including bare singulars (*ES* = 5.92, *SE* =.36, *z = 16.46, p* <.001), definite singulars (*ES* = 5.30, *SE =.35, z = 15.2, p* <.001), and definite plurals (*ES* = 5.6, *SE* =.35, *z = 16, p* <.001). The mean rating of the bare plural differed significantly from the ratings of the indefinite singular (*ES* = 1.30, *SE* =.33, *z* = 3.9, *p* =.003), which was also a target-like NP in this context. To ensure that the native control ratings conformed to the predicted target-like ratings in Table 7.6, the pairwise comparisons between the indefinite singular mean ratings and other NPs' ratings in the characterising generic context are presented in Table 7.8.

Table 7-8 Indefinite singular NP ratings compared to other characterising generics NPs in Tukey-corrected pairwise comparisons for native baseline AJT.

contrast	**Estimate**	**Std. Error**	***Z ratio***	***p***
Definite singulars	4.00	.29	13.75	< .001*
Definite plurals	4.30	.29	14.75	< .001*
Bare singulars	4.61	.30	15.26	< .001*

Note: * significant at the .05 level

Table 7.8 shows that the mean ratings of the indefinite singular in characterising generics significantly differed from those of the other NPs in this generic context. No significant differences were found between the mean ratings in the pairwise comparisons between the definite plural, the bare singular, and the definite singular as the *p* >.05 (as shown in pairwise comparisons in Appendix D). To summarise, the results of the CLMM showed that the native controls accepted both bare plural and definite singular NPs in kind reference generics and bare plural and indefinite singular NPs in characterising generics.

7.2.2.2 L2 learners' Results

To report how the L2 learners rated the mapping of the NPs to the generic meaning before the intervention, the same procedures were used as for analysing the native control responses. The L2 learners' ratings were described by plotting the rating distributions, mean ratings, and standard deviation for both the L2 learners and native controls in Figure 7.12.

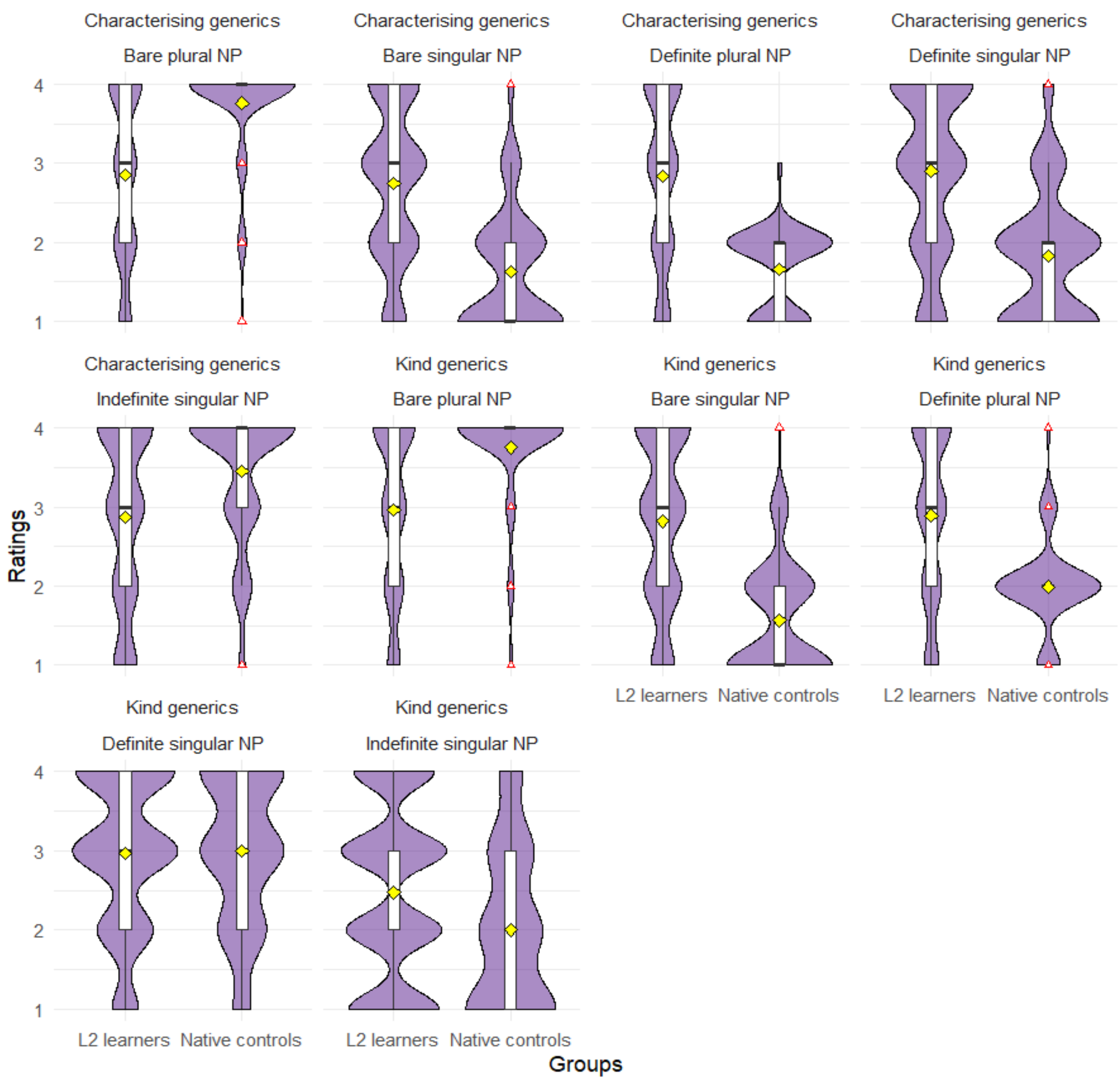


Figure 7.12 Raw ratings densities, mean ratings, and variance in pretest AJT for the L2 learners, compared to the native benchmark. The yellow diamond indicates the mean rating for each condition.

Figure 7.12 shows that the L2 learners' ratings were similar in almost all conditions, as indicated by the mean ratings and densities. For all conditions in both contexts, while the native controls' mean ratings of expected acceptable NPs in generic meanings were higher than three, and their average ratings of unacceptable NP types were lower than 2.5, L2 learners seemed to accept all NPs, as the mean rating of all conditions is higher than the midpoint of the scale and close to 3 (likely acceptable). Mapping definite singulars to kind generics is the only condition in which L2 learners' rating density and mean were similar to the native benchmark. [12]

The main question this AJT seeks to answer is whether the L2 learners show target-like form-meaning mappings in both generic contexts. The within-group differences were modelled using a CLMM, following the same procedure as the one outlined in the native controls data analysis. After fitting the intercept-only model and the model with *ratings* as the dependent variable, *conditions* as the fixed effect, and by-participant and by-item random intercepts, the likelihood ratio test revealed that the model with the fixed effect provided a significantly better fit to the data (*LR* (9) = 65.57, *p* <.001). Table 7.9 presents the odd ratios, estimates, standard errors, *and z- and p-values* of the L2 learners' ratings, mapping bare plurals to kind generics as the reference level of comparisons.

Table 7-9 Odd ratios, estimates, standard errors, *z*- and *p value*s of the L2 learners' ratings in AJT pretest.

Predictors	**Odds Ratios (CI)**	**Estimate**	**SE**	**Z value**	***P* value**
Kind generics [bare singular]	0.74 (0.57 – 0.96)	−0.30 |	0.13	−2.24	.03*
Kind generics [Definite plural]	0.86 (0.66 – 1.12)	−0.15	0.13	−1.14	.25
Kind generics [Definite singular]	0.91 (0.70 – 1.18)	−0.09	0.13	−0.69	.49
Kind reference [Indefinite singular]	0.41 (0.32 – 0.53)	−0.89	0.13	−6.69	<.001*
Characterising sentence Bare plural	0.79 (0.61 – 1.03)	−0.23	0.13	−1.72	.09
Characterising sentence Bare singular	0.63 (0.49 – 0.82)	−0.45	0.13	−3.44	.001*
Characterising sentence Definite plural	0.76 (0.59 – 0.99)	−0.27	0.13	−2.00	.05

[12.] Both groups' data was modelled using the clmm model to compare the L2 learners to the nati ve controls, and all comparisons were significant except for the ratings of definite singular NP i n kind generics (L2 to NS *ES* =-0.04, *SE*=0.23, *z*= -0.188, *p* =.99).The model is shown in appendix D.

Predictors	**Odds Ratios (CI)**	**Estimate**	**SE**	**Z value**	***P* value**
Characterising sentence Definite singular	0.86 (0.66 – 1.12)	−0.15	0.13	−1.09	.28
Characterising sentence Indefinite singular	0.84 (0.64 – 1.09)	−0.18	0.14	−1.31	.19

Note: model: Ratings ~ Conditions + (1 | id) + (1 | item), Hess=T, Observations:3835, number of participants sixty-four, number of items, 64, * significant at the .05 level

Table 7.9 shows the model results that compared all the NP types in both generic contexts to the reference level (bare plural NPs in kind generics). The data in the table revealed that, in kind generics, the L2 learners rated bare plurals the same way they rated definite singulars and definite plurals (the L1 form–meaning mappings) since the model results in the table above did not show any significant differences between these ratings in this context. Bare plurals in kind generics were accepted significantly more than bare singulars and indefinite singulars (the non-L2 and non-L1 target-like forms, respectively) by the L2 learners, as indicated by the negative estimate and higher z values in the model above. As for the singular form, the definite singular is the acceptable singular form in kind generics. Further post hoc Tukey-corrected pairwise comparisons that compared the definite singular to other unacceptable forms revealed that there was no significant difference between the ratings of definite plural and definite singular in this context (*ES* = 0.07, *SE* = 0.13, *z* = −.51, *p* = 1); no significant difference between the ratings of the definite singular and the bare singular (*ES* =-0.21. *SE* = 0.13, *z* = -1.62, *p* = .48), and the definite singular was significantly rated higher than the indefinite singular in kind generics (*ES* = 0.82, *SE* = 0.13, *z* = 6.25, *p* <.001). These results suggest that L2 learners accepted definite singulars, definite plurals, and bare singulars similarly with kind generics. I can therefore conclude that L2 learners showed statistically similar ratings for the L2 form (bare plural) and their L1 form-meaning mappings (definite singular and definite plural) in the kind generic meaning. They accepted the definite singular in the same way they accepted definite plurals and bare singulars in kind generics, which reflects a challenge in the acquisition.

Turning now to the characterising generic context, Table 7.9 compares the L2 learners' ratings of bare plurals in this context to bare plurals in kind generics. The difference between mapping bare plurals to the two contexts was not significant (*p* =.09, *p* >.05), showing that the L2 learners similarly accepted the bare plural NPs in both generic contexts. Post hoc comparisons reveales that the L2 learners accepted bare plurals, definite singulars, definite plurals, bare singulars, and indefinite singulars similarly, as the differences in rating means were not significant according to the pairwise comparisons of the bare plural in the characterising generic context, as shown in Table 7.10.[13]

Table 7-10 Post hoc pairwise comparisons of mapping bare plural NP to the characterising generics and other NP typos by L2 learning in AJT pretest.

Predictors	**Estimate**	**SE**	**z-value**	**p**
Bare singular characterising generic	0.22	.13	1.72	.78
Definite plural characterising generics	0.04	.13	0.29	.99
Definite singular characterising generics	−0.09	.13	−0.64	.99
Indefinite singular characterising generics	−0.05	.13	−0.40	.99

As for how L2 learners rated indefinite singulars (the acceptable singular form in this context), the pairwise comparisons related to this condition revealed that the ratings of the indefinite singular were not significantly different from the ratings of all the unacceptable forms, including definite plurals (*ES*= 0.09, *SE*= 0.13, *z*=0.69, *p*= .99), definite singulars (*ES*= −0.03, *SE*= 0.13, z= −0.23, *p*= 1), and bare singulars (*ES*=0.28, *SE*=0.13, *z*=2.12, *p*=.5). In summary, unlike the native controls, the L2 learners accepted all the NP types in characterising generic meaning, which reflects a challenge in the acquisition.

In the previous analysis, I investigated how L2 learners map different NPs to the generic meanings in the pretest AJT. It was necessary to establish that the comparison and the experimental groups were homogeneous and that their responses prior to the intervention in the AJT is not a confounding factor that affected the post-intervention results. To do so, the same CLMM modelling procedures used to analyse the L2 learners' ratings were applied to the data of

[13] To avoid type I errors in pairwise comparisons, I tried a separate model for the characterising generic meaning conditions, which gave the same results with no significant difference between all conditions at the .05 level. The model results are as follows: Bare plural to bare singular (*OR* = .80, *CI* = 0.62-1.04, *ES*= −0.22, *z* = −1.86, *p* =.09), to Definite singular (*OR*= 1.09, *CI*= 0.84-1.42, *ES*= 0.09, *z* = 0.68, *p* =.50), to definite plural (*OR* = .97, *CI* = 0.75-1.26, *ES* =−0.03, *z* = −0.22, *p* =.83), and to indefinite singular (*OR* = 1.06, *CI* = 0.82-1.38, *ES* =0.06, *z* = 0.45, *p* =.65)

these two groups. Before presenting the result of the CLMM, Figure 7.13 displays the density of ratings per condition comparing both groups.

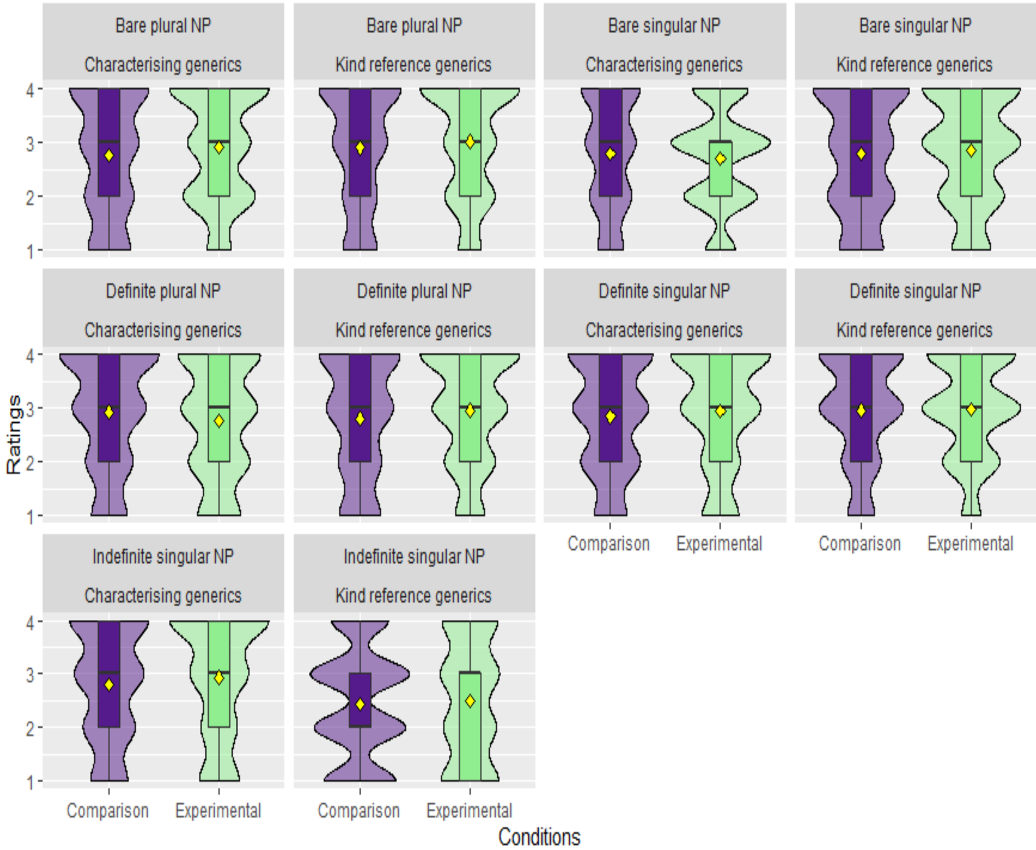


Figure 7.13 Violin plots of the L2 learner experimental and comparison groups' ratings per condition in the AJT pre-test, the diamond shape presents the mean rating.

This figure shows that the groups had similar mean ratings and densities. To ensure group similarity using CLMM, the model started with conditions, groups, and their interaction as the fixed effect with by-participant and by-item random intercepts. The results of the likelihood ratio test revealed that the interaction model provided a significantly better fit to the data than the null model (LR (18) =77.11, p <.001). Single-term deletion showed that the interaction between the groups and conditions was not significant (χ^2 = 11.59, df = 9, p =.24, p >.05). The results of the interaction model revealed no significant difference between the experimental and comparison groups in their ratings in the pretest AJT (ES =.92, SE = 0.22, z = 77, p =.44). All the interaction results showed no significant difference between the interaction between conditions and groups in different conditions (see Appendix D for full model results). This result suggests that the group difference in the pre-test was not a confounding factor.

7.2.3 Forced-choice Task Results

This task evaluated the L2 learners' selection of articles in different generic conditions. The participants read a series of dialogues and afterwards completed the gaps by selecting appropriate articles, namely *a, an, the*, or *Ø*, based on contextual cues, as described in Section 6.4.4. Table 7.11 summarises the conditions and the target choices for this task based on the semantic literature on genericity and native speakers' data.

Table 7-11 Forced-choice task conditions and expected target choices based on data from the native speaker study and the semantic literature.

Conditions	**Items and target choices**
Characterising generics	Plural form: (*Ø*) bare plural NP is the target choice. Singular form: *(a)* is the target choice.
Kind generics	Plural form: (*Ø*) bare plural NP is the target choice. Singular form: (*the*) is the target choice.

A binary scale was used to grade this task. Each test item was categorised as one if the participant correctly selected the expected form choice for the tested condition and zero otherwise. Then, for each participant, I calculated a total score out of four for each generic condition in the table above. Then, a percentage of the total expected choice score for those by-participant totals was produced per condition by multiplying the participant's total score of expected choice by one hundred and dividing it by the total number of test items in each condition.

Data analysis was conducted using binomial logistic regression analysis; a generalised linear mixed-effect model (GLMMs) was used to model the relationship between a binominal categorical response variable and one or more predictor variables while accounting for the correlation among observations within the same group. This allows for a more accurate analysis of the effects of the predictor variables on the response variable, considering any potential variability introduced by diverse groups or clusters in the data (Linck & Cunnings, 2015). The *glmer* function of the *lme4* package achieved this goal (Bates et al., 2015). The package *emmeans* (Lenth, 2023. Version: 1.8.7) was used for post hoc Tukey-corrected pairwise comparisons. The dependent variable in all models included the choice of expected form (binary). The possible maximal structure was the aim of each model. A detailed description of the structure of each model is provided when reporting results. The statistical analysis of the native speakers' data, followed by the results of the L2 learners, is reported in the following sections.

7.2.3.1 Native Controls' Results

The native speakers' performance in the FCT was plotted using violin plots to describe the diversity in the participants' total rates of hitting expected form choice and bar graphs to compare the proportion of expected and other NP choices between conditions.

The variability in the total choice of expected forms across individuals in each condition was examined separately by computing each participant's percentage of expected forms choice total score and visualising the densities, as shown in Figure 7.14.

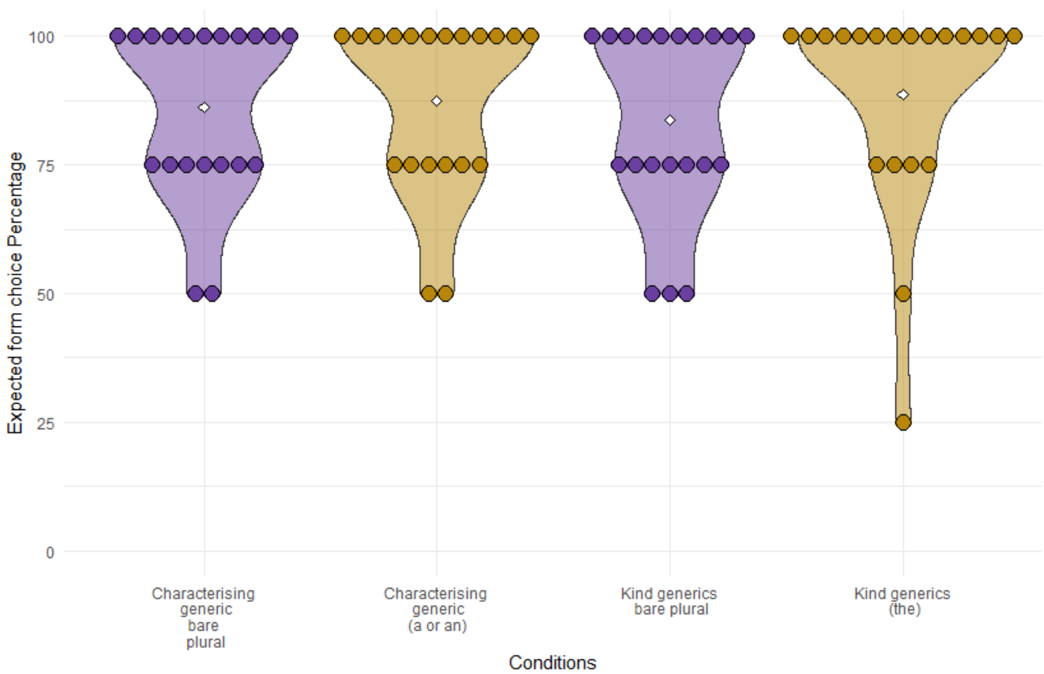


Figure 7.14 The native controls' total percentage of making the expected choice per condition in the FCT. The diamond shape indicates the group's mean percentage in each condition.

As we can see, Figure 7.14 demonstrates that most of the native speakers made the expected choices at the 75% rate or above. Only a few cases out of twenty participants demonstrated a lower than 75% overall score in hitting expected choices in any given condition. In particular, 2/20 showed lower performance in choosing the bare plural 'Ø' and the singular indefinite article 'a' in characterising generic conditions. As for kind generics, 3/20 showed a similar lower rate of expected form choices in choosing the bare plural '*Ø*' and 2/20 in choosing the definite article '*the*'. Overall, the native speakers slightly differed in the average rates of choosing the expected form in different conditions, as shown in the following table.

Table 7-12 Descriptive statistics of the English native controls' total rates of making expected choices in the FCT pre-test

Condition	**Mean**	**Standard Deviation**
Characterising generics (Bare plural NP)	86.25	17.16
Characterising generics (*a/an*)	87.50	17.21
Kind generics (Bare plural NP)	83.75	18.63
Kind generic (*the*)	88.75	20.64

Turning to the per-condition choices of the native controls, the proportion of expected form choice and other form choice revealed that all conditions have a high proportion of expected form choice by native speakers compared to the other form choice, as Figure 7.15 shows.

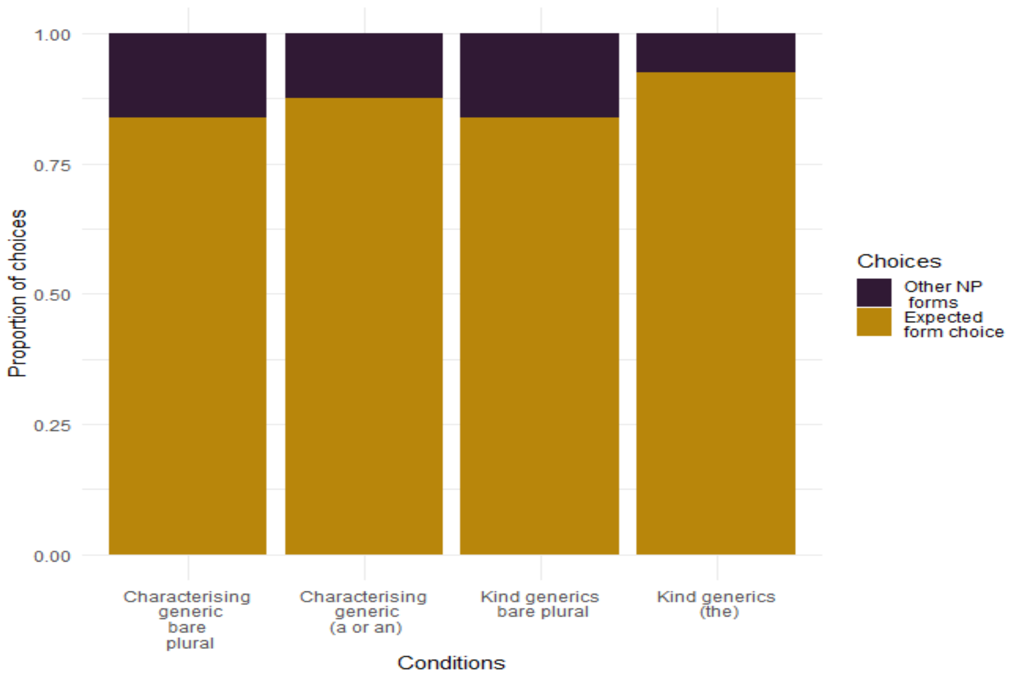


Figure 7.15 Proportions of expected and other forms choices made per condition in the FCT by native control group.

This figure presents the proportions of expected and unexpected choices by the native controls in the FCT. As expected, 93% of choices were for the definite article in kind generics. In the kind and characterising bare plural generic conditions, 84% of choices were for bare plurals, as expected. Finally, the choice of the expected form (the indefinite article) in the characterising generic condition was chosen 88% of the time. However, it is essential to present what unexpected choices were provided by the native benchmark, which is visualised in Figure 7.16.

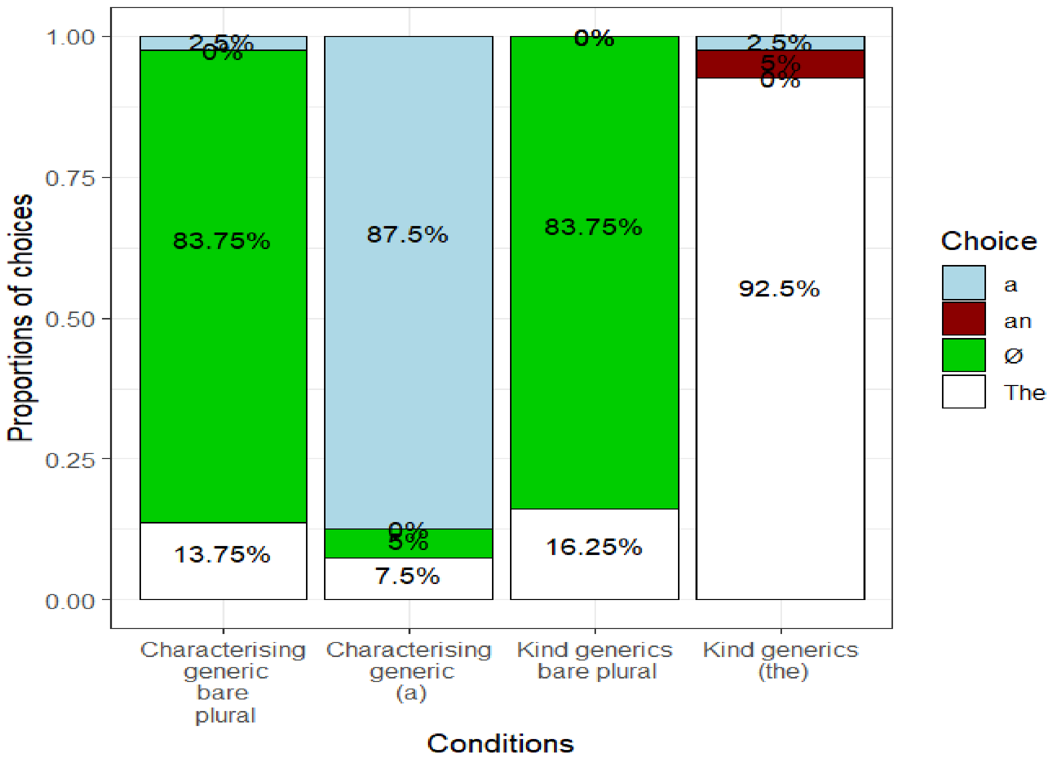


Figure 7.16 Native control's proportions of article choices in the FCT by condition.

The figure shows the proportions of the choices made by the native controls in the FCT for each condition. It shows that using the definite article was the most frequent unexpected choice in both types of characterising generic condition and in the plural kind generic condition.

These descriptive findings reveal that the native control group consistently performed well in associating articles with generic meanings across all conditions, even though occasional unexpected choices were made. However, since the rate of using expected forms varies across conditions, it is necessary to verify that their choices did not exhibit statistically significant differences across all conditions. Consequently, the inferential statistics for the native control group are presented in the subsequent analysis.

Statistical analysis using GLMMs was conducted on the native controls' choices to assess whether there were any significant differences across generic conditions. The model analysed the native controls' responses, focusing on the choice of expected forms (as the binary dependent variable) and conditions as the fixed effect. All four conditions were included in the model, with the mapping of bare plurals in the characterising generic meaning as the reference level. The analysis began with adding the fixed effect of condition and random effects for both items and participants. A likelihood ratio test was used to compare this model to the null random intercepts only model, and it found no significant difference between the two models (*LR* (3) =1.75, *p* =.62). The model with fixed effect AIC = 248 and BIC = 270 show goodness of fit. The model results can be found in Table 7.13. The final model did not reveal a significant main effect of the condition as a fixed effect ($X^{2=}$1.75, *df* = 3, *p* =.63).

Table 7-13 Native speakers' model estimates, standard errors, *z,* and *p* values for the fixed effects in FCT.

Predictors	**Odds Ratios** (CI)	**Estimate**	**Std. Error**	***z* value**	***p-value***
(Intercept)	9.20 (4.80 – 17.62)	2.22	0.33	6.69	<.001*
Characterising generics (a/ an)	1.54 (0.32 – 7.30)	0.43	0.80	0.54	0.59
Kind reference generics (Bare plural)	1.15 (0.25 – 5.31)	0.14	0.78	0.18	0.86
Kind reference generics (the)	2.83 (0.54 – 14.91)	1.04	0.85	1.23	0.22

Note: Number of observations: 320; id, 20; item, 16, Formula (choice of expected forms ~ Condition + (1 | id) + (1 | item), Family= binomial, glmerControl (optimizer = "bobyqa", optCtrl = list (maxfun = 20000)). *Significant at the .05 level.

Table 7.13 summarises the native controls model when including condition as a predictor variable. The significant intercept in the table above indicates that the expected form choices differed considerably from zero in the reference condition (bare plural characterising generics), but the lack of significant difference for the other conditions indicates that there is no evidence that these conditions are different from the reference condition. This suggests a significant level of choosing expected forms even without the influence of the condition as a fixed effect. The model coefficients indicated that the native speakers of the present research were more likely to map English forms to generic meaning as expected in all conditions.

Considering the random effects, the random effects in the model revealed variability across participants (σ^2 = 0.07, *SD* = 0.27) and across items (σ^2 =0.74, *SD* =0.86). However, based on the obtained intraclass correlation coefficient (ICC) value of 0.20, around 20% of the overall variance in native controls' choices can be attributed to variations across participants and between items. The results from the native control group therefore confirm the suitability of the FCT to elicit data about the L2 learners' accuracy and set the benchmark level.

7.2.3.2 L2 Learners' Results

Shifting our focus now to the results of L2 learners, this section reports how the L2 learners respond to the FCT by presenting the descriptive visualisation and the inferential statistical significance of any variance in their accuracy.

The L2 learners' responses were analysed using the same procedures for analysing the native controls' data. Initially, the between-participant total accuracy per condition and its density were plotted. Subsequently, the L2 learners' expected choice proportions across conditions were plotted and compared to the benchmark (native controls). A graphical representation of the proportions of unexpected article choices made by L2 learners was also provided. The aim of this visualisation was to shed light on any challenges encountered by these learners.

To compare how the L2 participants differ in their total accuracy when choosing articles across different generic contexts, Figure 7.17 presents the total accuracy percentage per participant and plots the distributional density of participants' responses.

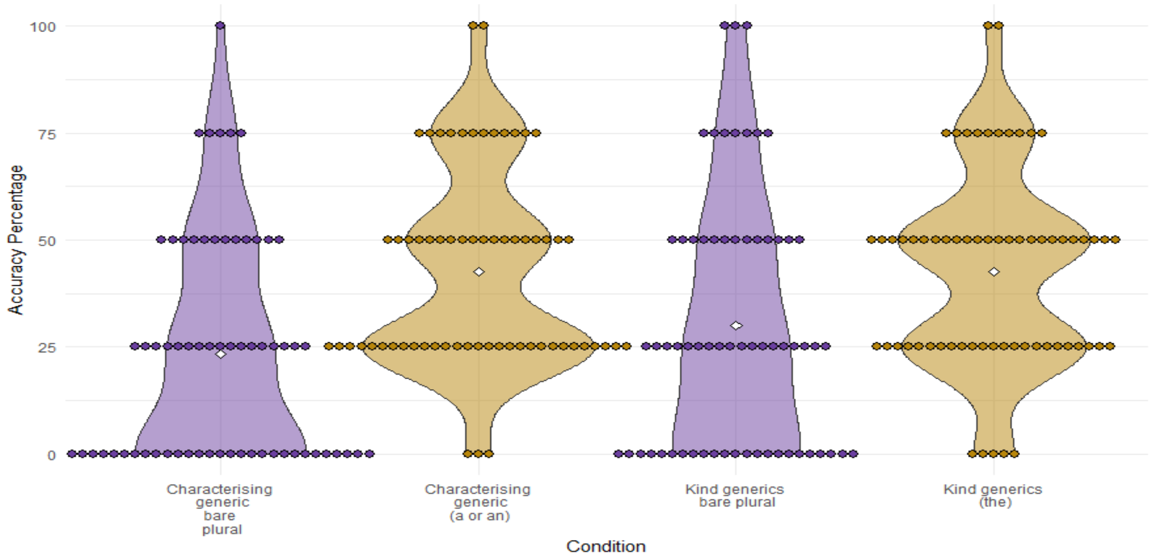


Figure 7.17 L2 learners' total accuracy in % in the FCT pre-test. The white diamond shape represents the mean.

This figure displays the L2 learners' total percentage of accuracy rate per participant in each condition. As this figure shows, most L2 participants showed a 50% or less accuracy rate across all conditions, in contrast to the native controls who showed above 75% accuracy. Only 6 out of 64 showed a native-like accuracy percentage in choosing the bare plural (no article option) in the plural characterising generics. This number was increased to 15 out of 64 when choosing the indefinite article 'a' in the singular characterising condition. As for kind generic conditions, only 9 out of 64 and 11 out of 64 learners reached ceiling accuracy in mapping the plural and singular forms to kind generics, respectively. These results suggest that mapping English articles to generic contexts posed challenges for Arabic-speaking learners of English as L2. A comparison between the mean accuracy rates of L2 learners and native controls revealed lower accuracy rates among the L2 learners, as shown in Table 7.14.

Table 7-14 The pretest total accuracy average rates of L2 learners in the FCT, compared to the native controls' choice of expected form total rates.

Condition	**Mean % (SD)**	
	Native controls	L2 learners
Characterising generics (Bare plural NP)	86.25% (17.2)	23.4% (26.3)
Characterising generics (*a/an*)	87.50% (17.2)	42.6% (23.4)
Kind generics (Bare plural NP)	83.75% (16.8)	30% (29.6)
Kind generic (*the*)	88.75% (20.6)	42.6% (23.4)

As for between conditions, the proportions of L2 learners' choices of expected and other NP forms are presented in Figure 7.18.

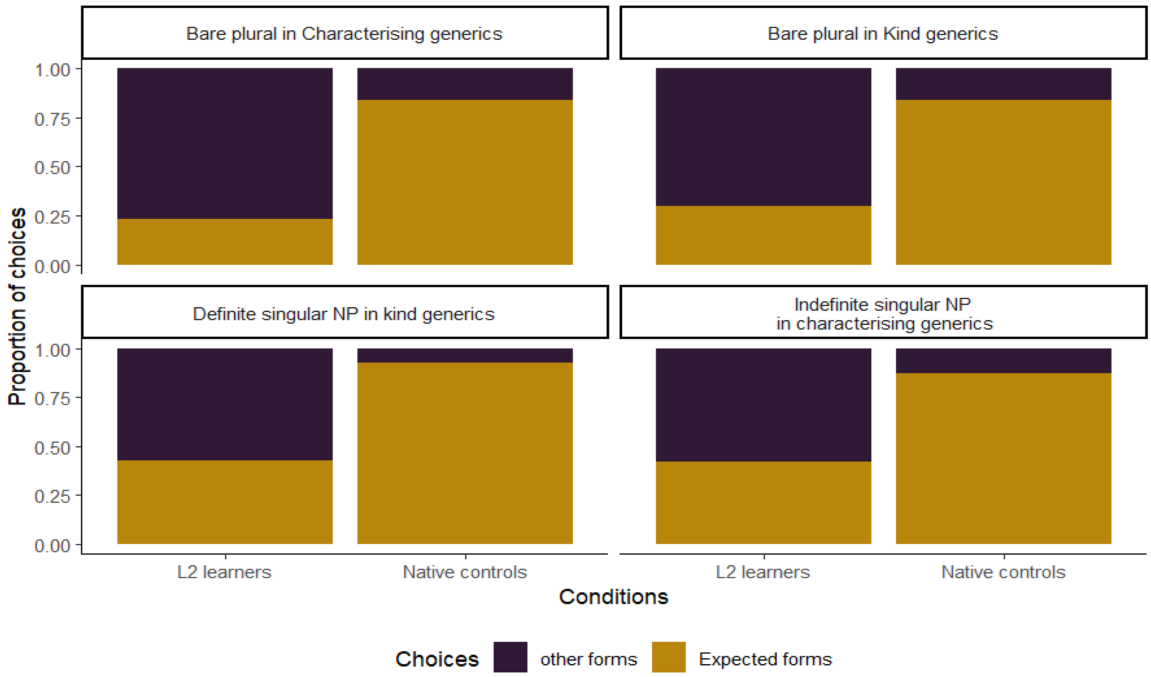


Figure 7.18 L2 learners' proportions of expected and other forms across conditions in the FCT pre-test compared to the native controls' choices.

This figure compares the L2 learners' choices of the expected forms to those of the native controls. It shows that L2 learners' expected form choice proportions were lower than those of the native speakers and lower than the proportions of other form choices across all four conditions. Less than one-third of the choices in both plural generic contexts were bare plurals as expected, which was only accurately chosen in 23% of characterising generics and 30% of the kind generics.

The choice of expected singular forms in both singular generic conditions was higher than the choice of expected plural forms in both plural generic conditions. The proportion of expected choices was 42% for the indefinite singular in characterising generic contexts and 42.5% for the definite article in the kind generics. These results are highly different from those of the native speakers. The L2 learners showed a lower proportion of expected choices than the native speakers (about 42% for the L2 learners in both singular generic conditions vs. 93% of the expected form choice of '*the*' and 88% of the expected form choice of '*a*' by the native speakers). The L2 learners' proportion of correct choices of bare plurals in characterising and kind generics were 30% and 23%, respectively, which were also low compared to the native controls' 84% choice of expected forms in both plural generic conditions. This indicated that the L2 learners faced challenges mapping articles to different generic meanings.

However, it is essential to provide a detailed account of the unexpected choices made by L2 learners. Figure 7.19 illustrates the specific proportions of each unexpected form choice per condition by the L2 learners.

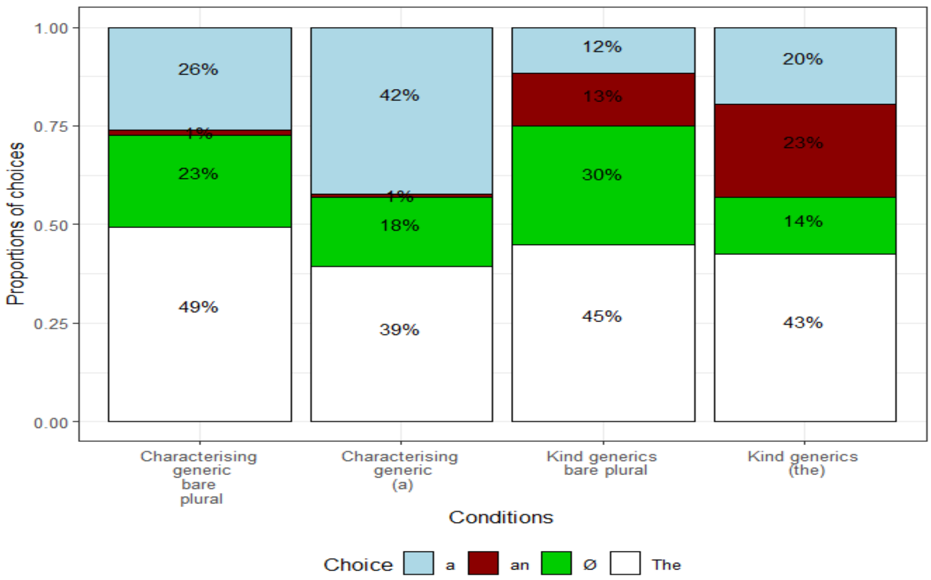


Figure 7.19 Proportions of each form choice per condition by the L2 learners in the FCT pre-test.

Figure 7.19 displays the unexpected form selection proportions by the L2 learners in the pre-test FCT. The figure shows that the selection of the definite article ‘the’ was more frequent in both plural generic contexts. Specifically, the unexpected choice of using the definite plural form (the L1 form) was highest in characterising generics, accounting for 49% in that context and 45% in kind generics with plural forms. The use of indefinite articles followed, comprising 26% and 25% in these two conditions, respectively. Shifting the focus to mapping forms to generic meanings in singular contexts, the most common error involved selecting the definite article ‘the’ (L1 form), constituting 39% of usage when mapping singular forms to characterising generic meanings, where the expected form was the indefinite article. Additionally, 18% of unexpected choices involved using the bare singular form. When mapping singular forms to kind generics, 43% of the unexpected choices were indefinite singulars and 14% were bare singulars.

These descriptive statistics reveal that L2 learners exhibited relatively low proportions of expected article choice in each generic condition. Comparing L2 learners to the native controls using GLMM shows that the groups significantly differ in their choices in the FCT (ES = 2.70, SE = 0.20, z = 13.62, p <.001).[14] This finding underscores the importance of investigating whether these low proportions of expected choices are statistically significant across different conditions.

Analysis was conducted using GLMM, as done for the native baseline data. In this analysis, accuracy was the dependent variable with two levels (accurate choice of expected forms and choice of unexpected form, i.e., binary) and *conditions* (with four levels including the target production described in Table 7.11), by-participants and by-items as random effects. The AIC of this model was 1281, and its BIC was 1315, with a significant effect of condition as indicated by the single term deletion test (X^2 = 8.034, df = 3, p =.04). Table 7.15 summarises the model's odds ratios, estimates, standard errors, and z and p values.

[14] The model and pairwise comparisons are shown in Appendix D.

Table 7-15 L2 learners pre-test FCT *glmer* model odd ratios, estimates, standard errors, and z and *p* values of the fixed effects

Predictors	Odds Ratios (CI)	Estimate	St. Error	z- value	p-value
(Intercept)	0.13 (.06 – .29)	−2.064	0.42	−4.89	< .001*
Characterising generics (a/ an)	2.44 (1.26 – 4.73)	0.89	0.34	2.65	.008*
Kind reference generics (Bare plurals)	1.40 (.72 – 2.73)	0.33	0.34	0.98	.328
Kind reference generics (the)	2.45 (1.27 – 4.76)	0.90	0.34	2.66	.008*

Note: Formula: Accuracy ~ Condition + (1 | id) + (1 | item), Data= L2 FCT, glmerControl (optimizer = "bobyqa", optCtrl = list (maxfun = 20000)); Number of observations: 1024, id, 64; item, 16; *Significant at the 0.05 level.

This table displays the results of the GLMM on the L2 learners' FCT pre-test. The intercept of this model is significant, indicating that L2 learners' accuracy in choosing the expected form is considerably different from chance in the reference condition, namely bare plural in characterising generics. Moreover, the L2 learners mapped the bare plural to both kind and characterising generics with the same accuracy, as there was no significant difference between the odd ratio of choosing the expected bare plurals in the kind generic condition and the reference level bare plurals in characterising generics. However, their accuracy in mapping bare plurals to characterising generic conditions was significantly lower than their accuracy in mapping the singular forms (a) to characterising generics and (the) to kind generics, (p = .008) in both conditions. This highlights the need for more post hoc comparison to determine whether the L2 learners' accuracy in mapping singular forms to both generic conditions significantly differed from each other and from using plural forms in kind generics. To achieve this, Tukey-corrected post hoc pairwise comparisons were utilised, and Figure 7.20 presents its relevant results.

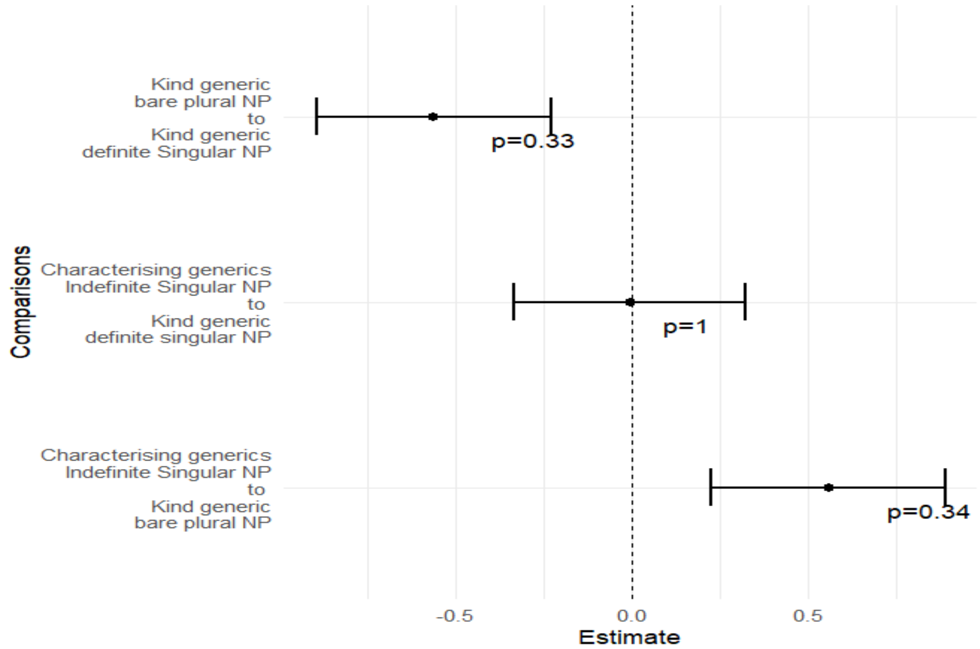


Figure 7.20 L2 learners FCT post-hoc pairwise comparisons.

The pairwise comparisons revealed that the L2 learners' accuracy in mapping the bare plural to the kind generic meaning was not significantly different from their accuracy in mapping the definite singular to this meaning. Moreover, there was no significant difference between mapping the indefinite singular to characterising generic meaning and mapping bare plural and definite articles to the kind generics.

In summary, statistical modelling and pairwise comparisons revealed that L2 learners struggled to map the English forms to generic contexts in the FCT pre-test. Although they demonstrated better accuracy in mapping singular forms to kind and characterising generics, their accuracy was lower when mapping plural forms to kind generics. Notably, mapping bare plurals to characterising generics posed the most significant challenge for L2 learners during the FCT pretest, as evidenced by lower accuracy than other conditions.

Recall that after the pre-test, the L2 learners were divided into two groups (experimental and comparison) before the intervention. It is crucial to ensure that both groups exhibited the same accuracy in choosing the expected article during the FCT pretest. This step prevents pre-teaching differences between the groups from confounding the treatment post-test results. Regarding the L2 learner experimental and comparison groups' by-participant accuracy, Figure 7.21 summarises the participants' total accuracy rates in each group.

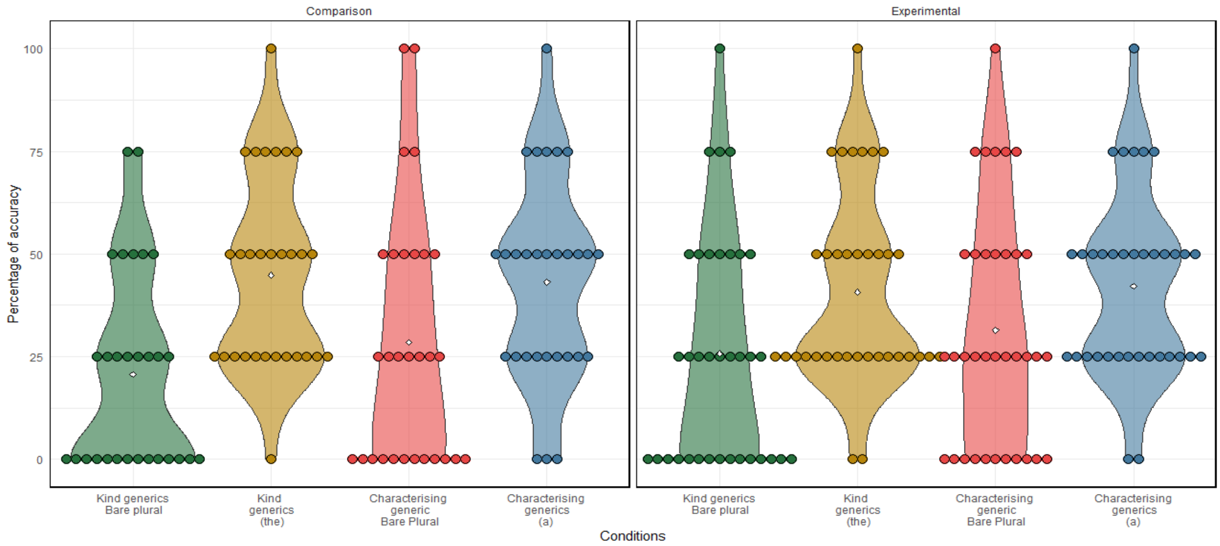


Figure 7.21 L2 learners' comparison and experimental groups' accuracy rates in the FCT pre-test. The white diamond shapes indicate mean accuracy rates for each condition.

This figure shows the accuracy rate for each participant and the mean accuracy rate for the groups per each condition. It demonstrates that almost all the comparison group participants have either 50% or less accuracy in kind generic conditions, except for 2 out of 29 participants who had 75% and above accuracy in choosing bare plural in kind generics and 7 out of 29 who likewise had 75% or higher accuracy in choosing '*the*' in kind generics. As for the experimental group, few participants showed 75% or above accuracy in kind conditions: 4 out of 35 in choosing the bare plural in the plural condition and 7 out of 35 in choosing '*the*' in the singular condition. The rest of the experimental group showed either 50% or less accuracy.

As for the characterising generic conditions, both groups showed a similar density of accuracy rates, with most members of both groups having 50% or less accuracy across all conditions. Only six learners in each group showed 75% or above accuracy rates in choosing 'a' in characterising generics, and less than seven participants in each group were at the ceiling in choosing bare plural in characterising generics. The mean accuracy rates are indicated by the diamond shape in Figure 7.21 above, and it showed that both groups had approximately similar mean scores in each condition. The mean accuracy percentage and the standard deviations are summarised in the following table:

Table 7-16 Mean accuracy rates per L2 learner groups in the pre-test FCT.

Condition	**Comparison**	**group**	**Experimental**	**group**
	Mean	**SD**	**Mean**	**SD**
Characterising generic (Bare plural)	20.68	24.15	25.71	28.11
Characterising Generics (a)	44.83	23.51	40.71	23.55
Kind generics (Bare plural)	28.45	31.14	31.42	28.66
Kind generics (the)	43.10	24.90	42.14	22.50

To describe the differences between conditions in both groups of L2 learners, Figure 7.22 presents the proportions of correct and incorrect choices in each condition per group.

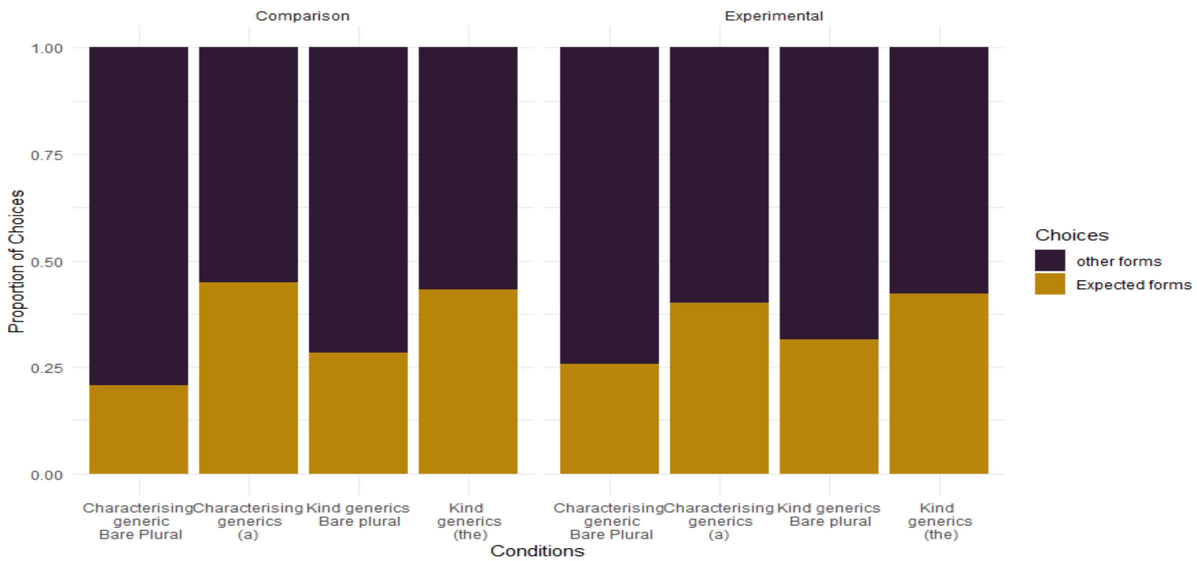


Figure 7.22 Proportions of expected and unexpected choices made by the comparison and experimental groups of L2 learners by condition in the FCT pre-test.

This figure presents the proportions of expected and unexpected choices by both the experimental and comparison groups of L2 learners. It demonstrates that the comparison and experimental groups were less than 50% accurate in all conditions. Both groups showed higher proportions of accurate choices in choosing singular forms in both singular generic conditions compared to choosing bare plural forms in both plural generic conditions. The comparison group's proportion of choosing the expected indefinite singular article in singular characterising generics was 44%, and their proportion of choosing the definite article in singular kind generic contexts was 43%. Also, the experimental group's proportion of choosing the target-like indefinite singular in singular characterising generics was 40%, and they chose the definite article in singular kind generics 42% of the time. As for choosing plural forms with generic meanings, the figure showed that both groups had lower proportions of choosing bare plural NP in kind generics, with the comparison group choosing the expected bare plural 28% of the time, and the experimental group choosing it in 31% of the time in this meaning. Finally, both groups had their lowest proportions of choosing the expected form for bare plural forms in characterising generics (about 20% in the comparison group and 25% in the experimental groups).

Although the descriptive statistics showed similarities between both L2 learners' groups, this similarity needs to be tested statistically. To this end, a binary GLMM was constructed to check if there was any significant difference in accuracy between the L2 learners' groups. The model structure used accuracy as the dependent variable, the interaction between conditions and groups as the fixed effect, and item and participant as random effects. The fitted model did not show a significant effect of the interaction between the generic conditions and groups as determined by the single-term deletion test (X^2 =1.82, *df* = 3, *p* =.61, *p* >.05)[15]. To improve the model fit; the model was reduced to have the group as a fixed effect, and then both models were compared using the likelihood ratio test, which revealed that the difference between the interaction model and the group-only model was not significant (*LR* (6) = 9.8, *p* = .13). However, the group-only model had a lower AIC, resulting in a better fit. In the group-only model, the groups were similar in their accuracy in the FCT pre-test, as there was no significant effect of group as a fixed effect (X^2 = .03, *df* = 1 *p* =.86, *p* > .05), see the model in Appendix D. Hence, L2 learners' experimental and comparison groups' generic form–meaning mappings were similar before the intervention; therefore, the variations may not affect post-test results, as Figure 7.23 shows.

[15] Tukey-corrected pairwise comparison results were checked before the interaction between groups and conditions was removed. None of the comparisons was significant, indicating that the groups were homogenous in each condition before teaching, as attached in Appendix D.

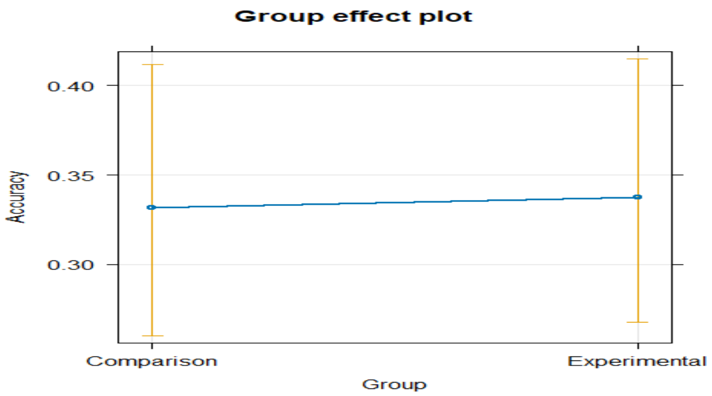


Figure 7.23 Effects plot showing the estimated differences between the two L2 learner groups' accuracy in the pre-test FCT.

7.3 Post-intervention Results

This section presents the results on whether teaching about genericity for eight weeks impacted the L2 experimental group compared to the L2 comparison group by providing the post-test and delayed post-test results for the three experimental tasks. For each post-test task, descriptive and inferential statistical results are presented.

7.3.1 Elicited written production task

The pretest results for this task indicated that the L2 learners did not demonstrate correct form-meaning mappings in producing NP-types in characterising and kind generics. With characterising generic meanings, bare plurals were challenging for L2 learners, as they were significantly less likely to produce these than the native baseline, instead producing a considerable proportion of definite plurals (the L1 form) in addition to the expected production of bare plurals. Likewise, the low proportion of producing the correct indefinite singular form reflects a challenge in this condition, with the L2 learners using definite singulars (the L1 form) and bare singulars at significantly higher rates than the expected form. Similarly, for plural kind generics, the results demonstrated that even though their probability of using bare plurals was significantly higher than other forms, it was still challenging because their production of this form was significantly lower than its production by the native controls. There were high proportions of erroneous production of bare singulars and definite plurals in this condition. Also, the pretest results revealed difficulty in producing the definite singular NP for singular kind generics. The proportion of responses using this form in this condition was significantly lower than the proportion of responses in which the learners used bare singulars.

The pretest results demonstrated that the experimental and comparison groups had similar rates in producing all singular forms within each generic context before the intervention.

However, they differ significantly in producing bare plurals in both generic contexts, with the comparison group, which did not receive the intervention, showing higher accuracy in these contexts. The post-test responses of each group were analysed to determine whether the previously mentioned challenges in using the expected plural and singular forms improved after the intervention. What follows reports the post-test results compared to the pretest results for each group.

In the post-test, each group's production proportions for each NP type were compared to their pretest proportions. Figure 7.24 displays the proportions per test for each group.

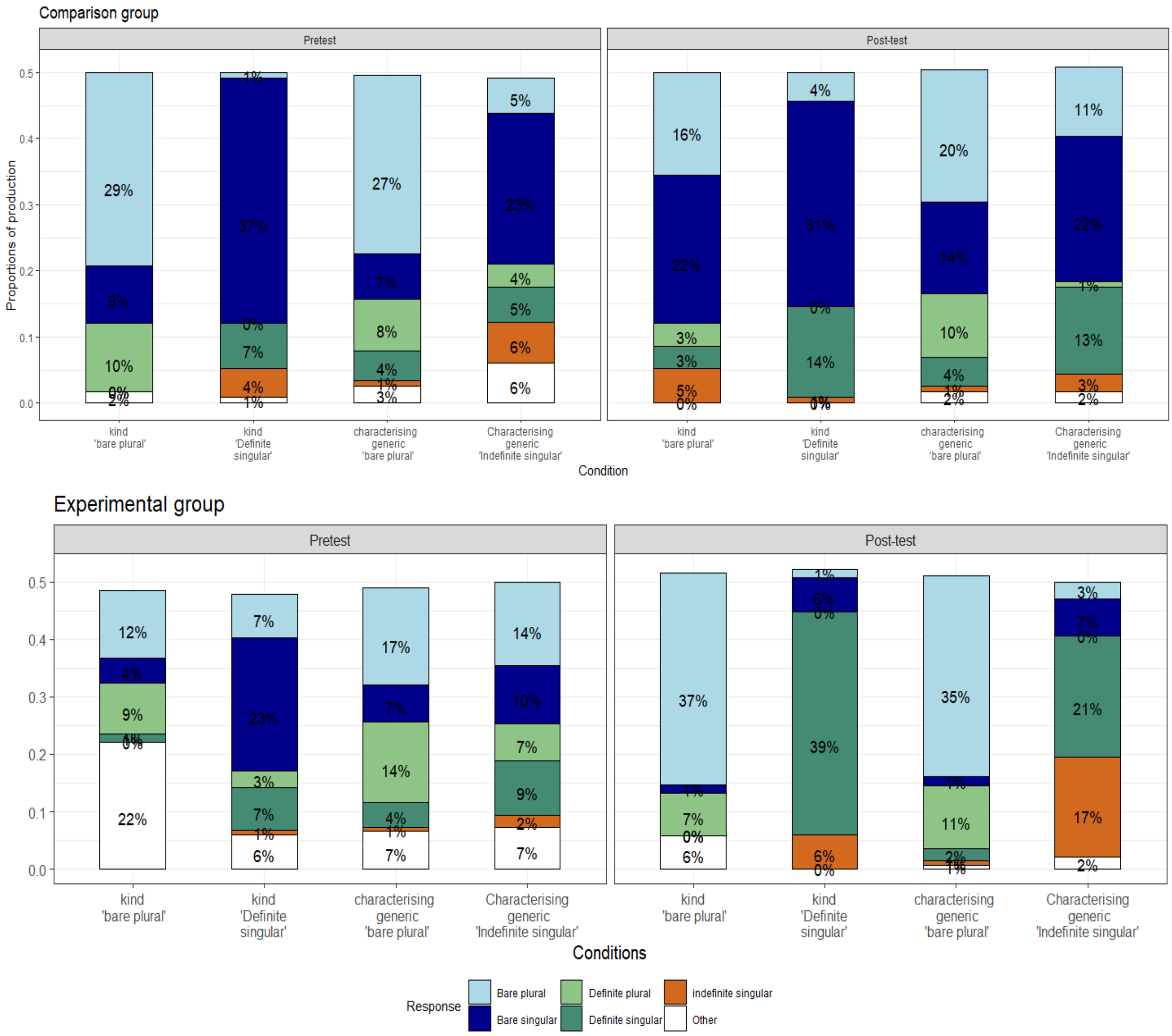


Figure 7.24 Proportions of producing different verb forms in the written production task in the pre-test and post-test by the experimental and comparison groups.

Figure 7.24 compares the proportions of all NPs produced by the experimental and comparison groups in the post-test and the pre-test. The Top half of the figure shows the proportions of the comparison group. The comparison groups showed approximately lower proportions of producing bare plurals in the post-test than in the pre-test. For mapping singular kind generic meaning to the definite singulars, they showed a higher proportion of definite singulars in the post-test. For singular characterising generic meaning, their production of the indefinite singulars was similar in the two tests. There was a notable increase in the production of bare singulars with singular kind generic meaning.

On the other hand, the bottom half of the figure showed a visible increase in the experimental group's proportions of producing the expected plural forms in the post-test with both generic contexts; bare plurals production increased from 12% to 37% in the plural characterising generic meaning and from 17% to 35% with plural kind generic meaning in the post-test. The experimental group's proportions of expected singular forms showed a similar increase in the post-test; the proportion of indefinite singulars increased from 2% to 17% for singular characterising generic meaning, and the proportion of definite singulars increased from 7% to 39% for singular kind generic meaning. The figure shows a notable decrease in the experimental group's use of bare singulars in both generic contexts.

As for the difference in proportions between the groups in the post-test, the figure shows that, after the intervention, the experimental group's correct production of the expected form proportion was higher than those of the comparison group in all contexts. In other words, while the experimental group produced the expected plural form in 37% of plural characterising generic contexts and in 35% of plural kind generic contexts, the comparison group's proportion of using the expected plural form in these contexts was 16% and 20%, respectively. The production of singular forms in both contexts showed a similar pattern. To further examine these differences, between-groups and within-group differences were analysed using multinomial logistic regression models. Three multinomial logistic regression models were constructed to analyse the post-test data, using the same procedures used in pre-test analysis. First, a model was used to compare the groups' differences in the post-test (between-groups difference). Then, a model was run for each group separately to test the improvement in the post-test compared to the pre-test (within-group difference). The maximal likelihood was followed for each model by using all possible predictors. Then, a model reduction was conducted using the *lrtest* function from the *lmtest* package to ascertain the significance of predictors to the model (Hua et al.,2021) and *anova* to compare the maximal model to the reduced models.

The differences between the experimental and comparison groups were tested to see if the groups significantly differed in the post-test production. The model structure included the response as the nominal categorical dependent variable with six levels of NPs, and the bare plural was set as the reference level. The fixed effects included the group, the condition, and their interaction. The interaction was kept in the model as the *anova* function results between the models showed that the interaction was significantly adding to the model (*LR* (15) = 47.26, *p* <.001). The final model was significant, with a better fit to the data, as shown by the likelihood ratio test results (X^2= 404.8, *df* = 5, *p* <.001). This model's results are presented in Appendix D.2, and Table 7.17 summarises post-hoc pairwise comparisons.

Table 7-17 Between-group Tukey-corrected pairwise comparisons of the post-test EWPT task for the two groups of L2 learners.

	Characterising generic meaning				**Kind generic meaning**			
Experimental to comparison	**Bare plural**		**Indefinite singular**		**Bare plural**		**Definite singular**	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Bare plural	11.60	.001*	6.13	.01*	12.35	.001*	1.89	.17
Definite plural	0.12	.73	1.02	.31	.96	.33	1.38	.24
Bare singular	15.92	.005*	15.39	<.001*	18.90	<.001*	46.58	<.001*
Definite singular	0.9	.33	3.82	.06	2.15	.15	35.33	<.001*
Indefinite singular	0.0	.89	21.21	<.001*	3.35	.07	5.42	.03*

Note: * significant at the .05 level

Table 7.17 shows all the related pairwise comparisons between the experimental and comparison groups in the post-test. What stands out in Table 7.17 is that the groups significantly differed in the post-test production of the correct form in each condition after the intervention, as shown by the shaded cells. In particular, the experimental group showed more accuracy in using bare plurals with both generic meanings and indefinite singulars with characterise generics. The experimental group's accuracy in producing the definite singular with kind generics was significantly higher than the comparison group. In addition, being in the experimental group was associated with a significant decrease in using the ungrammatical bare singulars in all conditions. Figure 7.25 presents the effect plot of group comparisons, which shows the difference between both groups after instruction.

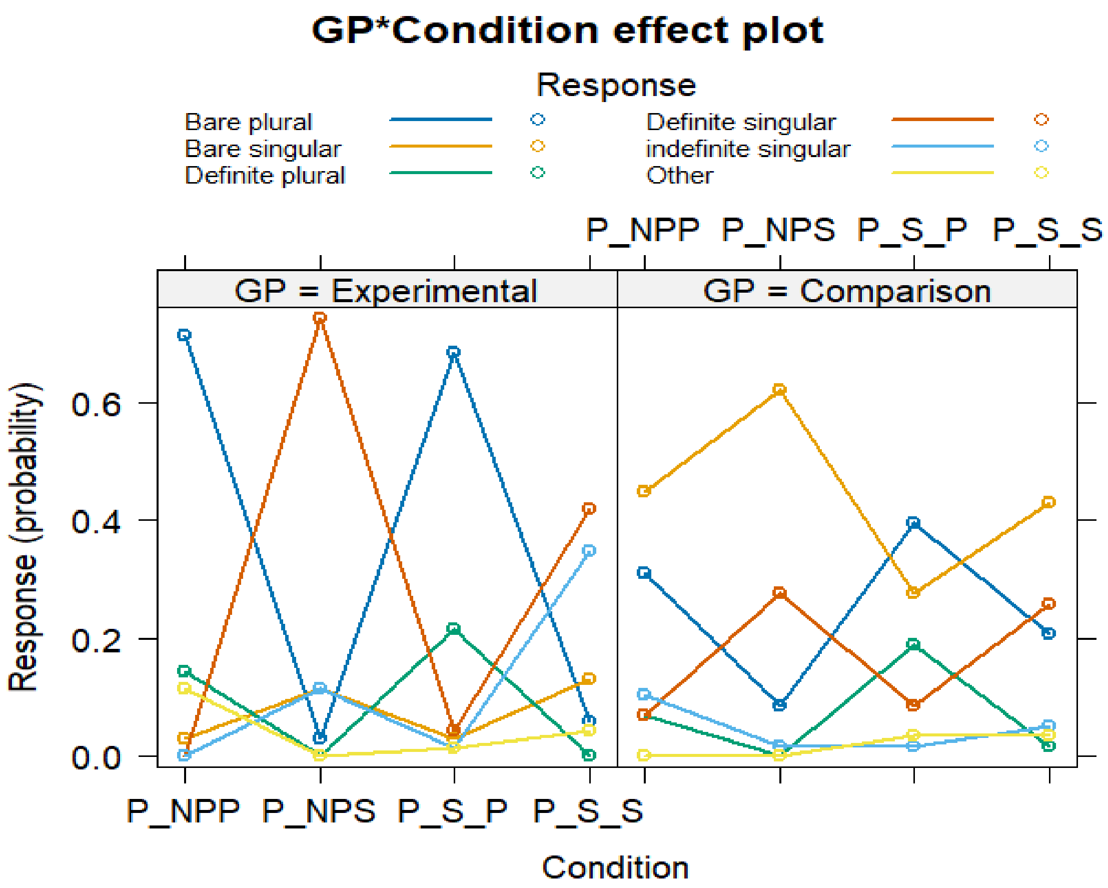


Figure 7.25 Effects plot for L2 learners' groups in the post-test by condition. [16]

Figure 7.25 shows the predicted effect for group differences. What stands out in this figure is that bare plural NPs, definite singular NPs, and bare singular NPs were the forms that showed the most apparent difference in probability between groups. To have a clearer view of the improvement in each group, I now turn to test within-group differences in the probability of using each NP in the post-test and their proportions in the pre-test for each group.

Starting with the comparison group, the comparison group's model started with the maximal structure using the interaction between the test and conditions as the fixed effect. The significance of adding the interaction was tested using the *anova* function, which revealed that adding the interaction was significant to the model (*LR* (15) =29.02, *p* =.02). The final model was compared to the null model using *the* likelihood ratio test *(lrtest* function) and the results showed that the final model was a significantly better fit (X^2 (5) =48.08, *p* <.001). The model results showed that the comparison group production of bare plural in the pre-test significantly

[16] P_NPP refers to plural forms in kind generics, P_NPS refers to the singular form in kind generics, P_S_P refers to the plural form in characterising generics, and P_S_S refers to the singular form in characterising generics.

differed from each NP type in the post-test except for producing definite plural NP (see Appendix D for the whole model). Further post hoc Tukey-corrected pairwise comparisons were conducted to ensure that the production was similar in both tests for each generic condition for each NP.

Table 7-18 Comparison group's pretest to post-test multinomial logistic regression model's Tukey-corrected pairwise comparisons.

	Characterising generic meaning				**kind generic meaning**			
	Bare plural		**Indefinite singular**		**Bare plural**		**Definite singular**	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Pre-test to post-test: Bare plural	.33	.57	1.17	.29	0.40	.53	2.09	.16
Pre-test to post-test: Definite plural	.86	.36	1.24	.27	0.82	.37	0.00	.99
Pre-test to post-test: Bare singular	2.21	.14	1.03	.32	2.60	.11	.563	.46
Pre-test to post-test: Definite singular	2.32	.21	7.72	.02*	1.18	.18	4.54	.04*
Pre-test to post-test: Indefinite singular	.40	.53	1.28	.26	0.49	.49	1.94	.18

Note: * Significant at the .05 level.

Table 7.18 presents the pairwise comparison results for the comparison group's production of different NPs forms in the post-test and pretest. It shows that the comparison group production in all NPs in the post-test was generally not significantly different from their production in the pre-test, except for their use of the definite singular (the L1 form) with both singular characterising and kind generic meanings. This means that the comparison group showed no improvement except that they produced a more target-like form in using definite singulars for singular kind generics. To ensure this is an improvement in the post-test, the comparison groups' production of definite singular with kind reference meaning was further compared to the production of all other singular NPs in this condition. Pairwise comparisons showed that definite singular production with kind generic meaning was significantly lower than the probability of producing the erroneous bare singular (F (40) = 15.75, p < .001). This can be interpreted as suggesting that the comparison group faced the same difficulty reported in the pretest results (section 7.2.1). To visualise the production of the comparison group in the pre-test and post-test, Figure 7.26 presents the effect plots before and after the intervention. What stands out in this figure is that the comparison group did not improve in the post-test compared to the pre-test; they encountered similar erroneous production with bare singular as the highest error and decreased accuracy of producing bare plurals in both contexts.

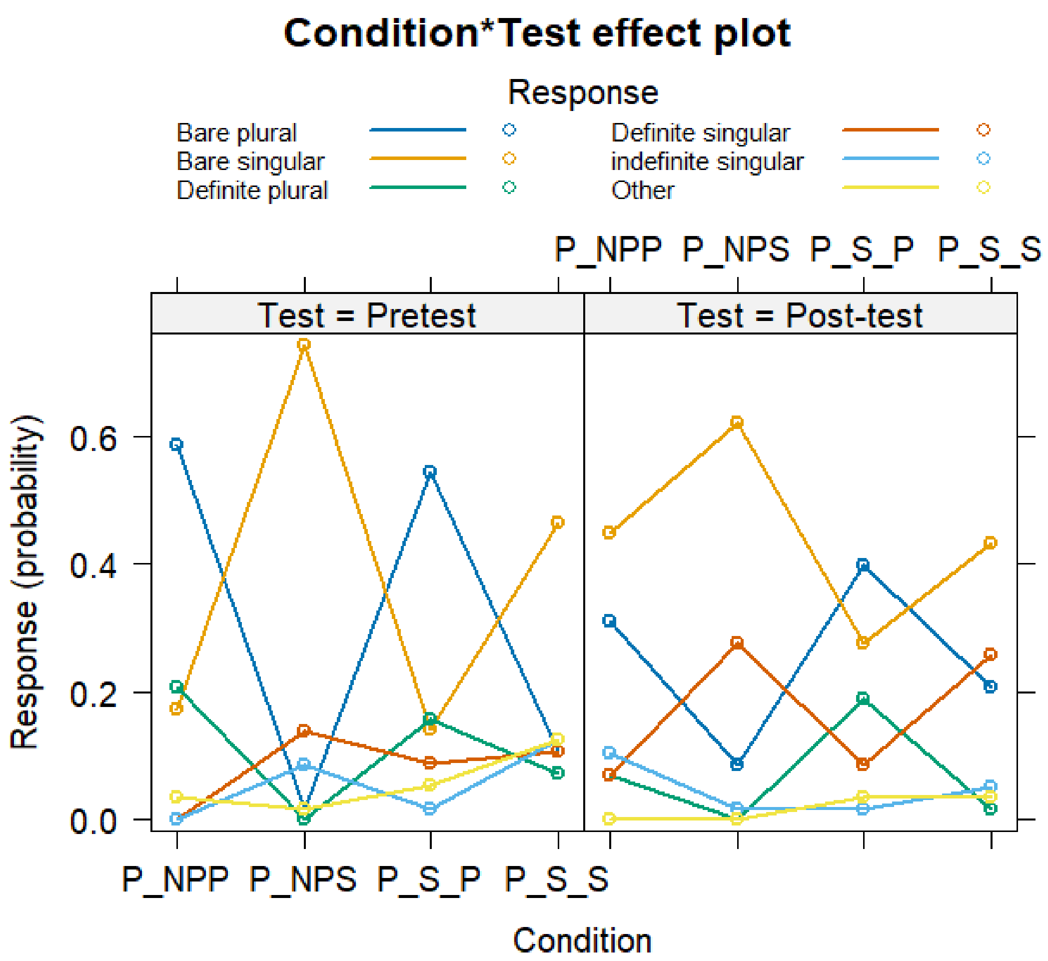


Figure 7.26 Effects plot for the use of different forms by the comparison group in the pre-test and post-test by condition.[17]

Turning to the experimental group, their responses in the post-test were modelled using the same procedures as used for modelling the data from the comparison group. The *test*context* interaction significance was tested using the *lrtest*, which revealed that adding the interaction was significant to the model (*LR* (15) =53.99, *p* <.001). The model was compared to the null model using the likelihood ratio test, and the results showed that the model with the fixed effects was significant with a better fit (X^2 (5) =416 *p* < .001). The construction of the model was followed with Tukey-corrected pairwise comparisons, the results of which are presented in Table 7.19.

---

[17] P_NPP refers to plural forms in kind generics, P_NPS refers to the singular form in kind generics, P_S_P refers to the plural form in characterising generics, and P_S_S refers to the singular form in characterising generics.

Table 7-19 Summary of Tukey-corrected pairwise comparisons of effects in a multinomial logistic regression model of the experimental group's use of each response in the pretest and post-test by condition

	Characterising	**generic**	**meaning**		**Kind**	**generic**	**meaning**	
contrast	**Target Bare**	**form plural**	**Target Indefinite**	**form singular**	**Target Bare**	**form plural**	**Target Definite**	**form singular**
	F	P	F	P	F	P	F	P
Pre-test to post-test: Bare plural	7.95	.009*	4.42	.05*	26.63	<.001*	1.88	.12
Pre-test to post-test: Definite plural	6.52	.02*	8.25	.008*	.59	.45	3.64	.06
Pre-test to post-test: Bare singular	2.63	.11	21.33	< .001*	0.19	.66	34.93	< .001*
Pre-test to post-test: Definite singular	9.42	.005*	18.88	< .001*	1.03	.31	48.66	< .001*
Pre-test to post-test: Indefinite singular	2.04	.16	23.43	< .001*	8.77	.006*	8.71	.006*

Note: * Significant at the .05 level.

Table 7.19 displays the pairwise comparison results of comparing the pre-test to the post-test production per each NP in each condition in the experimental group data. It showed that, regardless of the generic conditions, the production of bare plural NPs was significantly higher in the post-test compared to the pre-test, as shown in the shaded cells. This improvement is further confirmed by being significantly higher than the comparison group's production and by finding that bare plurals in both conditions were significantly higher than all other NPs in the post-test of both conditions, as shown in the effect plot in Figure 7. 27, below. To recapitulate, the pre-test results showed that the production of bare plurals was challenging for the experimental group in the pre-test as their production of bare plurals was statistically similar to their production of ungrammatical definite plurals and bare singulars. The post-test results showed that the L2 learners in the experimental group succeeded in increasing the use of bare plurals, which was significantly higher than both bare singulars and definite plurals ($F(40)=25.19$, $p<.001$) and when compared to definite plural NPs in the post-test ($F(40) = 54.33$, $p<.001$). The use of bare plurals with characterising generics showed a similar pattern in the post-test.

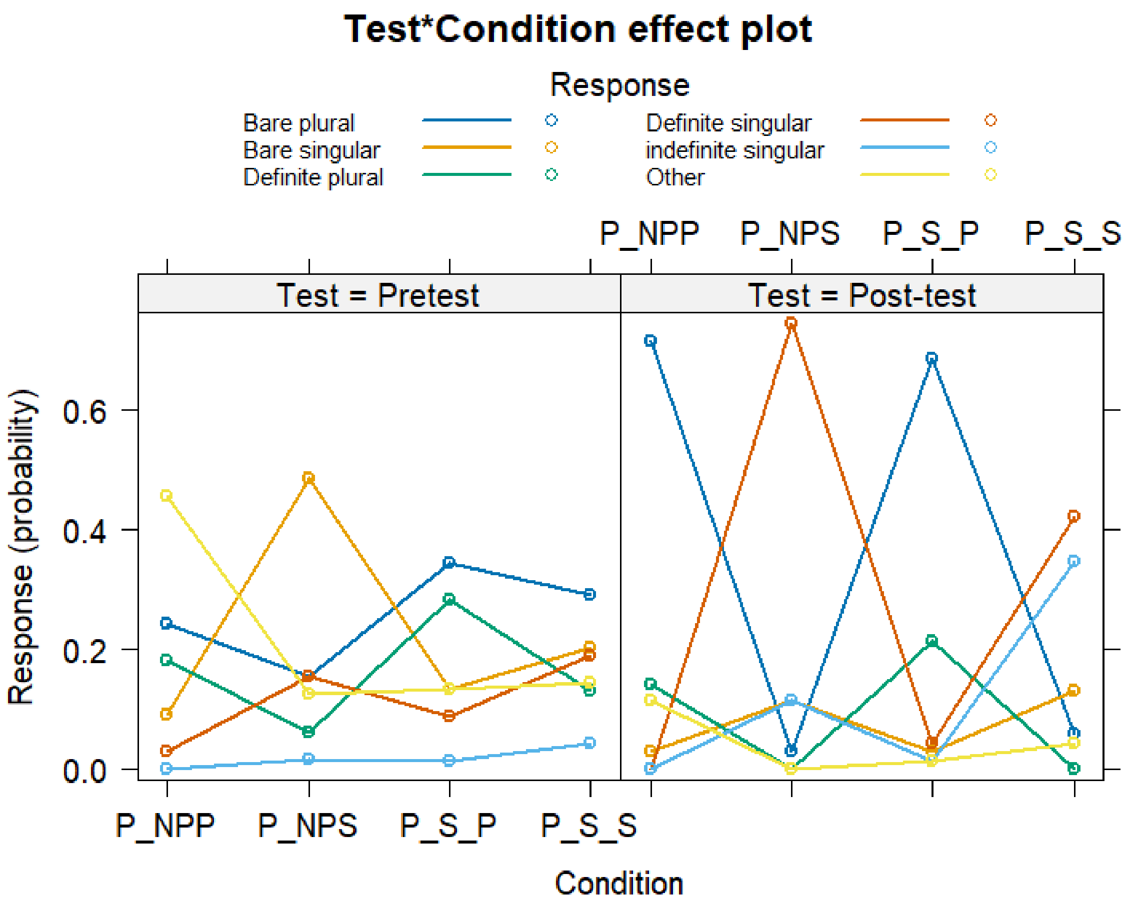


Figure 7.27 Effects plot for the use of different forms by the experimental group in the pre-test and post-test by condition. [18]

The experimental group's production of target-like definite singulars with kind generics and indefinite singulars with characterising generics significantly differed in the post-test when compared to the pre-test, as the shaded cells show in Table 7.19 above. Figure 7.27 demonstrates that this change constitutes a significant increase in producing expected singular forms in the post-test compared to the pre-test in both conditions and a significant decrease in

[18] P_NPP refers to plural forms in kind generics, P_NPS refers to the singular form in kind generics, P_S_P refers to the plural form in characterising generics, and P_S_S refers to the singular form in characterising generics.

producing the ungrammatical bare singulars with both characterising and kind generic singular conditions in the post-test.

It is also important to report the pairwise comparison results comparing the production of definite singulars and indefinite singulars to all other ungrammatical forms for each singular generic context in the post-test. Starting with mapping characterising generic meaning onto indefinite singulars, the pairwise comparisons showed that producing indefinite singulars was significantly higher than producing bare singulars (F (40) =17.60, p <.001). However, it was not significantly lower than definite singulars (F (40) =0.85, p =.36). This suggests residual difficulty in associating indefinite singulars with characterising generic meaning after the intervention.

Moving to consider whether the experimental group succeeded in producing the predicted singular form, the definite singular, after the intervention, Figure 7.27 showed that the probability of producing the definite singular was far higher than all ungrammatical forms in the post-test, which suggests that the experimental group improved in this condition after the intervention. Pairwise comparisons of the post-test data in this condition provide statistical evidence of this improvement. Production of definite singulars was significantly higher than all other ungrammatical forms in the post-test (F (40) =23.3, p <.001, for bare singulars; F (40) = 50.29, p <.001 for indefinite singulars).

In a nutshell, the experimental group's post-test results showed that they improved in the kind generics, as indicated by the successful use of the predicted singular and plural forms, which were significantly higher than the ungrammatical forms. They also improved in the production of the plural form in characterising generics. However, the results showed that using the singular form in characterising generics is still challenging even after instruction.

To assess whether this improvement was maintained 12 weeks after the post-test, the experimental group’s production in the post-test was compared to their production in the delayed post-test using a multinomial logistic regression model, following the structure used to compare the pretest and post-test above. Before presenting the model results, Figure 7.28 describes the proportions of each NP type in the experimental production.

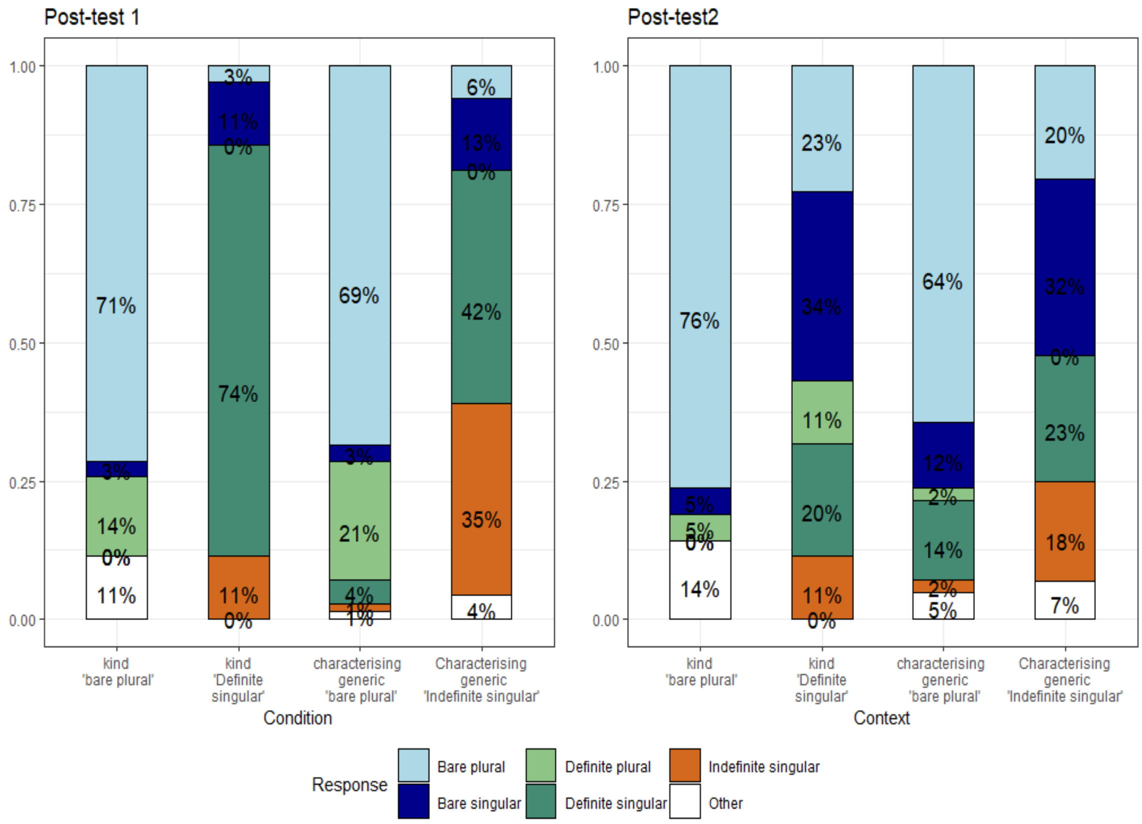


Figure 7.28 A comparison between the experimental group's delayed post-test and post-test proportions of production in the production task.

Figure 7.28 compares the experimental group's production rates for each form in the delayed post-test (post-test2) to their production in the post-test. A closer inspection of the figure shows that the experimental group's use of the bare plurals was similar in the post-test and delayed post-test with both kind and characterising generic contexts. However, there was a reduction in the production rates of the predicted singular form with kind generics in the delayed post-test compared to the post-test. These descriptive results indicate that the improvement in the post-test was maintained only for the plural form, i.e., bare plurals, with both kind and characterising generics. To test whether the improvement noticed in the descriptive analysis was statistically significant, the results of the multinomial logistic regression model follow.

The model structure included *response* as the dependent variable, and *test* and *context*, and their interaction, as the fixed effects. This model likelihood ratio test revealed that the model provided a significantly good fit to the data and that it is better than the null model (X^2= 313.74, *df* =5, *p* <.001). This model was not reduced as it was significantly better than the model with no interaction, as indicated by the results of the likelihood ratio test for multinomial models using *anova* (*LR* (15) = 39.91, *p* < .001). The model was followed with Tukey-corrected pairwise comparisons, which results are presented in Table 7.20.

Table 7-20 Summary of Tukey-corrected pairwise comparisons of effects in a multinomial logistic regression model of the experimental group's production task in the post-test and the delayed post-test.

	Characterising generic meaning				Kind generic meaning			
contrast	**Target Bare**	**form plural**	**Target Indefinite**	**form singular**	**Target Bare**	**form plural**	**Target Definite**	**form singular**
	F	P	F	P	F	P	F	P
Post-test to post-test2: Bare plural	0.22	.64	6.14	.02*	.16	.70	8.99	.004*
Post-test to post-test2: Definite plural	12.26	.001*	1.92	.17	1.60	.12	5.64	.02*
Post-test to post-test2: Bare singular	2.83	.10	5.36	.03*	0.12	.72	7.83	.007*
Post-test to post-test2: Definite singular	8.33	.007	4.38	.03*	11.39	.003*	22.25	< .001*
Post-test to post-test2: Indefinite singular	0.12	.73	4.13	.052	7.06	.01*	.00	.99

Note: * Significant at .05 level.

Table 7.20 displays the pairwise comparisons of predicted differences between each NP in the post-test and delayed post-test by the experimental group according to the regression model. The data in this table shows that the post-intervention increase in producing bare plurals with both kind and characterising generic contexts is maintained in the delayed post-test. The probability of using bare plurals in the delayed post-test was significantly similar to the probability of using bare plurals in the post-test with characterising and kind generic contexts, as shown in the shaded cells. Moreover, the use of bare plural was also significantly higher in the delayed post-test in contexts that target the use of singular forms, which were produced in these contexts by the native baseline. The experimental group showed a significant decrease in producing definite plurals to express characterising generic meaning, and the production of the definite plural was significantly lower than the correct production of bare plurals in the delayed post-test (F (40) = 23.3, p < .001), as shown in Figure 7.29. These results confirm the improvement in mapping bare plural to characterising generics as the L2 learners in the experimental group continued to prefer bare plurals to definite plurals, the L1 form, in the delayed post-test.

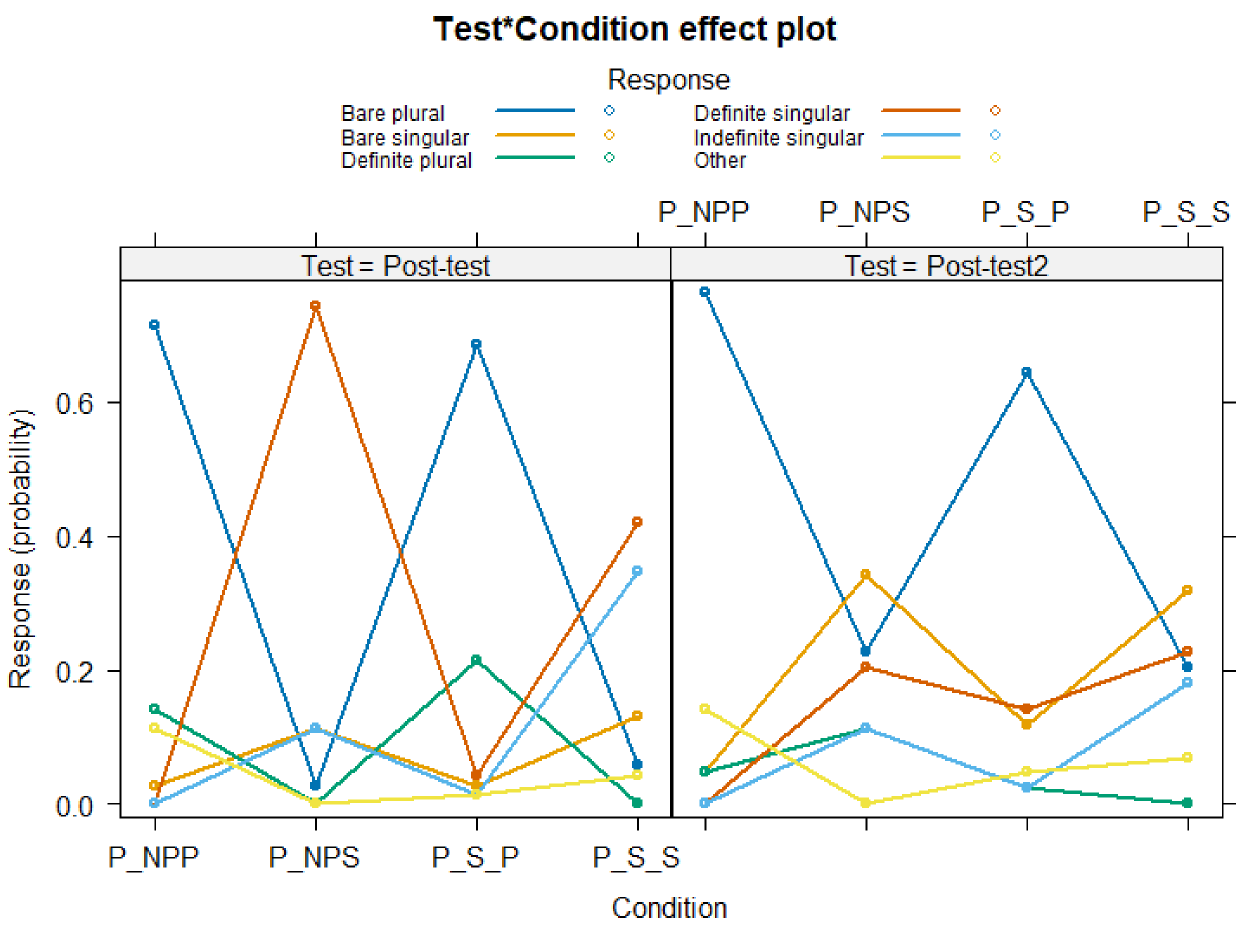


Figure 7.29 Effects plot showing the use of different forms by the experimental in both post-tests by condition. [19]

[19] P_NPP refers to plural forms in kind generics, P_NPS refers to the singular form in kind generics, P_S_P refers to the plural form in characterising generics, and P_S_S refers to the singular form in characterising generics.

As for the predicted singular forms, producing indefinite singulars for characterising generic meaning did not show improvement in the post-test, and the production of these NPs was significantly decreased in the delayed post-test, as the shaded cell shows in the table above. For kind generics, the probability of using definite singulars decreased significantly in the delayed post-test compared to the post-test (see the shaded cell in the table above). In addition, using bare singulars in production was significantly higher in the delayed post-test for this condition.

In sum, after the intervention, the experimental group showed significantly higher accuracy in producing the expected form in all conditions but the indefinite singular for characterising generic meaning in the immediate post-test. However, twelve weeks later, they maintained the improvement only in producing bare plurals in both types of generic meaning.

7.3.2 AJT post-tests Results

Turning now to the post-test AJT, an overall description of the results in each generic context is presented. Then, the groups' mean ratings in each condition after the intervention were statistically compared to determine whether there were any differences between the groups in the post-test. After that, the results of the within-group changes in the post-test compared to the pre-test ratings are presented for each group separately. Finally, the results on whether, 12 weeks later, the experimental group retained the post-test improvement are presented.

In terms of descriptions, Figure 7.30 compares the experimental and comparison groups' rating distributions and mean ratings in kind generic context.

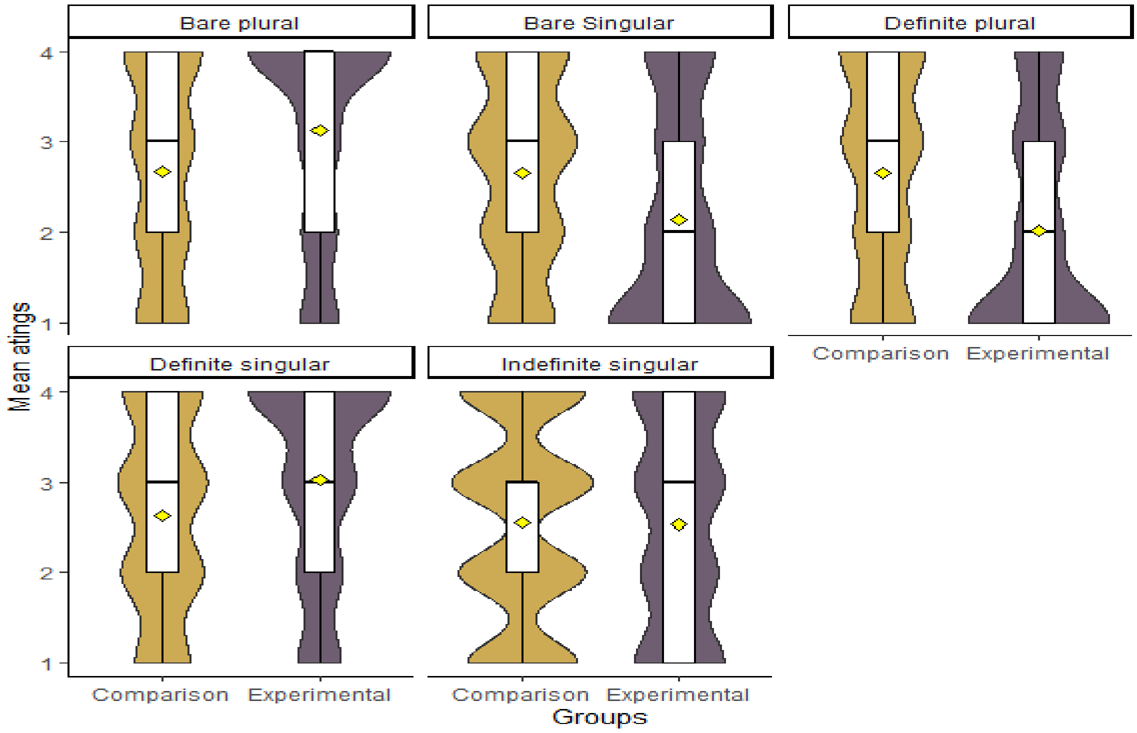


Figure 7.30 Groups' ratings distributions and means in the AJT post-test in kind generics. The yellow diamond shape indicates the group means for each verb form in the condition.

Considering kind generics, while the experimental group's mean rating of the acceptable forms, bare plurals and definite singulars, was (*M* = 3, *SD* = 1.1), the comparison group's mean rating for these acceptable forms was (*M* = 2.6, *SD* =1.1). Moreover, the experimental group's mean rating for the unacceptable forms in this generic context (bare singulars and definite plurals) was (*M* = 2.1, *SD* = 1.1) for each condition, which was lower than the comparison group's ratings (*M*= 2.6, *SD* = 1.1) for each. Both groups rated the indefinite singular in kind generics similarly low (*M* =2.5, *SD* =1.1 for each group). The experimental group's ratings of acceptable NPs in kind generics were higher than those of the comparison group; their ratings of the unacceptable forms were lower than those of the comparison group, with the exception of the indefinite singular form where both groups had the same mean rating.

If we now turn to describe the group ratings in the characterising generic context, the distributions and mean ratings are plotted in Figure 7.31.

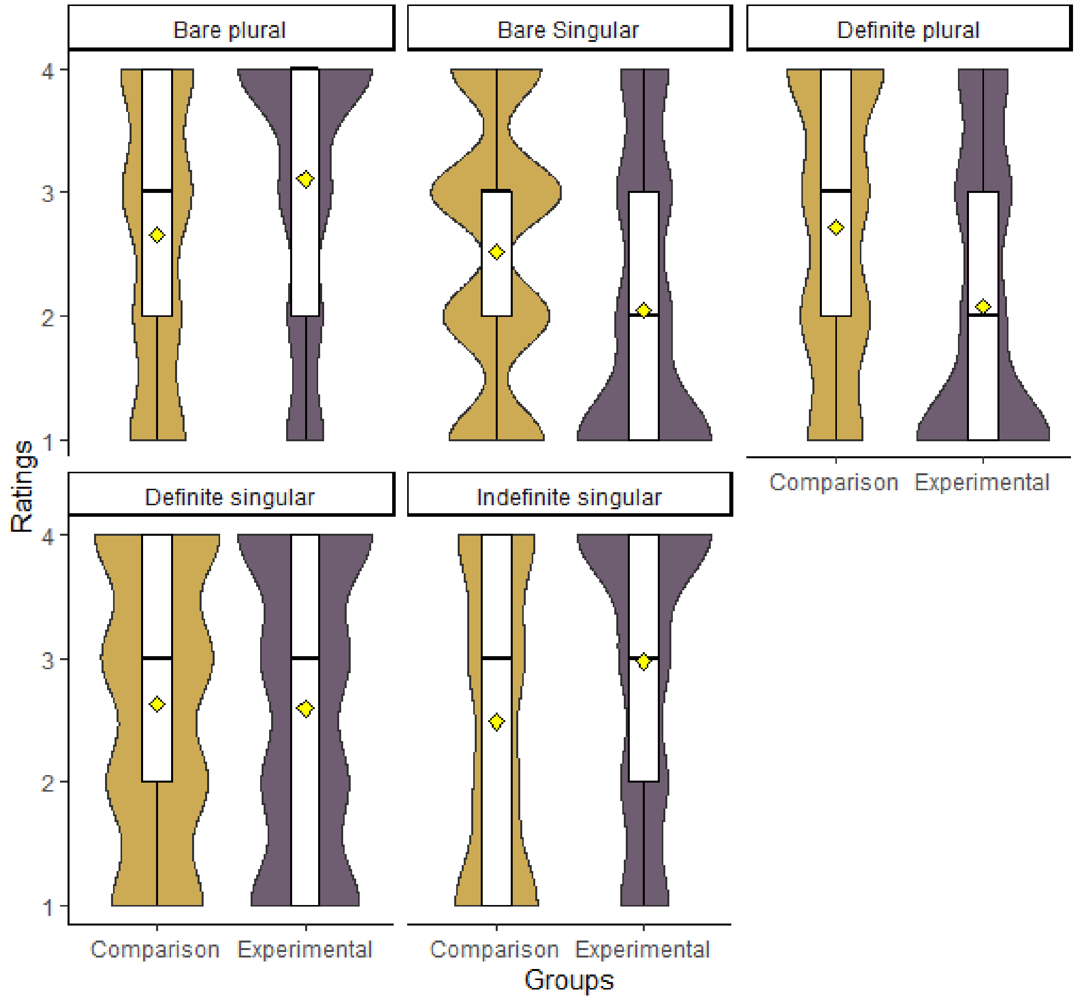


Figure 7.31 Rating distributions and means in the AJT post-test characterising generic context by condition. The yellow diamond shape refers to the mean rating.

Figure 7.31 displays both groups' ratings' distributions and means in the characterising generic context. What can be seen is that the experimental group's mean rating of the acceptable forms, bare plurals and indefinite singulars, were (*M* = 3.10, *SD* = 1.1) and (*M* = 2.97, *SD* = 1.2), respectively. The comparison group's mean ratings of the bare plural and indefinite singular were (*M* = 2.62, *SD* = 1.1) and (*M* = 2.48, *SD* = 1.2, respectively). The experimental groups' mean ratings in the acceptable conditions were higher than their mean ratings in other conditions, and they were higher than the comparison group's mean ratings in these conditions. As for the unacceptable forms, the experimental group's mean rating of bare singulars and definite plurals in characterising generics were (*M* = 2.04 and *M* = 2.07, *SD*= 1.1 respectively), which were lower than their mean ratings in other conditions as well as the comparison groups' mean ratings in these conditions (*M* = 2.5, *M* = 2.7, *SD*= 1.1 respectively). Both groups had similar ratings regarding using definite singulars in characterising generic contexts (*M* = 2.6, *SD* =1.1 in both groups). Overall, the descriptive results suggest that the experimental and comparison groups differed in their ratings of conditions in each generic context after the intervention. However, these results are merely descriptive and require further statistical analysis.

Considering the results of the difference between the experimental and the comparison groups in the post-test, a CLMM was constructed to compare the groups' ratings per context. For each context, the model structure included starting with a model that consists of *ratings* as the ordinal dependent variable, the *group* (with two levels: experimental and comparison), the *conditions* (five NPs) and the interaction between the *group* and *conditions* as fixed effects, with by-participant and by-item random intercepts. Then, the model was reduced using a single-term deletion test, which showed a significant impact of the interaction between group and condition in both generic context's models (χ^2 = 88.18, *df* = 4, *p* =<.001) in the characterising generic conditions model, and (χ^2 = 87.55, *df* = 4, *p* =<.001) in kind generics model. Therefore, the final model in each context included the interaction between *groups* and *conditions* as the fixed effect and the random effects described above (see Appendix D). Tukey-corrected post hoc pairwise comparisons followed each model. In each model, the *comparison×bare plural* interaction was the reference level.

For characterising generic contexts, the model results revealed a significant difference between the comparison and experimental groups in their mean ratings of the acceptable conditions; for the bare plural condition, the results suggest that the experimental group rated bare plurals significantly higher than the comparison group (*ES* =−7.98, *SE* =2.27, *z* = −3.05, *p* = <.001). They did so in the indefinite singular condition (*ES* =-8.27, *SE* = 0.23, *z* =-3.59, *p* = 0.01). As for unacceptable conditions, the post hoc pairwise comparisons revealed that the groups differed significantly in their ratings of the bare singular condition (*ES* =7.98, *SE* =.23, *z* = 3.56, *p* = .01), and the definite plural condition (*ES* =1.12, *SE* =.23, *z* =4.92, *p* =<.001). This suggests that the experimental group rated the bare singular and the definite plural significantly lower than the comparison group. There was no significant difference in the group's ratings of the definite singular in the post-test (*ES* =7.89, *SE* =.22, *z* =.35, *p* =1). These results suggest a positive effect of the intervention in accepting acceptable NP conditions in characterising generics.

As for kind generics, the model results revealed a significant difference between the comparison and experimental groups in their mean ratings of the bare plural condition (*ES* =.86, *SE* =.21, *Z* = 3.98, *p* = <.001). Interestingly, there was no significant effect of the interaction between group and condition in the definite singular condition (*ES* = -0.18, *SE* = 0.27, *z* =-0.68, *p* = 0.49). As for the unacceptable conditions, there was a significant difference between the groups in these conditions except for the indefinite singular. Considering the bare singular (*ES* = 0.86, *SE* = 0.21, *z* = 4.08, *p* = .001) and definite plural (*ES* = 1.12, *SE* = 0.21, *z* = 5.2, *p* <.001), being in the experimental group was associated with a significantly lower rating in these conditions. However, the groups rated the indefinite singular similarly low in the post-test (*ES* = 0.03, *SE* = 0.21, *z* = 0.13, *p* =1).

To summarise, the post-test results highlighted some significant differences between the comparison and experimental groups in the AJT. The experimental group rated the bare plurals significantly higher than the comparison group in both characterising and kind generics. They also rated definite plurals and bare singulars significantly lower than the comparison group in the characterising generics. In kind generics, the groups differ in their ratings of definite plurals, bare singulars, and bare plurals. The experimental group rated the acceptable bare plurals significantly higher and the bare singular and definite plural significantly lower than the comparison group.

Having examined the results in terms of group differences, I now move on to analysing how each group's ratings change in the post-test compared to the pre-test. To understand how each group rated the five conditions in the post-test and whether any with-in-group differences described above were significant, CLMMs were constructed. A separate model was fitted for each group in each generic context, and post hoc Tukey-corrected pairwise comparisons were conducted for each model. As for the model structure, in all CLMMs, each model started with a model that included the Test (with two levels: pre-test and post-test), the conditions (five NPs), and the interaction between the test and conditions as fixed effects, with by-participant and by-item random intercepts. Then, each model was reduced using a single-term deletion test.

Starting with the comparison group's ratings in the post-test, Figure 7.32 displays the mean ratings by condition, comparing the pre-test to the post-test in each generic context.

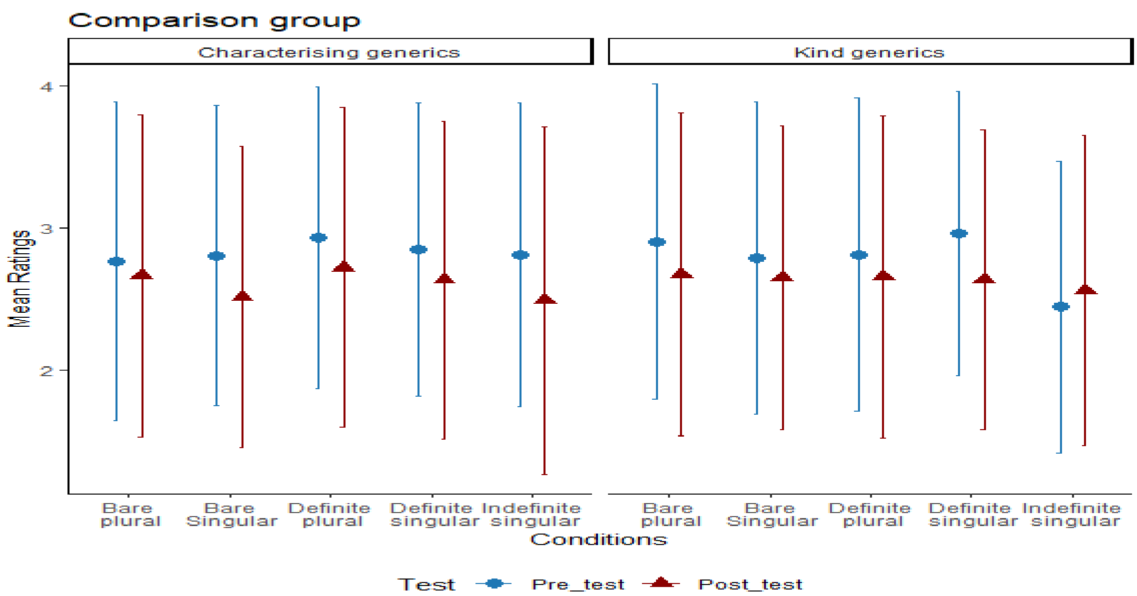


Figure 7.32 Comparison group's pre-test and post-test mean ratings in both contexts.

Looking at Figure 7.32, it is evident that there was not a noticeable change in the mean ratings of conditions in the post-test as compared to the pre-test, except that the post-test mean ratings were generally slightly lower than the pre-test mean ratings in both contexts. Considering the characterising generic context, the single-term deletion test for the CLMM model revealed no significant effect of the interaction between the test and conditions (χ^2 = 1.60, *df* = 4, *p* = .81). Therefore, none of the differences in the mean ratings in characterising generic in the figure above were significant. The model revealed that the difference between the Tests as a fixed effect was not significant, suggesting that there was no difference in the comparison group's mean ratings between the pre-test and post-test (*OR* = 1.21, *ES* =.19, *SE* = .19, *z*= .95., *p* = .34) (sees Appendix D for the model). Moreover, post hoc pairwise comparisons revealed no significant differences between the comparison group's mean ratings of each NP type in post-test and pre-test in the characterising generics, as shown in Table 7.21.

Table 7-21 Post hoc comparisons in the characterising generic context NPs mean ratings in the post-test compared to the pre-test for the comparison group.

Contrast	**Estimate**	**Standard error**	**Z ratio**	**p. value**
Post-test to pre-test [Bare plural]	−0.18	0.20	−0.96	.99
Post-test to pre-test [Indefinite singular]	−0.50	0.20	−2.50	.27
Post-test to pre-test [Definite singular]	−0.31	0.20	−1.61	.85
Post-test to pre-test [Definite plural]	−0.33	0.20	−1.69	.80
Post-test to pre-test [Bare singular]	−0.46	0.19	−2.39	.33

As for kind generics, the kind generic context model revealed no significant effect of the interaction between the test and conditions (χ^2 = 7.75, *df* = 4, *p* = .10). This means that there was no significant difference between each condition pre-test and post-test mean rating in the comparison group's rating in the kind generic conditions. Although the *test* effect coefficients were significant in the model (*OR* = 1.52, *ES* =.41, *SE* = .19, *z*= 2.08, *p* = .04), the comparison group's mean ratings were lower in the post-test for all conditions. This estimate of the *test* as a fixed effect considered the overall test data without taking the conditions into account, and did not reflect any improvement. Post-hoc pairwise comparisons provided the statistical evidence for this. They revealed no significant difference between the pre-post mean ratings in each condition at the 95% significance level, as shown in Table 7.22.

Table 7-22 Post hoc comparisons in the kind generic context NPs mean ratings in the post-test compared to the pre-test for the comparison group.

Contrast	**Estimate**	**Standard error**	**Z ratio**	**p. value**
Post-test to pre-test [Bare plural]	−0.42	0.20	−2.08	.53
Post-test to pre-test [Indefinite singular]	0.16	0.19	−0.89	.99
Post-test to pre-test [Definite singular]	−0.53	0.20	−2.78	.14
Post-test to pre-test [Definite plural]	−0.26	0.20	−1.26	.96
Post-test to pre-test [Bare singular]	−0.25	0.20	−1.26	.96

Considering the experimental groups' ratings before and after the teaching intervention, Figure 7.33 displays the mean ratings in each context, comparing the pre-test to the post-test.

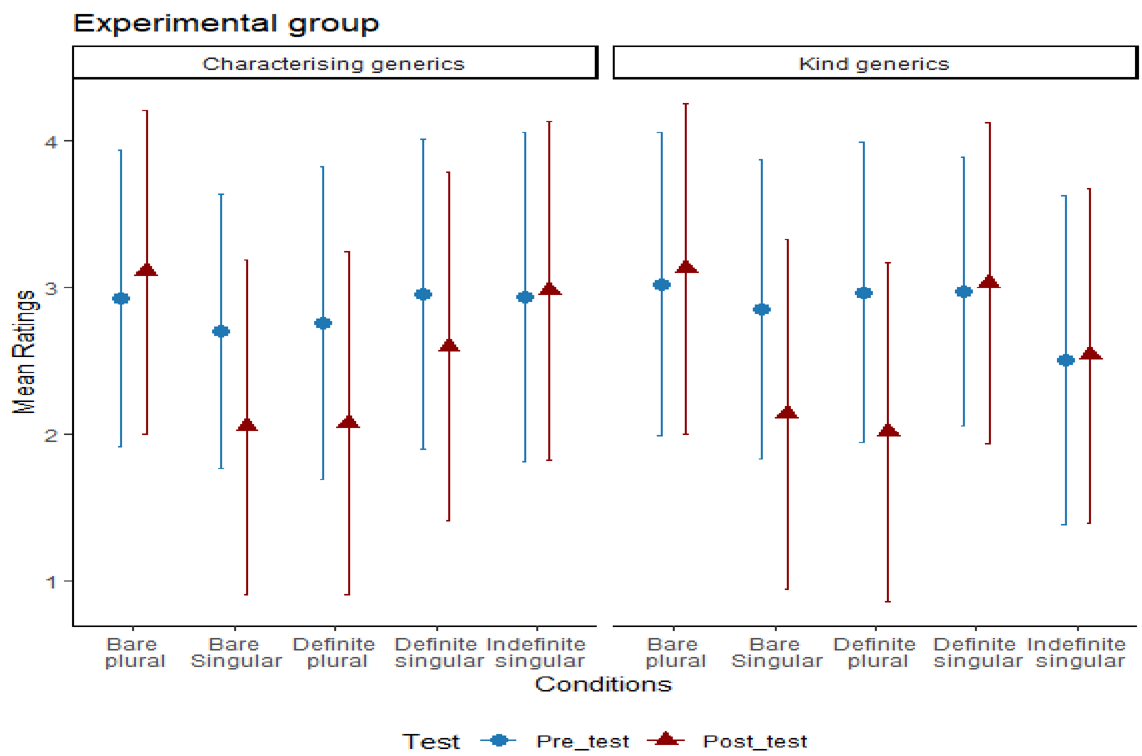


Figure 7.33 The experimental group's pre-test and post-test mean ratings in both contexts.

Closer inspection of the figure shows that, in the characterising generic context, the mean ratings of the acceptable NP types, namely bare plurals and indefinite singulars, showed a slight increase, and the mean ratings of unacceptable NPs, namely bare singulars, definite singulars, and definite plurals, decreased in the post-test as compared to the pre-test. Further analysis using CLMM for the characterising generic context revealed a significant interaction effect for test with condition, meaning that these differences are worth further investigation (χ^2 = 58, *df* =

4, *p* <.001). Table 7.23 presents the model estimate, odds ratio, standard error, and z- and *p*-values.

Table 7-23 Experimental group characterising generics pre-post comparison model estimate, odd ratios, *z,* and *p*- values.

Conditions	**Odds ratios (CI)**	**Estimate**	**St. Error**	**Z value**	***p-value***
Test [pre-test]	.67 (0.47 – 0.96)	−0.3997	0.18	−2.19	.03*
Conditions [Bare singular]	.16 (0.11 – 0.23)	−1.85	0.19	−9.78	< .001*
Conditions [Definite plural]	.16 (0.11 – 0.23)	−1.83	0.19	−9.66	< .001*
Conditions [Definite singular]	.40 (0.28 – 0.58)	−0.91	0.18	−4.95	< .001*
Conditions [Indefinite singular]	.80 (0.55 – 1.15)	−0.23	0.18	−1.216	.22
Test [Pretest] × Conditions [Bare Singular]	4.30 (2.62 – 7.05)	1.46	0.25	5.76	< .001*
Test [Pretest] × Conditions [Definite plural]	4.71 (2.85 – 7.79)	1.5496	0.2570	6.029	< .001*
Test [Pretest] × Conditions [Definite singular]	2.61 (1.59 – 4.31)	0.96	0.25	3.77	<.001*
Test [Pretest] × Conditions [Indefinite singular]	1.35 (0.81 – 2.24)	0.30	0.25	1.16	0.24

Note: Model: clmm (Ratings~Test*Conditions+(1|id) +(1|item), data = EXPrepstch Hess=T, threshold = "flexible"), Observations, 2099; Number of groups: id 35, item 6. *Significant at the .05 level.

The data in this table presents the results of the CLMM for the experimental group's rating for the characterising generic context in the post-test and compares the post-test ratings to the pre-test ratings. Considering the ratings in each condition in the characterising generics, the results reveal that the experimental group's post-test mean ratings of the bare plural condition (the reference level) were significantly higher than all the post-test conditions except the acceptable indefinite singular condition, where the difference in ratings was not significant.

Moreover, Tukey-corrected post-hoc pairwise comparisons revealed that the indefinite singular (acceptable form) mean ratings were significantly higher than the bare singular (*ES* = 1.61 *SE* = 0.18, *z* = 8.63, *p* < .001), definite plural (*ES* = 1.60 *SE* = 0.18, *z* = 8.51, *p* < .001) and definite singular (*ES* = 0.69, *SE* = 0.18, *z* = 3.73, *p* = .007), all of which are not acceptable forms. This indicates that the experimental group ratings for the acceptable NP types, i.e., the bare plural and indefinite singular, were significantly higher than the unacceptable forms in the characterising generic condition in the post-test. The experimental group participants successfully similarly gave low ratings for the unacceptable forms in the post-test as there was no significant difference between the mean ratings of the bare singular and definite plural (unacceptable NP types in the post-test, *p* >.05). The mean ratings of the definite singular NP in the post-test were significantly higher than the other unacceptable forms, yet significantly lower than the acceptable forms. For the intervention's impact in characterising generics, Table 7.24 compares the average ratings of each condition in the post-test to those of the same condition in the pre-test (the post hoc pairwise comparisons).

Table 7-24 The experimental group's post-test and pre-test pairwise comparisons in characterising generic context.

Contrast	**Estimate**	**SE**	**Z ratio**	***p*.value**
Post-test to pre-test [bare plural]	0.39	0.18	2.19	.45
Post-test to pre-test [bare singular]	−1.06	0.17	−6.03	<.001*
Post-test to pre-test [definite plural]	−1.15	0.18	−6.35	<.001*
Post-test to pre-test [definite singular]	−0.56	0.18	−3.15	.05
Post-test to pre-test [indefinite singular]	0.10	0.18	0.54	.99

Note: * Significant at the .05 level.

The table above showed that the post-test mean ratings of the unacceptable forms (definite plurals and bare singulars) were significantly lower than the pre-test rating means for these conditions, as shown by the negative estimates. The mean rating for the definite singular condition decreased, too, but this change was not significant. From This, I can conclude that in the characterising generic context, the experimental group successfully rated the acceptable NPs higher and the unacceptable NPs lower after the teaching intervention, reflecting the intervention's positive impact.

Turning now to consider the kind generic context, Figure 7.33 (above) showed that the experimental group's mean ratings of the acceptable NPs (bare plural and definite singular) showed a slight increase in the post-test, and their mean ratings of the unacceptable bare singular and definite plural NPs clearly decreased in the post-test. Of interest here is that the experimental group rated the indefinite singular as unacceptable, in both the post-test and pre-test (*M* = 2.5 in both tests). Further analysis used a CLMM for the kind generics and the model structure included rating as the dependent variable, the interaction between tests and conditions as the fixed effects, and the random by-item and by-participant intercepts. This model revealed a significant impact of *test*condition* interaction, meaning that these differences are worth further investigation (χ^2 = 96.3, *df* = 4, *p* =<.001). Table 7.25 displays the kind generics pre-test and post-test comparison model estimates, standard errors, odd ratios, and *z*- and *p*-values.

Table 7-25 Kind generic context pre-post-tests comparison model estimates, standard errors, odd ratios, z- and *p* values for the experimental group data.

Predictors	**Odds ratios (CI)**	**Estimate**	**St. Error**	***z* value**	***p-value***
Test [Pretest]	0.73 (0.51 – 1.05)	−0.31	.18	−1.68	.09
Conditions [Bare Singular]	0.16 (0.11 – 0.24)	-1.80	.19	−9.44	<.001*
Conditions [Definite plural]	0.13 (0.09 – 0.19)	−2.04	.19	−10.58	<.001*
Conditions [Definite singular]	0.77 (0.53 – 1.11)	−0.26	.18	−1.39	0.16
Conditions [Indefinite singular]	0.33 (0.23 – 0.47)	−1.11	.18	-6.02	<.001*
Test [Pretest] ×Conditions [Bare Singular]	4.41 (2.65 – 7.32)	1.48	.25	5.72	<.001*
Test [Pretest] ×Conditions [Definite plural]	6.88 (4.12 – 11.49)	1.92	.26	7.36	<.001*
Test [Pretest] × Conditions [Definite singular]	1.10 (0.67 – 1.82)	0.09	.25	0.38	.70
Test [Pretest] ×Conditions [Indefinite singular]	1.29 (0.78 – 2.13)	0.25	.25	0.99	.32

Note: Model: clmm (Ratings~ test*Conditions+(1|id) +(1|item), data = KGENPST,Hess=T, threshold = "flexible"), Observations, 1918; Number of groups:2; id 64, item 6. *Significant at the .05 level.

This table displays the results of the CLMM used to analyse the experimental group's post-test ratings compared to their pre-test AJT ratings. The results revealed that after the teaching intervention, the experimental group participants' mean rating of the bare plural condition was significantly higher than that of all the unacceptable conditions (bare singulars, definite plurals, and indefinite singulars) in the post-test. Interestingly, the mean rating of the bare plural was statistically similar to the mean rating of definite singulars (the mean ratings of both acceptable forms were high in the post-test). A closer examination of the post-hoc pairwise comparison showed that the experimental group's mean rating of the definite singular NP was significantly higher than their ratings of all unacceptable forms, including bare singulars (*SE*= −1.54, *SE* = 0.18, *z* = −8.27, *p* < .001), indefinite singulars (*ES* = −0.8, *SE* =0.18, *z* = −4.27, *p* <.001) and definite plurals (*ES* = −1.78, *SE* =0.18, *z* = −9.47, *p* <.001). To compare the difference between the experimental group's pre-test and post-test mean ratings, the related post hoc Tukey-corrected pairwise comparison results are tabulated in Table 7.26.

Table 7-26 The experimental group's kind generics AJT post-test ratings to pre-test ratings pairwise comparisons results.

contrast	**Estimate**	**SE**	**z. ratio**	***p***
Post-test to Pre-test [Bare plural]	0.31	0.19	1.68	.80
Post-test to Pre-test [Bare Singular]	−1.17	0.18	−6.52	< .001*
Post-test to Pre-test [Definite plural]	−1.61	0.18	−8.81	< .001*
Post-test to Pre-test [Definite singular]	0.22	0.18	1.22	.97
Post-test to Pre-test [Indefinite singular]	0.0579	0.176	0.330	1

Note: *significant at .05 level.

Table 7.26 displays post-hoc pairwise comparisons of the experimental group's mean pretest ratings compared to the post-test in kind generics. What can be seen in this table is that, after the teaching intervention, the mean ratings of the acceptable conditions (Bare plural and definite singular) slightly increased in the post-test as compared to the pretest. However, this increase is not statistically significant. Looking at the unacceptable NP types, the mean ratings of the unacceptable bare singular and definite plural conditions significantly decreased in the post-test as compared to the pre-test in kind generics. The indefinite singular maintained the low ratings in the post-test.

To establish whether the experimental group retained the improvement in the generic form-meaning mapping reported above, this section reports how this group rated the same AJT in the delayed post-test, which took place 12 weeks after the first post-test. In terms of description, we report the distributions and mean ratings of the experimental group in both post-tests per generic context.

Concerning characterising generics, the experimental group showed improvement in the post-test. As a reminder, they rated the acceptable conditions significantly higher than the unacceptable ones. It is essential to see if the experimental group retained this improvement three months after the post-test. Figure 7.34 compares the distributions and mean ratings of the post-test and the delayed one (post-test2).

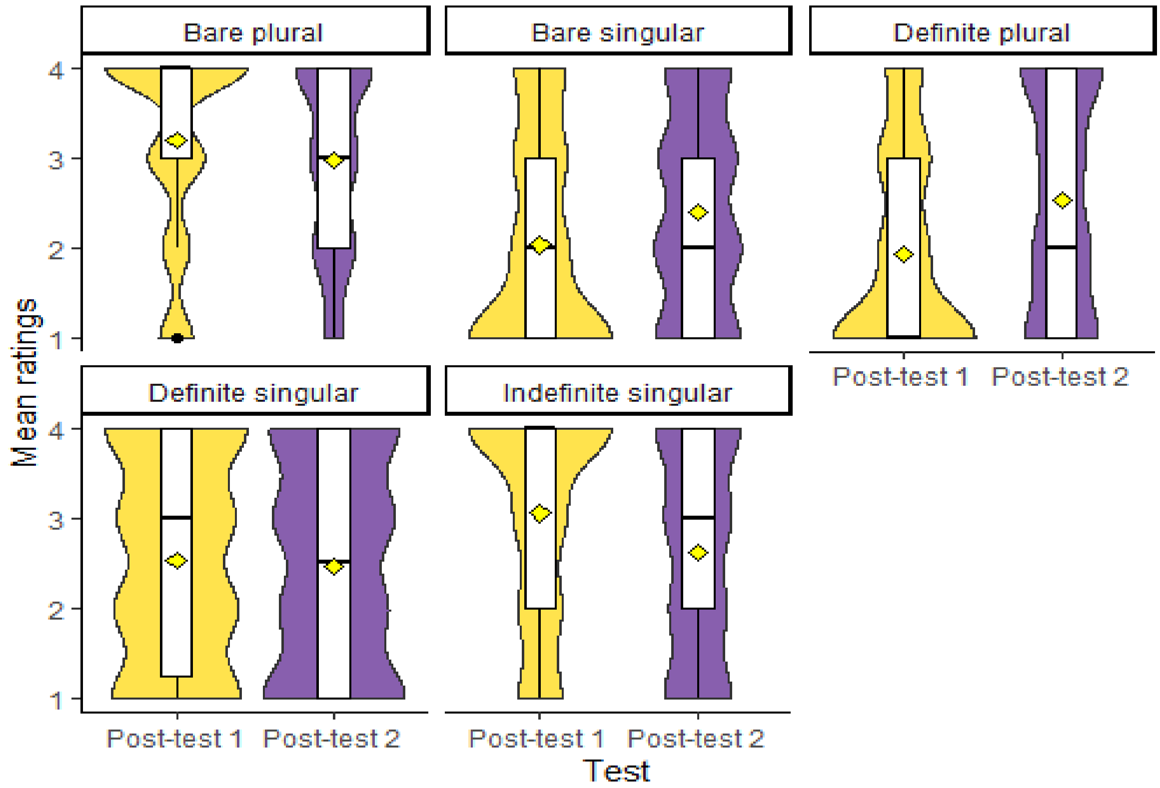


Figure 7.34 Rating distributions and means in the experimental group's AJT post-tests characterising generic context by condition. The yellow diamond shapes represent the group's mean rating in each condition.

This figure shows the experimental group's distributions and mean ratings in the delayed post-test compared to the post-test ones in characterising generics. Examining the figure, the experimental group's mean ratings of the acceptable conditions (bare plural and indefinite singular) decreased in the delayed post-test (post-test 2) compared to the first post-test. In addition, their mean ratings of unacceptable conditions (definite plural and bare singular) increased in post-test 2. Interestingly, both tests showed similar distributions of their ratings in the definite singular condition.

To further understand this change, a CLMM was constructed to test the experimental group's ratings of the generic conditions in the delayed post-test (between-conditions differences in post-test 2) and to compare these ratings in both post-tests (between-tests differences). The model included rating as the dependent variable and conditions, test, and their interaction as fixed effects. In addition, it included by-participant and by-item random effects. Using the *drop1* function to conduct a single-term deletion revealed a positive impact of the interaction between *conditions* and *test*, (χ^2 =47.19, *df* = 4, *p* <.001). Therefore, the final model included this interaction as a fixed effect in addition to the random effects. Table 7.27 shows the model results.

Table 7-27 Characterising generic context post-tests comparison model estimates, standard errors, odd ratios, and z- and *p*-values for the experimental group data.

Predictors	**Odds Ratios CI**	**Estimate**	**Standard error**	**Z value**	**p-value**
Test [Post-test 1]	1.60 (1.26 – 3.11)	0.47	0.22	2.12	.03*
Conditions [Bare singular]	0.41 (0.27 – 0.64)	−0.88	0.22	−4.13	< .001*
Conditions [Definite plural]	0.50 (0.33 – 0.78)	−0.69	0.22	−3.11	.002*
Conditions [Definite singular]	0.65 (0.30 – 0.71)	−0.77	0.22	−3.55	< .001*
Conditions [Indefinite singular]	0.57 (0.37 – 0.87)	−0.56	0.22	−2.58	.01*
Conditions [Bare singular] × Test [Post-test 1]	0.34 (0.19 – 0.63)	−1.07	0.32	−3.46	.001*
Conditions [Definite plural] × Test [Post-test1]	0.22 (0.12 – 0.42)	−1.49	0.32	−4.74	< .001*
Conditions [Definite singular] × Test [Post-test 1]	0.70 (0.38 – 1.27)	−0.35	0.31	−1.17	0.24
Conditions [Indefinite singular] × Test [Post-test 1]	1.41 (0.76 – 2.60)	−0.54	0.31	−1.09	0.28

Note: model: clmm (Ratings ~ Conditions+ Test+ Test*Conditions + (1 | id) + (1 | item), data = pos t_delAJTch, Hess = TRUE, threshold = "flexible"), observations 1439; id 25; items 6.
*Significant at .05 level.

Table 7.27 uncovered that the improvement in rating the bare plural as acceptable in characterising generic condition was retained in the delayed post-test; the mean rating of bare plurals was significantly higher than the ratings of all other conditions in the delayed post-test. Moreover, post-hoc pairwise comparisons showed that there was no significant difference between the experimental group's mean ratings of the bare plural in the delayed post-test compared to the first post-test (*ES* = −0.47, *SE* = 0.22, *z* = − 2.12, *p* = .99). Unlike the bare plural condition, the descriptive statistics showed that the mean rating of the indefinite singular (acceptable condition) decreased in the delayed post-test and was significantly lower than the mean ratings of the acceptable bare plural condition. Moreover, post-hoc pairwise comparisons showed that the delayed post-test mean rating of the indefinite singular condition was not statistically significantly different from the mean ratings of the unacceptable condition, as (*p* > .05) in all comparisons.

Considering the differences in mean ratings between post-tests per each generic condition in the characterising generic context, Table 7.28 summarises the results of the related post hoc pairwise comparisons.

Table 7-28 Tukey-corrected post hoc pairwise comparisons of mean ratings in both post-tests for the experimental group by condition.

Contrast	**Estimate**	**SE**	**Z ratio**	***p*-value**
Post-test 2 - Post-test 1 [Bare plural]	−0.47	0.22	−2.12	.99
Post-test 2 - Post-test 1 [Bare singular]	0.60	0.22	2.78	.75
Post-test 2 - Post-test 1 [Definite plural]	1.02	0.22	4.59	.004*
Post-test 2 - Post-test 1 [Definite singular]	−0.11	0.21	−0.54	1.00
Post-test 2 - Post-test 1 [Indefinite singular]	−0.81	0.22	−3.67	.12

Note: *Significant at the .05 level.

Table 7.28 compares the mean ratings in the delayed post-test to the post-test by condition. The results show that the mean ratings increased in unacceptable conditions, including the bare singular and definite plural conditions; yet, only the definite plural was rated significantly higher in the delayed post-test as compared to the post-test even though the mean rating did not exceed the mid-scale (*M* =2.53, *SD* = 1.22).

I now turn to investigate whether the experimental group retained the improvement reported in the post-test in kind generics three months after the intervention. To recall, the experimental group showed improvement in accepting the bare plural and definite singular conditions in the post-test, as they rated them significantly higher than the unacceptable conditions. Figure 7.39 shows the distributions and mean ratings for the conditions in the kind generics delayed post-test.

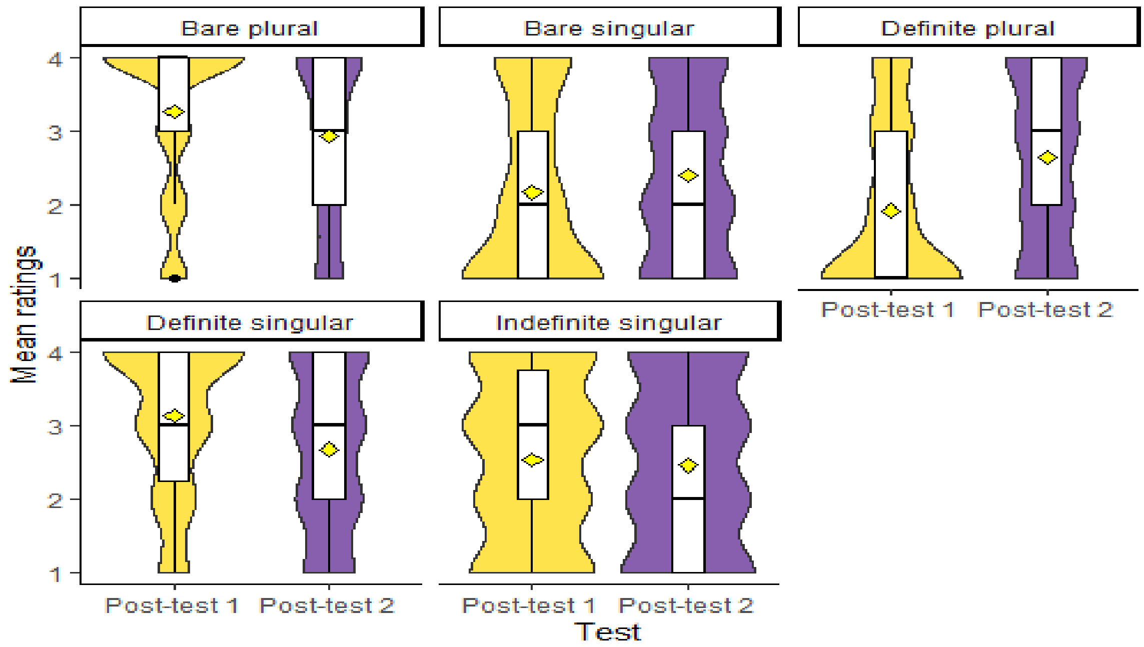


Figure 7.35 The experimental group's mean ratings and distributions per condition in the kind generics post-test and the delayed post-test.

Figure 7.35 compares the experimental group's mean ratings and distributions per condition in the post-test and the delayed post-test for kind generics. A closer inspection of the figure showed a change in the ratings' means and distributions in all conditions. As for acceptable conditions, the figure showed a decrease in the mean ratings of bare plural and definite singular. In contrast, the mean ratings of the unacceptable definite plural and bare singular conditions all increased in the delayed post-test. No change was observed in the mean rating of the indefinite singular condition.

A CLMM was constructed to determine any significance of the abovementioned change. The model structure included the conditions, test and their interaction as fixed effects, with rating as the dependent variable, and by-participant and by-item as random effects. The bare plural condition and the delayed post-test were the reference levels of the condition and test categories, respectively. The single-term deletion test showed a significant effect of the interaction between condition and test (χ^2 = 56.2, *df* = 4, *p* <.001). Table 7. 29 summarises the model results.

Table 7-29 Kind generics post-tests comparison model estimates, standard errors, odds ratios, and z- and *p*-values for the experimental group data.

Predictors	**Odds Ratios CI**	**Estimate**	**Standard error**	**z-value**	***p*-value**
1|2	0.18 (0.13 – 0.26)	− 1.70	0.19	− 8.97	< .001*
2|3	0.49 (0.34 – 0.70)	− 0.71	0.18	− 3.85	< .001*
3|4	1.35 (0.94 – 1.94)	0.30	0.18	1.62	.10
Conditions [Bare singular]	0.42 (0.27 – 0.64)	− 0.88	0.22	−3.96	< .001*
Conditions [Definite plural]	0.61 (0.40 – 0.95)	−0.49	0.22	−2.21	.03*
Conditions [Definite singular]	0.65 (0.42 – 1)	−0.43	0.22	−1.98	.05*
Conditions [Indefinite singular]	0.45 (0.29 – 0.70)	−0.79	0.22	−3.60	< .001*
Test [Post-test 1]	1.98 (1.26 – 3.11)	0.68	0.23	2.94	.003*
Conditions [Bare singular] × Test [Post-test 1]	0.34 (0.18 – 0.63)	−1.09	0.32	−3.43	.001*
Conditions [Definite plural] × Test [Post-test1]	0.15 (0.08 – 0.27)	−1.92	0.32	−5.99	< .001*
Conditions [Definite singular] × Test [Post-test 1]	1.05 (0.57 – 1.95)	0.05	0.32	0.17	.87
Conditions [Indefinite singular] × Test [Post-test 1]	0.59 (0.32 – 1.08)	− 0.53	0.31	−1.72	.09

Note: model: clmm (Ratings ~ Conditions+ Test+ Test*Conditions + (1 | id) + (1 | item), data = post_delAJTK, Hess = TRUE, threshold = "flexible"), observations 1440; id 25; items 6.

*Significant at .05 level.

Table 7.29 summarises the CLMM for kind generics in the two post-tests . The post-test results showed that the experimental group rated bare plural and definite singular significantly higher than unacceptable conditions in the post-test. An inspection of the data in Table 7.29 reveals that the experimental group's mean rating of the bare plural, an acceptable condition, was significantly higher than the mean ratings of all other conditions in the delayed post-test. The experimental group's high ratings of the bare plural condition were reserved in the delayed post-test as post-hoc pairwise comparison showed no significant difference between the post and delayed post-tests in bare plural (*ES* = −0.68, *SE* =.23, *z*= −2.94, *p* =.63). As for the definite singular, the model showed that the mean ratings of the acceptable definite singular were significantly lower than the mean ratings of the bare plural. A post-hoc test showed that it was not significantly different from the mean ratings of all the unacceptable conditions ($p > .05$, for all comparisons between definite singular and bare singular, indefinite singular, and definite plural in the delayed post-test). Moreover, post-hoc pairwise comparisons showed a significant decrease in the mean rating of definite singular in the delayed post-test compared to the mean rating for the post-test (*ES* = −0.73, *SE* = 0.21, *z* = −3.41, *p* = .02).

As for the unacceptable conditions, post-hoc pairwise comparisons showed that the change in the indefinite singular and bare singular conditions was not significant, meaning that the experimental group maintained low ratings for these conditions in the delayed post-test (*ES* = 0.41, *SE* = 0.22 *z* = 1.873, *p* = 0.68 for bare singular condition and *ES* = −15, *SE* = 0.21 *z* = −0.7, *p* = 0.99 for the indefinite singular). The change in the mean rating of the definite plural in the delayed post-test was striking. The findings indicate a significant difference between the mean rating of the definite plural between the delayed post-test and post-test 1, with an estimated effect size of 1.24 and a high level of statistical confidence (*SE* = .22, *z* = 5.59, $p < .001$).

Overall, the post-test results for the AJT revealed an improvement in the experimental group's generic form-meaning mapping in both kind and characterising generics compared to their pre-intervention form-meaning mappings. Moreover, the results showed a significant effect of the intervention as there was a significant difference in mean ratings between the experimental and the comparison group in the post-test in both generic contexts. The experimental group maintained the improvement in plural generic form-meaning mappings in the delayed post-test as the mean ratings of the bare plural conditions in both contexts were significantly high, and there was no significant difference in the bare plural conditions' mean ratings between the delayed post-test and post-test. The delayed post-test showed that the singular generic form-meaning mappings were challenging for the L2 learners in this study.

7.3.3 FCT post-test Results

The FCT pretest showed that the L2 learners had difficulty accurately selecting English articles for characterising and kind generic meanings. Notably, both the experimental and comparison groups of the L2 learners exhibited low proportions of target-like choices in each generic condition. To investigate whether the experimental group's selection of expected forms improves after receiving instruction on genericity, both groups' responses to the post-test FCT were analysed using descriptive and inferential statistics, which results are presented below.

In terms of description, each group's proportions of expected target-like choices after eight weeks of teaching were compared to the group's pre-test proportions and plotted in the following Figure.

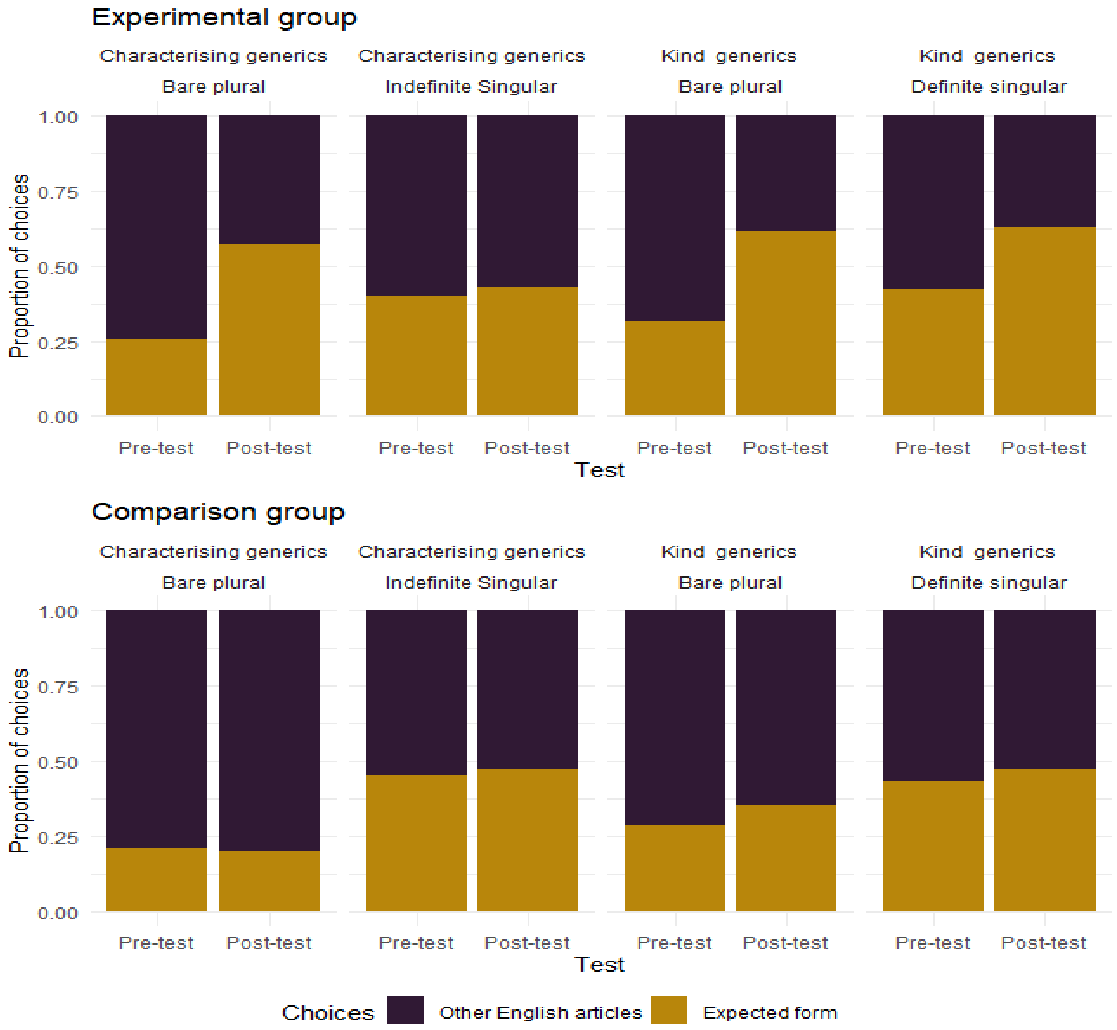


Figure 7.36 Proportions of expected and other English articles used in the post-test compared to the pre-test by group and condition.

Figure 7.36 compares the proportions of expected target-like choices for each group in the pre-test and post-test FCT tasks. This figure shows that the experimental group exhibited higher proportions of expected form selection in the post-test (after the teaching intervention) across all conditions (above 57% expected form selection), except for selecting indefinite singular forms in the characterising generic condition (43% expected form selection). In contrast, the comparison group demonstrated similar proportions of expected form selection before and after the intervention across all conditions. Comparing the post-test proportions of selections between the groups reveals that they had similar proportions of expected form when selecting indefinite singular articles in characterising generic contexts. However, in all other post-test conditions, the experimental group showed a higher proportion of expected form selection than the comparison group. These results are descriptive; subsequent analysis will confirm whether these differences are statistically significant.

GLMM was employed to analyse the FCT post-test data, with separate models constructed for each group. The purpose was to compare the post-test data to the pre-test data and thereby assess the effect of teaching. Between-group differences were then examined using a model that compared both groups in the post-test to determine if they differed significantly after the intervention. For each group, the model structure included accuracy in article selection as the binary dependent variable. Fixed effects comprised generic condition (a categorical variable with four levels, with characterising generic with bare plural as the reference level), test (a categorical variable with two levels: pre-test and post-test, with the latter as the reference level), and the interaction of condition and testing. Additionally, the model incorporated random effects for both items and participants. Each model began with the interaction structure and was subsequently reduced using single-term deletion tests.

Starting with the comparison group results, the single-term deletion test revealed no significant effect of the interaction *testing*condition* (χ^2 = .84, *df* = 3, *p* = .89). The results are summarised in the following table.

Table 7-30 The comparison group post-test and pre-test FCT GLMM model estimates, odd ratios, standard errors, z, and p-values.

Predictors	Odds Ratios (CI)	Estimate	SE	Z value	P value
(Intercept)	0.56 (0.41 – 0.77)	−0.57	.16	−3.64	< .001*
Kind generics [Bare plural]	2.22 (0.93 – 5.28)	0.79	.44	1.80	.07
Kind generics [the]	3.74 (1.46 – 9.64)	1.32	.44	2.99	.003*
Characterising generic [a]	3.78 (1.60 – 8.90)	1.33	.43	3.04	.002*
Test [Pre-test]	0.86 (0.65 – 1.15)	−0.14	.15	−1	.31

Note: glmer (Accuracy ~ Condition+ Test+ (1|id) + (1|item), data = pstFCTCo, control = glmerCon trol (optimizer = "bobyqa", optCtrl = list(maxfun=20000)), family = binomial); Id, 29; item, 16, nu mber of observations,928; * significant at .05 level. Marginal R2 / Conditional R20.071 / 0.136

Table 7.30 summarises the model assessing the comparison group's pre-test and post-test accuracy in selecting English articles with generic meanings. The evidence presented in this table confirms that the comparison group's accuracy in article selection was not statistically significantly different in the post-test compared to the pretest ($p = 0.31$). Moreover, as in the pre-test results (Section7.2.3), accuracy in bare plural selection in kind generics was not significantly different from accuracy in selecting this form in the characterising generic context post-test, and both were significantly lower than singular forms. Further Tukey-corrected post hoc pairwise comparisons showed that none of the conditions exhibited a significant difference between the pre-test and post-test ($p > 0.05$ in all comparisons), as shown in the following table.

Table 7-31 FCT post-test and pre-test Tukey-corrected pairwise comparisons by generic condition for the comparison group.

Contrast	**Estimate**	**SE**	***Z* value**	***P* value**
Bare plural in Characterising generics [post-test to pre-tests]	−0.05	0.33	−0.16	.99
Bare plural in kind generics [post-test to pre-tests]	0.33	0.29	1.16	.94
Definite singular in kind generics [post-test to pre-tests]	0.19	0.28	0.70	.99
Indefinite singular in characterising generics [post-test to pre-tests]	0.11	0.27	0.40	.99

The errors selected by the comparison group in the post-test are visualised in Figure 7.37. This figure shows that using the definite plural was the highest erroneous choice made by the comparison group with conditions that used the bare plural form.

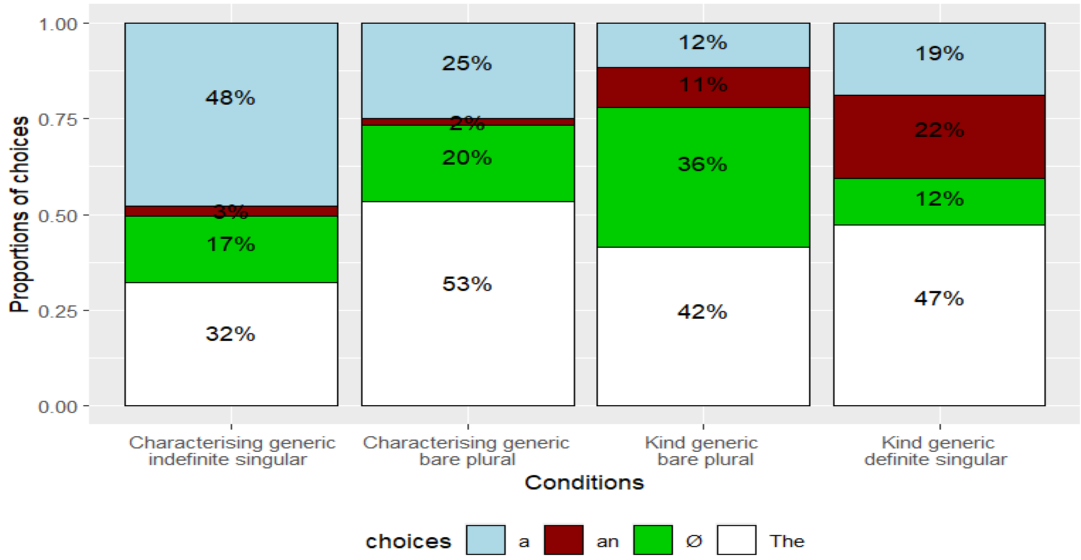


Figure 7.37 Post-test comparison group's choices for each condition.

Turning to the experimental group's article selection before and after the intervention, the descriptive results showed higher expected form choices in all conditions except for selecting the indefinite singular article to express characterising generic meaning in relevant contexts. To assess the significance of these differences, a GLMM was constructed using the same structure as the one used in modelling the comparison group above. The single-term deletion test revealed a significant effect of the interaction between generic conditions and testing fixed effects (χ^2 = 17.13, *df* = 3, *p* =.001). Table 7.32 presents the results of this model.

Table 7-32 The experimental group post-test and pre-test FCT GLMM model estimates, odd ratios, standard errors, z, and p-values.

Predictors	**Odds Ratios (CI)**	**Estimate**	**SE**	**Z value**	**P value**
Kind generics [Bare plural]	1.23 (0.56 – 2.70)	0.20	0.40	0.50	.61
Kind generics [the]	1.35 (0.61 – 2.98)	0.30	0.40	0.73	.46
Characterising generic [a]	0.52 (0.24 – 1.14)	−0.65	0.40	−1.63	.10
Test [Pre-test]	0.36 (0.28 – 0.47)	−1.02	0.14	−7.55	< .001*

Predictors	Odds Ratios (CI)	Estimate	SE	Z value	P value
Kind generics [Bare plural] × Test [Pre-test]	1.11 (0.52 – 2.34)	0.10	0.38	0.26	.79
Kind generics [the] × Test [Pre-test]	1.67 (0.79 – 3.52)	.51	0.38	1.33	.18
Characterising generic [a] × Test [Pre-test]	3.99 (1.91 – 8.35)	1.38	0.38	3.67	< .001*

Note: glmer (Accuracy ~ Condition+ Test+ Test* Condition+ (1|id) + (1|item), data = pstFCTex, co ntrol = glmerControl (optimizer = "bobyqa", optCtrl = list(maxfun=20000)), family = binomial); Id, 35; item, 16, number of observations,1120; * significant at .05 level. Marginal R2 / Conditional R 20.088 / 0.24; * significant at .05.

Table 7.32 compares the experimental group's accuracy in the post-test FCT with their accuracy in the pre-test. The results indicate a significant difference in accuracy between the post-test and the pre-test, as the shaded cell shows. To further evaluate this difference, Tukey-corrected pairwise comparisons were conducted. Table 7.33 summarises the comparisons that assess the significance of the difference between the two tests for each condition.

Table 7-33 The experimental group's FCT post-test and pre-test Tukey-corrected pairwise comparisons per the generic condition.

Contrast	**Estimate**	**SE**	***Z* value**	***P*-value**
Bare plural in Characterising generics [post-test to pre-tests]	1.51	0.27	5.54	<.001*
Bare plural in kind generics [post-test to pre-tests]	1.41	0.27	5.26	<.001*
Definite singular in kind generics [post-test to pre-tests]	1.03	0.27	3.74	.004*
Indefinite singular in characterising generics [post-test to pre-tests]	0.13	0.26	0.52	.99

Note: * Significant at .05 level

This table compares the experimental group's accuracy in selecting the expected form in the post-test FCT to their accuracy in the pre-test FCT to assess the effect of the teaching intervention on their accuracy. The data in this table shows that their accuracy significantly improved after the teaching in all conditions, with the exception of selecting the indefinite article (a) to express characterising generic meaning, where no improvement in accuracy was observed.

After assessing the significance of these differences, it is essential to explore the proportions of each unexpected form that was selected in each condition to understand the difficulty the experimental group faces. Figure 7.38 describes the detailed proportions for the experimental group in the post-test.

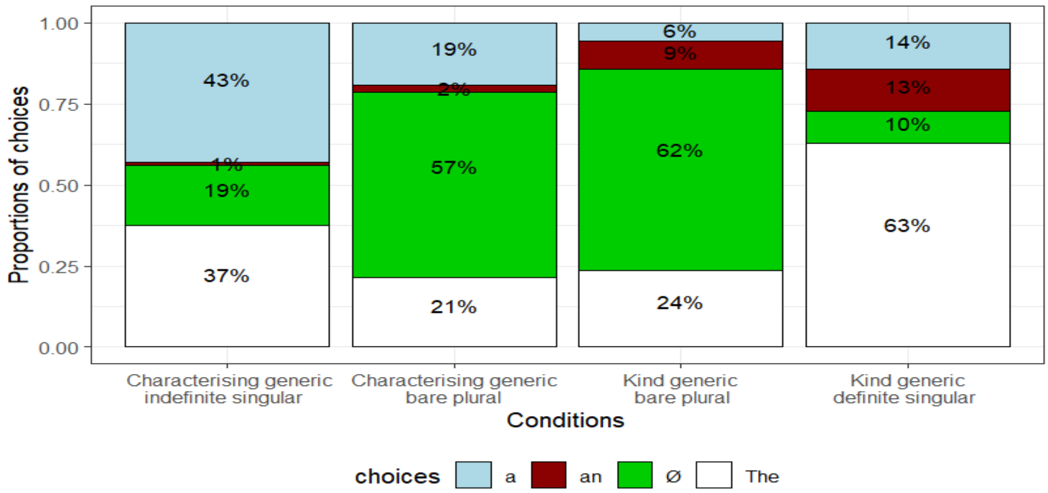


Figure 7.38 Post-test proportions of choices in the FCT by condition for the experimental group.

The figure displays the article choices made by the experimental group during the post-test. As already established, the experimental group exhibited significant improvement in their accuracy across all conditions between the pre-test and post-test except when selecting the indefinite singular form for characterising generic meaning. In the conditions where improvement was observed, the proportion of expected article forms exceeded 56%. However, the focus of the diagram lies on unexpected forms. Analysing the data, we find that when mapping 'a' to characterising generics, the most frequent unexpected choice was the definite singular article 'the' (L1 form), accounting for 37% of the choices in this condition. Additionally, 19% of the choices involved the ungrammatical use of bare singular forms. For the expression of characterising generics with bare plurals, 21% of the selections used the definite plural (the L1 form), while another 21% involved ungrammatical indefinite articles with plural nouns. Turning to kind generics, when bare plurals were expected, 24% of the choices reflected the L1 form by selecting definite plural forms, and the remaining 15% consisted of ungrammatical use of indefinite singular articles. Finally, when mapping kind generics to the singular definite article ‘the’, 27% of unexpected choices involved selecting indefinite articles, and 10% were ungrammatical bare singular forms. These results highlight the challenge posed by the L1 form.

The examination of within-group differences in accuracy so far reveals that while the experimental group exhibited significant improvement in selecting the expected form in generic contexts during the post-test compared to the pre-test, the comparison group did not. Section 7.2.3.2. reported that both groups had the same level of accuracy before the teaching, and now the focus shifts to whether the groups significantly differ after the intervention.

To assess between-groups differences in the post-test FCT, a GLMM was employed with accuracy as the dependent binary variable, the groups (a categorical variable with two levels with the experimental group as the reference), Conditions (a categorical variable with four levels with bare plural in characterising generics as the reference) and their interactions as the fixed effects. The model also includes by-participant and by-item random effects. This structure was the structure of the final model as the single-term deletion test showed a significant effect of the interaction group * condition in this model (χ^2 = 30.13, *df* = 3, *p* <.001). The model is summarised in Table 7.34

Table 7-34 Summary of a logistic mixed effects regression model used to examine between-group differences in the post-test FCT.

Predictors	**Odds Ratios (CI)**	**Estimate**	**SE**	**z**	***p***
(Intercept)	1.32 (0.98 – 1.79)	0.28	0.15	1.82	.07
kind generics [Bare plural]	1.23 (0.74 – 2.05)	0.20	0.26	0.78	.43
Kind generics [the]	1.32 (0.79 – 2.20)	0.27	0.26	1.05	.29
Characterising generics [a]	0.51 (0.31 – 0.86)	−0.67	0.26	−2.56	.12
Group [Comparison]	0.42 (0.27 – 0.65)	−0.87	.23	−3.80	<.001*
kind generics [Bare plural] × Group [Comparison]	1.98 (0.85 – 4.19)	0.63	.40	1.57	.12
Kind generics [the]× Group [Comparison]	3.02 (1.37 – 6.65)	1.10	.40	2.74	.01*
Characterising generics [a]*Group [Comparison]	7.73 (3.51 – 17.04)	2.04	.40	5.08	<.001*

Note: glmer (Accuracy ~ Condition+ Group+ Group* Condition+(1|id) + (1|item), data = pstFCTpst, control = glmerControl (optimizer = "bobyqa", optCtrl = list(maxfun=20000)), family = binomial); Number of observations: 1024, groups: 2; id, 64; item, 16; Marginal R^2 / Conditional R2, 0.098 / 0.405; * significant at the .05 level.

Table 7.34 summarises the between-groups post-test comparison GLMM model. The data in this table show that there was a significant difference between groups in the post-test (*p* <.001 in the shaded cell). Further Tukey-corrected post hoc pairwise comparisons were applied to the post-test data to understand the difference between the groups for each generic condition. The results indicate that after the intervention, both groups were not significantly different in their selection of the singular form ‘a’ in characterising generic meaning (*ES* = −0.24, *SE* = 0.34, *z* = −0.7, *p* =.99) for which the experimental group showed no improvement. Moreover, although the experimental group showed significant improvement in selecting ‘the’ for kind generics in the post-test compared to the pre-test, their selection accuracy was not significantly higher than that of the comparison group in the post-test (*ES* = .81, *SE* = 0.33, *z* = 2.30, *p* =.27). Turning to selecting the bare plural form in both generic meanings, the results revealed a significant difference in accuracy between the experimental and comparison groups after the teaching (*ES*

= 1.82, *SE* = 0.35, *z* = 5.16, *p* <.001 for characterising generics, and *ES* = 1.23, *SE* = 0.33, *z* = 3.6, *p* =.007 for kind generics). The between-group comparison results indicate that the experimental group exhibited significantly greater accuracy than the comparison group in selecting the plural form 'bare plural' for both meanings but not for the singular forms.

Let us now shift our focus to the empirical evidence regarding whether the experimental group sustained the accuracy improvement three months after instruction. By comparing the data from the delayed post-test with the initial post-test, we sought to ascertain whether there was any enduring accuracy enhancement within the experimental group. Figure 7.39 illustrates the proportions of expected and unexpected form selections by the experimental group in both the post-test and delayed post-test.

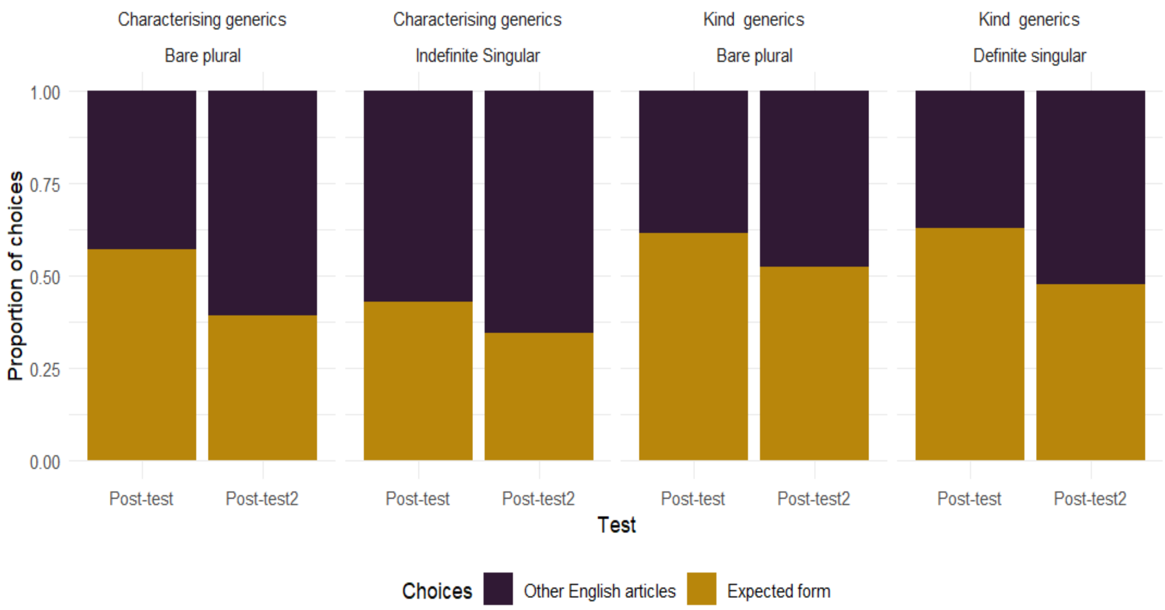


Figure 7.39 Accuracy proportions in articles selection in the post-test and delayed post-test by condition for the experimental group.

The figure illustrates a decline in the proportions of expected form selection across all generic conditions. Recall that upon comparing pre- and post-tests within the experimental group, improvements were observed in all conditions except for indefinite singular selection in characterising generics. A closer examination of the figure reveals specific changes; the proportion of selecting bare plurals in characterising generics decreased from 57% in the post-test to 39% in the delayed post-test. Similarly, the expected form proportions for selecting bare plurals in kind generics decreased from 62% in the post-test to 52% in the delayed post-test. Finally, the selection of definite singulars to express kind generic meaning decreased from 63% to 48% in the delayed post-test. However, the proportions of expected form selection in these conditions remained higher than those observed in the pre-test. Further analysis used GLMM to test the significance of this decrease.

Accuracy in selecting the expected form was the binary dependent variable of the model. Generic condition, test, and the interaction of these two variables were included as fixed effects, with characterising bare plurals and the delayed post-test as the reference levels of the condition and test variables, respectively. By-item and by-participants random effects were also included. A single-term deletion test revealed no significant effect of the interaction (χ^2 =6.4, *df* = 3, *p* =.09). The model estimates, odd ratios, standard errors, z, and *p* values are presented in Table 7.35.

Table 7-35 The experimental group's delayed post-test to post-test model results.

Predictors	Odds Ratios (CI)	Estimate	SE	z	p
(Intercept)	0.64 (0.40 – 1.02)	−0.45	0.24	−1.87	.06
Test [Post-test]	1.80 (1.53 – 2.99)	0.76	0.17	4.47	< .001*
kind generics [Bare plural]	1.89 (0.75 – 4.75)	0.64	0.47	1.36	.18
Kind generics [the]	1.49 (0.59 – 3.77)	0.40	0.47	0.85	.40
Characterising generics [a]	0.79 (0.31 – 2.00)	−0.24	0.7	−0.50	.62
kind generics [Bare plural] × Test [Post-test]	0.37 (0.28 – 1.53)	−0.99	0.43	−2.26	.02*
Kind generics [the]× Test [Post-test]	0.92 (0.39 – 2.17)	−0.08	0.44	−0.19	.85
Characterising generics [a]* Test [Post-test]	0.63 (0.27 – 1.49)	−0.46	0.44	−1.05	.30

Note: glmer (Accuracy ~ Condition+ Test+ Test* Condition+(1|id) + (1|item), data = delFCTpst, control = glmerControl (optimizer = "bobyqa", optCtrl = list(maxfun=20000)), family = binomial); Number of observations: 896; id, 35; item, 16; Marginal R^2 / Conditional R2, 0.059 / 0.296; * significant at the .05 level.

This table presents the results of the GLMM applied to the delayed post-test data. The model indicates a significant difference in accuracy between the delayed and initial post-tests. However, the reference level (bare plural in characterising generics in the delayed post-test) did not significantly differ from any delayed post-test conditions. Further Tukey-corrected pairwise comparisons reveal that the experimental group accuracy in the delayed post-test was not significantly different from the post-test in selecting the definite singular and bare plural NP with kind generic contexts, indicating that they maintained the improvement in these conditions. Their accuracy in selecting bare plural with characterising generic meaning differed significantly in the delayed post-test, indicating that this improvement was not maintained in the delayed post-test, as Table 7.36 Shows.

Table 7-36 Experimental group accuracy in FCT pairwise comparisons between the delayed post-test and post-test.

contrast	**Estimate**	**SE**	***Z* ratio**	***P value***
Post-test2 to post-test [bare plurals with characterising generics]	−0.98	.32	−3.09	.04*
Post-test2 to post-test [bare plurals with kind generics]	0.02	.32	0.05	.99
Post-test2 to post-test [the definite singular with kind generics]	−0.91	.33	−2.79	0.10

Note: *Significant at .05 level.

Figure 7.40 shows the effect plot of predicted probability of accuracy by the experimental group before and after the intervention and visualises the improvement in accuracy.

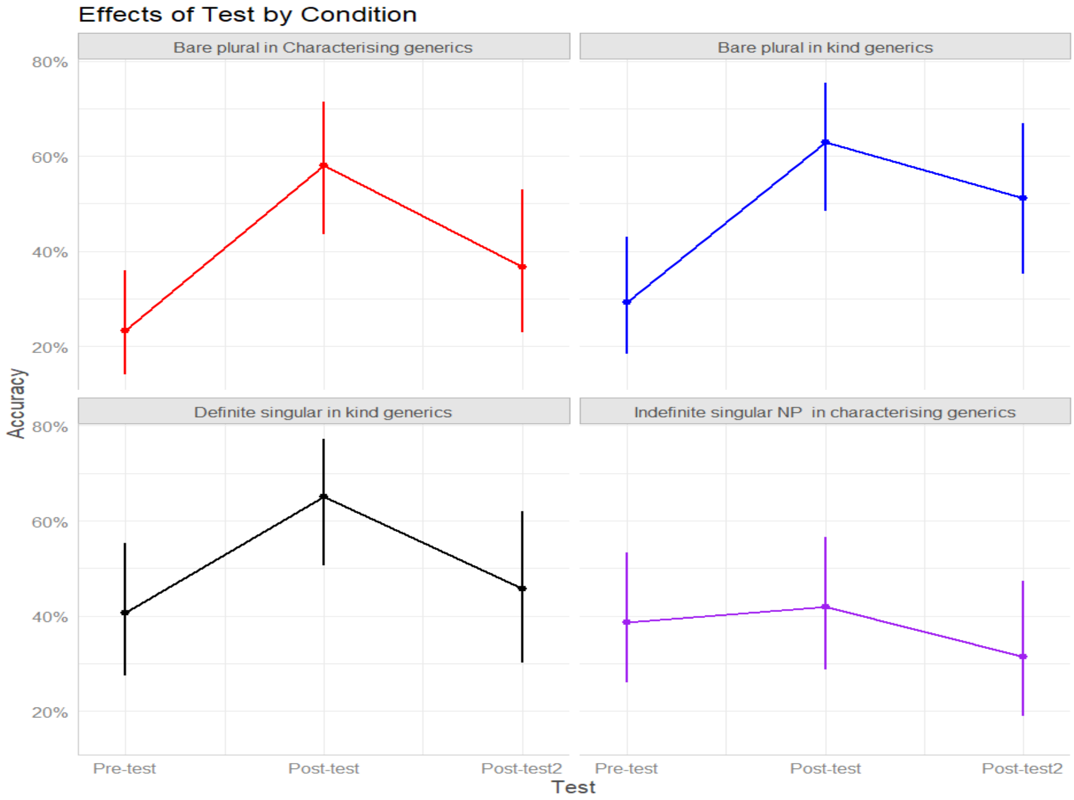


Figure 7.40 Effects plot for the accuracy of the experimental group choices in the FCT.

Examining the proportions of unexpected form selections is essential to gaining a deeper understanding of the delayed post-test responses. Figure 7.41 summarises the proportions for each article selection per condition in the delayed post-test.

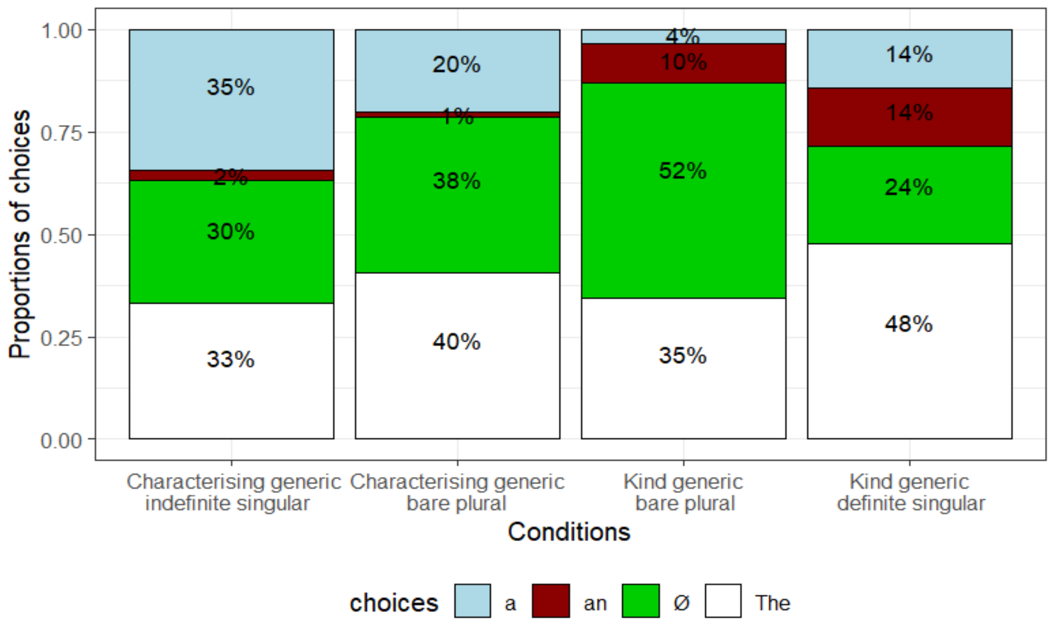


Figure 7.41 Proportions of choices made by the experimental group in the delayed post-test FCT.

Figure 7.41 depicts the distribution of choices made by the experimental group in the delayed post-test FCT. When considering kind generics with plural forms, 35% of the choices involved selecting the definite article ‘the’ with plural noun phrases (L1 form definite plural) in the delayed post-test, compared to 24% in the post-test. Additionally, 14% of choices were ungrammatical, involving singular articles with plural noun phrases. Regarding using singular forms with kind generics, the delayed post-test exhibited errors such as selecting the indefinite article in 28% of the contexts, and an increase in ungrammatical bare singular forms from 10% to 24%. Regarding characterising generics with plural forms, the rate of selecting the L1 form ‘the’ increased from 21% to 40% in the delayed post-test.

In summary, the FCT results demonstrated improvement in the experimental group across all conditions, except for using the indefinite article ‘a’ in characterising generic meaning. When comparing the experimental group’s response to the same task three months later, we observed a slight decrease in expected forms with kind generics. However, this decrease was not statistically significant, suggesting a considerable improvement was maintained in kind blural generics. However, the improvement in characterising generic contexts was not maintained in the delayed post-test.

7.4 Summary

This chapter presented the pre-intervention and post-intervention results of the study tasks: the production task, the acceptability judgement task, and the forced-choice task. Table 7.37 summarises the pre-intervention results.

Table 7-37 Pre-intervention summary of results.

TASK	**Generic conditions**	**Native control group**	**L2 learners (*n* =64)**
Production task	‘a’ in characterising generics	High preference for bare plural. Indefinite singular production was significantly higher than all other singular forms.	The production proportion of ‘a’ is significantly lower than definite singular and plural and bare singular NPs.
	Bare plurals in characterising generics	Bare plural was produced significantly higher than all other NP types.	The production proportion of the bare plural is significantly higher than other NP types but significantly lower than the baseline. The production proportion of producing definite plural is higher than the correct possible singular form (23%)

TASK	Generic conditions	Native control group	L2 learners (*n* =64)
	'the' in kind generics	High preference for bare plural. Definite singular production was significantly higher than all other singular forms.	Its production proportion is not significantly different from the baseline. Its production proportion is significantly lower than the production of bare singular NPs.
	Bare plural in kind generics	Bare plural was produced significantly higher than all other NP types.	The production proportion of bare plural NPs is significantly lower than the baseline. The production proportion of producing definite plural is higher than the correct possible singular form (19%)
AJT	'a' in characterising generics	Rated significantly higher than all unacceptable forms	The mean rating is statistically similar to the mean ratings of the other four NP types.
	Bare plurals in characterising generics	Rated significantly higher than all unacceptable forms	The mean rating is statistically similar to the mean ratings of the other four NP types.
	'the' in kind generics	Rated significantly higher than all unacceptable forms	Although it is similar to the target baseline, The mean rating is statistically similar to the mean ratings of the bare singular, which reflects difficulty.
	Bare plural in kind generics	Rated significantly higher than all unacceptable forms.	The mean rating is statistically similar to the mean ratings of the L1 form (definite plural NPs). It is rated significantly higher than indefinite singular and bare singular mean ratings.
FCT	'a' in characterising generics	This group showed a high mean of selecting the predicted article in each context. *M*= 87.50%, *SD* (17.2)	L2 learners' accuracy was significantly lower than the baseline group in the four conditions. *M*= 42.6%, *SD* (23.4). Significantly higher than Bare plural. Errors: 39% definite singular, 18% bare singular
	Bare plurals in characterising generics	*M*= 86.25%, *SD* (17.2)	*M*= 23.4%, *SD* (26.3) Errors:49% definite plural and 26% indefinite singular.
	'the' in kind generics	*M*= 88.75%, *SD* (20.6)	*M*= 42.6%, *SD* (23.4). Significantly higher than Bare plural. Errors: 43% indefinite singular, 14% bare singular.

TASK	Generic conditions	Native control group	L2 learners (*n* =64)
	Bare plural in kind generics	*M*= 83.75%, *SD* (16.8)	*M*= 30%, *SD* (29.6) Errors: 45% definite plural, 25% indefinite singular.

This table shows the difficulty L2 learners faced in each generic condition per task. The post-test results showed that the intervention's impact was significant. Table 7.38 provides a comprehensive summary of the improvement in each task post-test.

Table 7-38 Summary of post-test results.

Production task	AJT	FCT
<u>Comparison group:</u> • Post-test accuracy was significantly similar to the pretest accuracy in all conditions except the definite singular NP, for which production increased in the post-test. **<u>Experimental group:</u>** • With kind generic meaning, bare plural and definite singular production was significantly higher in the post-test than in the pre-test and significantly higher than all the ungrammatical forms in the post-test. • Bare plural NP production with characterising generics was significantly higher in the post-test than in the pretest, and higher than all other forms in the post-test. • Using the indefinite singular form in characterising generics was not significantly different from producing the definite singular NP.	**<u>Comparison group:</u>** • There were no significant differences between the pre-test and post-test mean ratings. All conditions were rated similarly in the post-test. **<u>Experimental group:</u>** • The experimental group successfully rated the acceptable NPs higher, and they rated the unacceptable NPs lower after the teaching intervention in both generic meanings.	**<u>Comparison group:</u>** • None of the conditions exhibited a significant difference in accuracy between the pre-test and post-test. **<u>Experimental group:</u>** • The post-test accuracy was significantly higher than the pre-test for all conditions, except selecting the indefinite article (a) in characterising generic meaning.

The delayed post-test results showed that improvement in mapping bare plural onto kind generic meaning was maintained in the three tasks, and mapping this form onto characterising generic was maintained in the EWPT and AJT. The improvement in mapping the definite singular form onto kind generic meaning was not maintained.

Chapter 8 Discussion and Conclusions

8.1 Introduction

This chapter discusses the empirical results of this study and its implications. Section 8.2 summarises the results concerning the research question. Section 8.3 considers the results on genericity acquisition, their implications for FRH and BH, and the hypotheses tested in this study regarding the difficulty in feature reassembly. The results concerning the short-term and long-term impact of explicit instruction on genericity acquisition are discussed in Section 8.4 in relation to previous intervention research literature. In Sectin 8.5, the study's implications are discussed. In Section 8.6, limitations are acknowledged and presented in addition to some proposed directions for future investigations on genericity acquisition.

8.2 Summary of the results

This study used three tasks to test the acquisition of genericity in English by Arabic–speaking learners: an elicited written production task, an acceptability judgment task, and a forced-choice task. This section briefly summarises these results in preparation for the discussion. The results are presented as answers to the study's research questions.

8.2.1 Research question 1 on the difficulty in acquisition

The first research question is repeated here:

Do intermediate Arabic–speaking English learners demonstrate a target-like mapping of English articles to kind and characterising generic meanings? If not, what is the difficulty and how can it be accounted for according to FRH and BH?

The answer to this research question is no, as the results of the three pretest tasks showed that Arabic–speaking L2 English learners significantly differed from the native baseline in the pretest. To answer the question about the nature of variability in L2 learners, a summary of the difficulty in each task follows.

In the elicited written production task, the baseline native control group produced bare plurals significantly more often than the predicted singular form in both characterising and kind generics, and the production of both predicted forms was significantly higher than the ungrammatical forms in all generic conditions. The L2 learners' production did not reflect a target-like production of bare plurals as the L2 learners significantly differed from the baseline in characterising and kind generic contexts, and the accuracy mean percentage was significantly low. As for characterising generics, the results showed that the L2 learners' production of bare plural was significantly higher than that of other forms, but there was a considerable proportion of the production of the definite plural (19%). The production of the correct indefinite singular form reflects a challenge in this generic meaning, too, as the indefinite singular production reflected significantly lower proportions compared to the L1 form (the definite singular) and the ungrammatical bare singular. In kind generics, the L2 learners' production of the bare plural NP was significantly higher than all other forms but significantly lower than the native baseline. 23% of the L2 learners' production was using definite plurals. They also showed difficulty in producing the expected definite singulars as it was significantly lower than the production of bare singulars.

As for the AJT, the baseline control mean ratings of the acceptable NPs were significantly higher than the mean ratings of the unacceptable forms in both the characterising and kind generics. The L2 learners' mean ratings differed significantly from those of the native baseline in all conditions except in the acceptance of the definite singular with the kind generic reading, where the group's ratings were similar. A closer look at the L2 learners' ratings revealed that in characterising generics, L2 learners' ratings of bare plural and indefinite singular NPs did not significantly differ from the mean ratings of the other NP types in characterising generics. For the kind generic meaning, the mean ratings of bare plurals were statistically similar to those of the definite plural NPs, reflecting a challenge. The mean rating of the definite singular NP was statistically like the mean ratings of the bare singular, which reflects difficulty.

Finally, the FCT results revealed that the native baseline control group showed a significant level of predicted form selection in all conditions, as indicated by the high mean of predicted form selection. The L2 learners' mean accuracy in selecting predicted forms in all the conditions of the FCT was significantly lower than the baseline's mean selection of target forms. When the correct choice is bare plural NP (no article), L2 learners tend to select definite plurals (L1 form) about 40% of the time, and some learners selected the indefinite singular 'a' about 25% of the time, which is ungrammatical for both characterising and kind generics in this task. When the predicted selection was the indefinite singular article 'a' in characterising generics, L2 learners selected the definite singular article 'the' (L1 form) 39% of the time, with 14% of the selections for the bare singular NP (no article). I will account for this difficulty in Section 8.3 by discussing each difficulty.

8.2.2 Research question 2 on the impact of instruction on this difficulty

The second research question is repeated here:

After being exposed to 8-week acquisition-informed explicit instruction on English genericity form-meaning mappings, can the experimental group improve their form-meaning mappings in kind and characterising generic meanings compared to the comparison group? If any, is this improvement retained three months after instruction?

To answer this question, I looked at the accuracy of the experimental and comparison groups after completing the instruction intervention for the experimental group, using the same tasks used in the pretest. Then, to account for the long-term effect of instruction on the experimental group's accuracy, they completed a twelve-week delayed post-test. Table 8.1 summarises the post-intervention improvement in each group per each generic condition in the three tasks.

Table 8-1 Post-intervention improvement in L2 learners' generic-form meaning mappings in the three tasks.

	Bare plural with kind generics		**Definite singular with Kind generics**		**Bare plural with characterising generics**		**Indefinite singular with characterising generics**	
	Comparison	Experimental	Comparison	Experimental	Comparison	Experimental	Comparison	Experimental
EWPT	x	✓	x	✓	x	✓	x	x
AJT	x	✓	x	✓	x	✓	x	✓
FCT	x	✓	x	✓	x	✓	x	x
	Long-term effect on the Experimental group							
EWPT		✓		x		✓		x
AJT		✓		x		✓		x
FCT		✓		x		x		X

Note: ✓ indicates improvement in form-meaning mapping, and x indicates the absence of improvement.

Table 8.1 summarises the results of the L2 group's improvement in the three post-test tasks. In each task, improvement refers to the L2 learners' accuracy in the target-like forms being significantly higher than the ungrammatical forms in the post-test. A close look at the top half of the table shows that the comparison group's results revealed no improvement in accuracy in the post-test session. This indicates that time alone is not sufficient to gain improvement in learning the generic form-meaning mappings. After the intervention, the experimental group's results of the three tasks post-test showed improvement in form-meaning mappings in all generic conditions except mapping the indefinite singular NP onto the characterising generic meaning. In this condition, only the results of the AJT showed improvement, as the experimental group ratings of the indefinite singular NP in this condition were significantly higher than all ungrammatical forms. In Section 8.4, I will account for possible interpretations of improvement in each condition.

The bottom half of Table 8.1 answers whether the abovementioned improvement was retained after twelve weeks. The results of the three tasks in the delayed post-test showed that the experimental group's accuracy in mapping the bare plural NP onto kind generic meaning was not statistically different in the delayed post-test as compared to the immediate one, a result that could indicate a long-term effect of the intervention on accuracy in this condition. Similar long-term accuracy was reported in the EWPT and AJT mappings of the bare plural NP onto characterising generic meaning, but not in the FCT. The experimental group's improvement in mapping the definite singular with kind generic meaning was not maintained in the delayed post-test, as the results of the three tasks showed a significant decrease in accuracy in this condition. Possible explanations for the long-term effect in some generic conditions might be provided considering the intervention's nature and length, as Section 8.4 will show later in this chapter.

8.3 Difficulty in learning generic form-meaning mappings: L1 transfer and Feature Reassembly.

This section considers the results answering the first research question, the implications of these results for FRH and BH, and the predictions tested in this study regarding the difficulty in feature reassembly. The summary above showed that the L2 learners' responses to the three tasks significantly differed from the baseline native controls, which could be evidence of acquisition difficulty. This thesis proposed predictions about the difficulty in light of the cross-linguistic feature contrast and reassembly proposed by Lardiere's FRH (2007, 2009) and Slabakova's (2009) cline of difficulty. What follows discusses the L2 learner's results for each of these predictions in turn.

8.3.1 Mapping [+Definite, +kind, −plural] onto the definite singular NP

It has been predicted that Arabic–speaking learners of English may face no difficulty in mapping the definite-singular NP to kind generics as the learners' L1 and L2 showed similarity in the features [+Definite, +kind, −plural], and no reassembly is required in the learning task. Table 8.2 links the pre-test results to this hypothesis.

Table 8-2 Results' support to H_1.

H1: Arabic–speaking learners of English may face no significant difficulty in mapping the definite-singular NP onto the reference to kind generic meaning due to the similarity between the learners' L1 and L2 form-meaning mapping.

Task	EWPT	AJT	FCT
support to H_1	Not supported	Not supported	Not supported

Table 8.2 reveals that L2 learners' results in the three tasks disconfirm this hypothesis and that L2 intermediate Arabic–speaking learners of English faced challenges in mapping the kind generic meaning onto the definite singular NP. Evidence for this difficulty can be found in the results of the three tasks. In the EWPT, L2 learners struggled to map the definite singular NP with the kind generic meaning. Although the L2 learners were not significantly different from native controls in this condition, this result must be interpreted cautiously as the native baseline preferred plural forms and only produced the definite singular in 15% of the contexts. Therefore, being not significantly different from the baseline in the production task is not a precise indication of accuracy. A conclusion on difficulty could be arrived at by comparing the definite singular production to the production of non–target–like singular forms by both groups in this condition. The native baseline results revealed that the baseline production of the definite singular, although lower than the bare plural, was significantly higher than that of non–target singular forms. In comparison, L2 learners' production of the definite singular was significantly lower than that of the ungrammatical bare singular NP. This could provide evidence for the difficulty faced by Arabic–speaking learners. Another evidence can be found in the L2 learners' ratings for the definite singular in the AJT, which showed that the mean ratings of the definite singular NPs were not significantly different from the mean ratings of bare singular NPs, reflecting difficulty. The FCT provides evidence for this difficulty as the L2 learners' accuracy was significantly lower than the choices of definite articles by the native baseline.

Although this result differs from some published studies (Sabir, 2015; Snape et al., 2013), the results on the difficulty in mapping kind referring generics onto the definite singular NP when the L1 and L2 use the same form are consistent with Alzamil (2019). Sabir (2015) reported the L1 transfer facilitative effect, stating that Hejazi learners' mean ratings of the definite singular in her study's AJT were not significantly different from the native baseline ratings, suggesting support for L1 transfer. A similar facilitative L1 transfer in this condition was reported by Snape et al. (2013). In this study, the researchers found that Spanish–speaking learners of English did not face difficulty in mapping the kind generic meaning onto the definite singular NP as Spanish mapped this meaning to definite nouns, and L1 transfer was helpful for the Spanish–speaking L2 learners. In contrast, Alzamil (2019) found that low- and upper-intermediate and advanced learners showed low accuracy in this condition in the FCT and the oral production tasks. He also found that low-proficiency learners relied on L1 transfer more than more proficient learners in this condition.

This prediction reflected Lardiere's (2009) FRH and Slabakova's cline of difficulty. To recapitulate, as discussed in 3.2.2.1 and 3.2.2.2, these hypotheses assumed Full Transfer and Full Access and claimed that when L1 and L2 form-meaning mappings are similar, reassembly of features is not required for the learning task and the learners find acquisition easier relying on L1 transfer, this condition requires no reassembly of features. Cho and Slabakova (2014) added that if the reassembly is from an overt morpheme in the L1 to an overt morpheme in the L2, the acquisition of the L2 is argued to be less challenging when the input provides enough evidence. The effect of the similarity between L1 and L2 and the reassembly requirement in the learning task for this condition is not found as predicted. It is difficult to explain this result, but it might be related to the developmental stages of the learners' mental grammar, which does not necessarily reject the assumptions of FRH. A possible explanation might be that L2 intermediate learners start to restructure their L2 mental grammar away from the transferred L1 system at this stage of acquisition development and, therefore, conduct omission and substitution errors, which they may overcome with more exposure to evidence in the L2 input. VanPatten (2020) highlights that the role of the L1 in L2 acquisition is constrained and that UG enables L2 learners to restructure their L2 mental grammar based on the available input. The difficulty in mapping kind generic meaning onto the definite singular could be attributed to the fact that this form-meaning mapping is not highly attested in the input. Slabakova (2019) suggested that form-meaning mappings that are less frequent in the input are more challenging to acquire. According to Biber et al. (1999), the definite generic article is rarely used, appearing in less than 2.5% of conversational and fictional contexts and about 5% of academic and news contexts. As for generic form-meaning, Czypionka and Kupisch (2020, as cited in Kupisch and Snape (2024)) found that mapping definite singulars to the kind generic meaning is less frequent in the corpus (7%). Ionin et al. (2011b) discussed the difficulty found in this condition suggesting that limited frequency in the input might affect the difficulty in mapping definite singular to kind generics. In addition to these possible explanations, the fact that the definite article has different functions in English—such as anaphoric, specific, situational, and generic readings—could create mixed positive input, which might confuse the L2 learners, as Snape et al. (2013) suggested.

A closer look at the nature of difficulty suggests that the task type effect cannot be ruled out. While Arabic–speaking learners' extensive use of bare singular in kind referring generics was noticed in the elicited written production task (43%) and acceptability judgement task (2.7 mean ratings), the elicited forced choice task showed more use of the indefinite article (43%) with singular NPs instead of the use of the correct definite article and lower proportions of using no article (14% bare singular NPs). A similar task-related error pattern was reported by Alzamil (2019), as the participants in his study used bare singular in the production task more often than the FCT. Several factors could explain this observation. Firstly, the three tasks employed in this study have different focuses: the EWPT focused on measuring the ability to use language in production without targeting the learners' attention to the form; the AJT focused on measuring the learners' intuitions to accept a sentence based on interpretation without highlighting a grammatical form; and the FCT focused on choosing the grammatical form. While the first two tasks are more implicit in nature, the FCT is more explicit as its focus is the use of metalinguistic knowledge about articles. A possible explanation is that the L2 learners might rely on their metalinguistic knowledge of the need for articles used with singular NPs, which leads them to select the ungrammatical indefinite article in the absence of the anaphoric reference in the provided context. This could explain the increase in the use of indefinite articles in this task compared to the other two tasks. Secondly, it seems possible that the way the elicited written production task is designed may encourage Arabic–speaking learners to use bare singular NPs more frequently. The test items that aimed to elicit the definite singular used a singular picture cue. It could be that the L2 learners rely on naming the item in the photo without worrying about the correct use of English articles.

8.3.2 Mapping [− definite, + generic, + plural] and [− definite, + kind, + plural] onto bare plural NPs.

The semantic literature in Chapter 2 and the native speaker results in Chapter 5 showed that English maps both meanings onto bare plural NPs, whereas MSA uses definite plural NPs. Bare plural in English can be ambiguous; although bare plurals usually give a generic reading, they can provide an existential reading in some contexts. Bare plurals in MSA are always indefinite and only give an existential reading. Moreover, using the definite plural in MSA can be ambiguous as it can encode anaphoric and generic meanings depending on the conditioning environment. In contrast, the definite plural NPs express maximality in English, only giving an episodic reading. Considering FRH and Slabakova's cline of difficulty, I predicted that Arabic–speaking learners of English may face difficulty in mapping the bare plural NPs to both generic meanings. Arabic–speaking learners are expected to overgeneralise the use of the definite plural NP with a generic interpretation under the effect of the L1 transfer.

The results of the native baseline were entirely consistent with the target-like predicted form-meaning mapping for this condition in the three tasks. In each task, the native baseline group use of bare plural was significantly higher than all other non–target forms. As for the L2 learners, Table 8.3 links the pre-test results of the three tasks to this hypothesis.

Table 8-3 Results' support to H_2.

H2: English maps characterising and kind generics onto bare plural NPs. The learners' L1 maps these meanings into the definite plural. Considering the FRH, Arabic–speaking learners are expected to face difficulty acquiring this form-meaning mapping in English and overgeneralise the L1 use of the definite plural NP with a generic interpretation under the effect of the L1 transfer.

Task	EWPT	AJT	FCT
support to H_2	Partially supported	Supported	Supported

The summary of the results in Section 8.2.1 revealed that, in the EWPT, the L2 learners faced challenges in producing bare plurals in both characterising and kind generic contexts, as the proportions of producing bare plurals were significantly lower than that of the native baseline (40% with characterising generics and 44% with kind generics whilst over 90% by the native baseline). Due to their low production proportions which did not reach 50%, the mid-scale of production, this result indicated difficulty regardless of the fact that their production of bare plural NPs was significantly higher than all other NP types in each generic context. Considering their erroneous production, it was surprising that L2 learners produced low proportions of the L1 form: 19% of definite plural with kind generic meaning and 23% with characterising generic meaning. Although L2 learners' production of the definite plural NP looks low, it was significantly higher than the possible correct definite singular NP in kind generics (p =.003) and higher than the production of the indefinite singular form in characterising generics (p < .001) which reflects a considerable production of the L1 definite plural, hence providing support for the hypothesis about difficulty overgeneralising the definite plural. The support provided by the EWPT is partial, meaning that it supports the hypothesis on the difficulty in mapping bare plural to both generic meanings, but the hypothesis was not fully supported by the results of the EWPT with respect to the predicted overgeneralisation of the production of the L1 form, definite plural NPs.

Comparing this unexpected result to the results of the difficulty in this condition obtained from the other two tasks is essential. The AJT and the FCT provided evidence for the difficulty in mapping bare plurals to characterising and kind generics. As might be expected, the AJT results revealed that the L2 learners' mean rating of the bare plural NP did not significantly differ from that of the definite plural NP with characterising and kind generic meanings. The FCT results revealed a low mean of total accuracy percentage in this condition (*M*= 23.4 with characterising generics and *M*= 30 with kind generics) and a high proportion of selecting the definite plural NP in both generic meanings (L1 form); 49% of the definite plural NP selection in characterising generics and 45% with the kind generic meaning. These results are unsurprising, considering that Arabic–speaking learners might rely on their L1 form-meaning mappings at this developmental stage, as assumed by the FT/FA hypothesis (Schwartz & Sprouse, 1994).

In accordance with the present results, previous studies have demonstrated that L2 learners whose L1 mapped characterising and kind generic meanings onto the definite plural found the acquisition of bare plural form mapping to these meanings challenging and overgeneralised the L1 form. Ionin and Montrul (2010) found that Spanish learners of English as L2 transferred the L1 from the definite plural NP in these conditions. MSA and Spanish map the characterising and kind generic meanings onto the definite plural NP, and L2 learners in both studies showed similar transfer of the L1 definite plural form in both generic meanings. In Ionin and Montul's (2010) study, even advanced Spanish learners in an EFL context faced difficulty in resetting the features of the L1 to those of the L2 under the effect of less exposure as being in an EFL context. This study's results on the difficulty in mapping bare plurals to generic meanings agree with those obtained by Alzamil (2019). Elementary and intermediate Arabic–speaking learners in Alzamil's study rated the definite plurals as generics, and he attributed the acceptance of the unacceptable definite plural to the reliance on L1 at an early acquisition stage. Aboras (2020) reported that L2 learners also showed difficulty mapping bare plurals to characterising generic meaning. This outcome is contrary to those of Snape (2013), who found that advanced Spanish–speaking learners did not show the predicted difficulty with bare plurals for both generic meanings, giving significantly higher ratings to bare plurals than to definite plurals. Snape's (2013) study used advanced Spanish–speaking learners of English, while this study reported results from low –intermediate L2 learners; hence, the results are contradictory.

It is likely that a number of factors, including L1 transfer, the complexity of the L2 and feature reassembly, play a role in explaining the results of the difficulty of mapping bare plurals to generic meanings. L1 transfer can neatly allow for accounting for the difficulty in mapping bare plural NPs to characterising and kind generic meanings shown in this study. The results of the AJT and FCT revealed clear evidence for L1 transfer. The learners' higher accuracy in the EWPT compared to the AJT and the FCT is unlikely to come from L1 transfer. The EWPT nature and design can account for such a result. It is possible that the picture cues might encourage the use of bare plural in the production. In other words, each picture in these generic condition test items includes a picture with more than one item; therefore, some Arabic–speaking learners might rely on naming the items in the pictures, resulting in higher proportions of this form in this task. As for the complexity of the L2 and feature reassembly, Lardiere (2009) distinguished between feature selection and feature reassembly and argued that L2 learners must do both in the L2 learning task. In acquiring the generic reading of the bare plural NPs, L2 learners must select the features for the bare plural in English and assemble the L2 features in an appropriate way. Whilst doing this, they need to reassemble the feature bundle that has already been assembled in MSA. According to Lardiere's FRH, Arabic–speaking learners preset grammar assembled [+kind] to the definite plural. In learning L2, they need to remove the [+ kind] feature from the definite determiner and retain only the [+ definite] feature before transferring it to the L2 mental grammar system. They must then assign the [+ kind] or [+generic] features to the [− definite, +plural] bundle in English based on identifying the grammatical contexts (Slabakova, 2016). Finally, they map the L2 features to bare plurals. Considering that English maps the characterising and kind generic meanings to multiple forms, the form–meaning mappings in the L2 input is complex. This feature reassembly and L2 complexity can account for the difficulty in these two generic conditions. These results are compatible with the FRH and the BH.

8.3.3 Mapping [− definite, + generic, − plural] onto indefinite singular NPs.

The semantic literature and the native speaker results showed that English maps the characterising generic meaning onto a singular form, indefinite singular NPs, in addition to bare plural NPs, whereas MSA uses definite singular NPs. Considering the FRH and Slabakova's cline of difficulty, I predicted that Arabic–speaking learners of English may face difficulty in mapping the indefinite singular NPs onto characterising generic meaning. Low intermediate Arabic–speaking learners are expected to overgeneralise using the definite singular NP with a generic interpretation under the effect of the L1 transfer.

The native speakers' response to the AJT and FCT is compatible with the target-like forms. In this generic condition, they rated the indefinite singular higher than the definite one and selected the indefinite article in the FCT significantly higher than ungrammatical forms. In written production, the native baseline group showed less preference for the singular form, producing the plural form significantly higher than the singular form in items that aimed to generate a single form production. This result is not surprising, as it can be accounted for by the corpus study done by Czypionka and Kupisch (2020, as cited in Kupisch & Snape, 2024). They found that bare plurals were the most common form used to encode the generic meaning (84.3%) and that the indefinite singulars were the least common forms with the generic meaning

(2.9%). The results of the L2 learners reflect a difficulty in this condition, as summarised in 1.2.1. The results of each task are linked to the hypothesis in Table 8.4.

Table 8-4 Results' support to H_3.

H3: English maps the characterising generic meaning onto the indefinite singular NPs. The learners' L1 maps this meaning into the definite singular. Considering the FRH and Slabakova's cline of difficulty, Arabic–speaking learners are expected to face difficulty acquiring this form-meaning mapping in English and overgeneralise the L1 use of the definite singular NP with a generic interpretation under the effect of the L1 transfer. This condition is expected to be more difficult as it is less evident in the L2 input.

Task	EWPT	AJT	FCT
Support to H_3	Supported	Supported	Supported

The L2 learners' results on the three tasks revealed variability in mapping the indefinite singular NP to this generic condition as the L2 learners' ratings were significantly different than the native baseline, as summarised in Section 8.2.1. To recapitulate, the EWPT found that the L2 learners' production of the indefinite singular NP in this condition was significantly lower than that of the native baseline and lower than that of the definite singular and bare singular NPs. In this task, 32% of the participants' production consisted of bare singular NPs and 15% consisted of the definite singular. Further evidence for variability in this condition comes from the AJT, where the L2 learners similarly rated the indefinite singular, the definite singular, and the bare singular NPs. It was not surprising to see a higher proportion of selecting the definite article 'the' in the FCT; 39% of the L2 learners' selection included 'the' and 18% of their selection was bare singular NPs. These results are not surprising, considering that learning this form–meaning mapping requires feature reassembly and that this mapping is less frequent in the L2 input, as stated earlier.

In accordance with the present results, previous studies have demonstrated that mapping the characterising generic meaning onto the indefinite singular NP is challenging for L2 learners. Alzamil (2019) reported difficulty in mapping the indefinite singular onto characterising generic meaning by Saudi Arabic–speaking learners of English. Aboras (2020) reported that Arabic–speaking L2 learners whose L1 map generics to the definite article accepted the L1 form higher than the L2 form in this generic condition. A similar conclusion about this difficulty was reported by Hermes (2020), who reported that Arabic–speaking transferred the definite singular to characterising generic meaning. These studies attributed the ungrammatical mapping of the definite singular to this condition to L1 transfer and L2 complexity. This outcome is contrary to that of Snape (2013) and Snape et al. (2013), who reported that advanced Spanish–speaking learners did not face difficulty in this condition, in contrast to what was expected. Studies on articleless L1–learners of English did not face difficulty in this condition, too. The Russian and Korean learners of English in the study conducted by Ionin et al. (2011b) faced no difficulty in this condition.

Recalling that MSA does not have an equivalent for the indefinite article and encodes indefiniteness using bare NPs, the feature selection and reassembly are more difficult in this condition. The L2 learners in this study need to learn the L1 form and its indefiniteness features before disassembling the [+ generic] feature from the L1 form's features and assemble it to the L2 feature bundle and then learn the contextual conditions in which the L2 form encodes genericity which enables them to accurately map the indefinite singular to characterising generics. The complexity of this reassembly is a potential source of the difficulty in this condition. This study showed that Arabic-speaking learners relied on the MSA generic form–meaning mappings during this complex reassembly. This might explain their higher use of the L1 form, the definite singular, in the three tasks. The role of L2 input in this difficulty cannot be ruled out. The English language maps the indefinite singular to characterising genericity. In parallel, English maps the definite singular to this meaning with a well-defined kind, as in "the Coke bottle has a narrow neck," as discussed in Chapter 2. Therefore, the L2 input includes using both forms with singular characterising generics, which makes the L2 input a source of confusion for the low-intermediate L2 learners in this study. These results could be attributed to the view that the L2 learners relied on their L1 form-meaning mappings in light of the confusion created by the L2 input and the required reassembly, adding that this form–meaning mapping is less frequent in English.

An unexpected finding is that the AJT rating and production proportion of the bare singular NP were noticeably high. Bare singular NPs were accepted the same way the definite singular and indefinite singular were accepted in the AJT. Bare singular NPs in production were significantly higher than the definite singular NPs (both ungrammatical). There are two possible explanations for this observation. First, the production task design could encourage more use of bare singular, as discussed in 8.3.1. The second possibility is that L2 learners opt for the indefinite reading and transfer the MSA indefiniteness form–meaning mapping, bare singular NP. The first possibility was observed in all conditions, and the FCT results revealed that the L2 learners transferred the MSA generic form-meaning mapping as only 18% of bare singular was selected in this condition. In light of these observations, this study considers the task effect as a more reasonable possible explanation of the production of the ungrammatical bare singular in his condition. This study predicted that an intervention designed to consider this difficulty and focused on providing a more precise L2 input in instruction can aid the Arabic–speaking low–

intermediate learners of English in their learning of the generic form-meaning mappings. What follows is an account of the post-intervention results.

8.4 The impact of instruction

This section considers the results of the post-intervention testing, which aimed at assessing the impact of instruction on L2 generic form–meaning mapping learning. Before discussing the results on the impact of instruction, it is essential to recall how this study conceptualised instruction and clarifies some possible misunderstandings of the nature of instructions and improvement in this thesis. As stated in Chapter 4, instruction in this study does not refer to the explicit explanation of generic form-meaning mapping only. However, it combines explicit explanation with maximising the L2 classroom input through example discussion, input-based classroom activities, and input-based communicative practice through activities that increase the learner's opportunities to process the form-meaning mappings. This study assumed that instruction is different from input and that input is the critical factor in L2 mental grammar development, and instruction's role is to support this process by maximising the input and its use in the L2 classroom as well as explicitly explaining the meanings, forms and mappings in this input. Considering this operationalisation of instruction, the positive impact in this discussion so far refers to the impact of the combination of explicit explanation, L2 classroom input and input-based activities.

Another possible misunderstanding is that improvement refers to native-like output. A fact about L2 acquisition is that acquisition goes through developmental stages, which are universal to L2 learners and could overlap and that L2 learners vary in their mental development of the L2 system due to learner-based factors, but this does not contradict the fact of the developmental stages in creating this mental grammar (VanPatten, 2020; White, 2020). Considering these facts and assuming a GenSLA perspective, improvement is defined as a development in the L2 learners' interlanguage that is impossible to measure directly (White, 2020). Therefore, this study collected data from three different measures, and improvement refers to the L2 learners' ability to show significantly higher accuracy in generic form-meaning mappings after the intervention compared to their accuracy before the intervention in their performance in the three measures. In this sense, accuracy means that the target-like form-meaning mappings are significantly higher than the ungrammatical form-meaning mappings per each generic condition.

Overall, the summary of the results in 8.2.1 revealed that the experimental group showed more target-like form–meaning mappings after being instructed compared to the pre-test results in all conditions except mapping the indefinite singular NP onto the characterising generic meaning as their performance in the production task and FCT did not show improvement in this condition. The following is an account of the results of both the comparison and experimental groups in each task. Starting with the EWPT, the comparison group production in all NPs in the post-test was not significantly different from their production in the pre-test, except for using the L1 form, definite singulars, with both singular characterising and kind generic meanings. This means that the comparison group showed no improvement except that they produced a more target-like form in using the definite singular with kind generic meaning. To ensure this is an improvement in the post-test, pairwise comparisons showed that definite singular production with kind generic meaning was significantly lower than the probability of producing the erroneous bare singular. This reflects that the comparison group faced the same difficulty reported in the pretest results in the post-test in EWPT. In contrast, the experimental group showed significantly higher accuracy in producing the expected form in all conditions, but the indefinite singular with characterising generics in the immediate post-test, as the production of the indefinite singular form in characterising generics was not significantly different from producing the definite singular NP in the post-test. In the AJT, the comparison group gave similar ratings to grammatical and ungrammatical forms, accepting both forms in the post-test, which echoes their performance in the pretest AJT. In contrast, the experimental group gave significantly higher ratings for singular and plural forms in each generic context. This group gave significantly lower ratings for the ungrammatical form, rating them as unacceptable. This is an improvement as the experimental group's pretest showed acceptance for both grammatical and ungrammatical forms in the AJT in both generic meanings and they overcame this pattern in the post-test. Finally, the FCT post-test performance of the comparison group was similar to their performance in the pre-test, whilst the experimental group showed more accuracy in all conditions except in selecting the indefinite article 'a' with characterising generic meaning.

I predicted that explicit instruction in this study might improve the generic form-meaning mappings in the experimental group. Overall, the immediate post-test results showed a positive impact of the explicit instruction in this study in all conditions except for mapping the indefinite singular to the characterising generic meaning. Some intervention-related factors can explain the lack of improvement in the latter condition. It is unsurprising that the experimental group found mapping the indefinite singular NP onto characterising generic challenging after the intervention. The learning task for this form–meaning mapping is more difficult, as shown in the proposed cline of difficulty. To recall, it requires learning the new article 'a' and its basic semantic features [−indefinite, − plural] and then disassembling the L1 [+generic] feature from the L1 definite article and reassembling [+generic] to its form in the L2, 'a'. In the cline of difficulty proposed in this study, it was expected that the lower frequency of this form in the input may have increased the difficulty. The reason for this is not clear but it may have something to do with the intervention planning. During the intervention, only one explicit instruction and practice session were devoted to teaching this condition after presenting the characterising generic meaning characteristics and how they are mapped to bare plurals for two weeks. Although learners received two revision sessions for all conditions, they showed no improvement. It is possible that the amount of teaching and practice might be insufficient for this condition. These results suggest an important question about how long L2 learners need to be engaged with instructional input that instantiates the form-meaning mappings to show improvement.

Apart from the difficulty in the indefinite singular NP, the post-test results support the prediction that explicit instruction that maximises the L2 learners' engagement with input and clarifies the semantic meaning characteristics and its mapping onto different forms might accelerate the feature reassembly required in the learning task. Evidence for this argument can be found in the comparison group's results, which revealed that time alone is insufficient for improvement. If it were not the instruction in this study that might increase the experimental group's accuracy, we would expect similar improvement in the comparison group post-test. This does not suggest that the feature reassembly was not happening in the comparison group, but it might be slower. The comparison group's results showed an improvement in mapping the definite singular onto kind generic meaning. This proves that feature reassembly was in progress in the comparison group in light of L1 transfer but was slower. However, this improvement was not enough as the comparison group showed higher ratings for the bare singular in this condition; therefore, it is not considered an improvement according to this study's definition. Another finding is that the experimental group significantly improved their understanding of the ungrammaticality of bare singular count nouns in English. After the intervention, there was a significant decrease in the use of bare singulars in all tasks. Since bare singulars are just ungrammatical in English, the treatment had an effect not just on the genericity form to meaning mappings but also on knowledge of article use.

Considering the impact of the linguistically informed input, the improvement observed in the experimental group's post-test results cannot be solely attributed to this intervention input. The fact that the control group did not receive any additional teaching as compared to the experimental group is a limitation, as it leaves open the interpretation that the gains of the experimental group were due to the additional instructional time rather than the specific linguistically informed input. This highlights the necessity for future studies to provide balanced additional teaching to both groups to more accurately isolate the effects of the linguistically informed input used in the intervention and ensure a robust evaluation of its impact. If the control group had received additional teaching, it would have been possible to more clearly distinguish between the effects of general additional input and the specific benefits of linguistically informed instruction. This methodological adjustment would help to eliminate any ambiguity regarding the source of the experimental group's improvements. By ensuring that both groups receive an equal amount of instructional time, future research can better attribute observed gains to the specific type of input provided, thereby strengthening the validity of the findings and offering more definitive conclusions about the efficacy of linguistically informed instruction.

Having accounted for the improvement shown by the experimental group in the immediate post-test, the present study was designed to consider the long-term effect of instruction in this study by using a 12-week delayed post-test. This thesis found that the improvement reported on the immediate post-test was maintained for some generic conditions. The three tasks demonstrated that the experimental group sustained accuracy in mapping bare plurals onto kind generic meaning twelve weeks later. Another result is that the improvement in mapping bare plurals onto characterising generic meaning was found in the elicited production and AJT. The FCT did not show this improvement. A conclusion to draw is that the experimental group showed improvement in mapping bare plurals to both meanings when tested 12 weeks after the intervention. The effect of instruction may partly explain these results. A note of caution is due here since we cannot exclude the possibility that factors other than instruction contribute to the long-term accuracy reported in this study due to the limitation of the delayed post-test data collection. Only three comparison group participants were willing to participate in the delayed post-test; hence, the small sample size data is not representative. Therefore, little is known about the improvement in the comparison group participants' generic form–meaning mapping. Consequently, the role of time in explaining the long-term effect on instruction cannot be ruled out.

It is difficult to compare these results to previous literature as the studies differed in many aspects, including the amount and quality of input, the measurements, the intervention lengths, and the time between the post-test and delayed post-tests. However, these factors possibly account for the different outcomes in this study, compared to Snape and Yusa (2013)–who reported that instruction in their study was not effective – and Sabir (2015), who reported no effect of instruction. These two studies shared a similar number of sessions devoted to teaching genericity (three sessions in each study), suggesting that short intervention length might be a possible factor. Similar to this study, Umeda et al. (2019) and Abumelha (2019) reported an improvement in the immediate post-test. Moreover, Abumelha (2019) reported long-term improvement in bare plurals mapping onto characterising generics.

The long-term effect must be interpreted with caution, as the improvement reported here is limited to 12 weeks delayed post-test, and no conclusion can be drawn about whether this improvement will be maintained after a longer period if we consider that the participants in Umeda et al. (2019) forgot what they had learnt after a year. Very little was found in the literature on the long-term effect of instruction, as discussed in Section 0. Umeda et al. (2019) suggested that the move from explicit to implicit knowledge is unlikely to take place for this complex property. However, this study cannot conclude about the type of knowledge improved in the learners' mental grammar, although a mixture of production, intuition and metalinguistics tasks were used. The results suggested that giving a conclusion about the role of instruction in this study requires further delayed post-tests that include a longer time frame and measures that tap into implicit knowledge. Thus, these results cannot support the weak interface position assumed in the operationalisation of this study instruction—which argues that the move from explicit to implicit knowledge is possible in light of sufficient input and practice. Nevertheless, the results do not suggest supporting the no interface position either. It suggests that more investigations on the quality and quantity of input needed for L2 acquisition and implicit knowledge measures are needed to inform intervention research design to be suitable for contributing to the debate on interface positions.

Furthermore, while the results are consistent with feature reassembly, alternative interpretations are possible. For instance, a usage-based interpretation could suggest that learners' attention was drawn to the specific form-meaning mappings, which they acquired without necessarily implicating feature reassembly. This perspective posits that the observed improvements might be due to increased exposure and practice with the specific linguistic forms rather than a deeper cognitive restructuring. Given that the experimental group did not maintain the improvement except for mapping bare plurals to generic meaning, it is crucial to acknowledge these alternative possibilities and the limitations of what can be claimed based on the current results. Future research should aim to disentangle these effects by incorporating balanced instructional time and considering both feature reassembly and usage-based explanations to provide a more nuanced understanding of the learning processes involved

8.5 Theoretical and pedagogical implications

The pre-test results of the three tasks reflect the predicted difficulty in mapping bare plurals onto characterising and kind generic meanings and mapping indefinite singulars to characterising generics. The results reflect the unpredicted difficulty in mapping the definite singular with kind reading. These results are significant in major respects. The findings of this research provided insights into the role of the learners' L1 in the acquisition task by adding additional evidence on the role of L1 and L1–L2 differences in acquisition difficulty. The findings suggested difficulty may exist even when the L1 and L2 are similar in form–meaning mappings. This highlights the importance of considering the L2 complexity and the learner's acquisition state when discussing acquisition difficulty. This raised important questions about the nature of the L1–L2 similarities and the conditions in which this similarity can facilitate L2 learning. If we assume that the results on difficulty in generic form meaning mappings are not anomalous, a key implication for L2 form–meaning mapping acquisition would be that variability in L2 learners' form–meaning mappings can be accounted for by the Feature Reassembly and the factors provided by Slabakova's (2019) updated Bottleneck Hypothesis. The results provide support for the argument that the need for feature reassembly in L2, as well as L2 complexity and frequency, are factors that contribute to the difficulty in the acquisition of L2 functional morphology. Moreover, the results provide support to the Bottleneck Hypothesis, as the difficulty in acquisition lies in mapping the morphosyntax (English articles) to the meaning whilst the semantics are available to the L2 learners.

The study results suggest a link between linguistic theory, SLA and L2 instruction. Results on the impact of SLA–informed instruction provide some support for the conceptual premise that calls for using the results of L1–L2 linguistic differences and results on variability in L2 learners' mental grammar in informing L2 teaching, providing an implication for conducting intervention studies in link to language acquisition theorisation. This study can provide a tentative methodological implication for the theorisation of intervention research in light of SLA findings. This study suggests that SLA results—about the difficulty source, the role of L1, and the role and type of input— can be beneficial in operationalising the instruction input, tasks, content, and its order, increasing the instruction impact on supporting L2 learning. Therefore, this study's results support the calls for strengthening the link between intervention research and SLA.

Moreover, the results of this thesis highlight the important question of the quality and quantity of input L2 learners need to move from explicit to implicit knowledge on genericity. This study's results do not have any theoretical implications for the interface positions due to the limitation of the design discussed above. However, it suggests that the Instruction role is to provide explanation and engagement with input. For the move from explicit to implicit knowledge, learners need enough input and practice inside and outside the L2 classroom long enough for the move to happen in a foreign language learning context.

In addition to these theoretical implications, this study draws our attention to some pedagogical implications. In light of this study's results, pedagogical grammar that includes the grammatical rule and its semantic constraints, which govern form–meaning mappings, is likely to be of value in L2 classroom input, textbooks, and curriculum design. Grammar teaching may benefit from highlighting the contexts of form–meaning mappings in L2 and how they differ from L1 mappings in instruction to create a more effective instructional input that may provide more support for the learning task. The results of this study suggest that increasing classroom activities and practice that maximise the processing of form–meaning mapping might be highly beneficial and is likely to be considered when teaching complex linguistic properties, like genericity, as such activities are more likely to be associated with the development of implicit knowledge, the ultimate goal of L2 acquisition and instruction. Finally, teaching the English article system in the classroom may benefit from explicit form–meaning mapping explanation in the L2 classroom input. Greater efforts are needed to communicate SLA research to L2 teachers simply and clearly. Communicating basic SLA findings about what is easy and difficult in L2 acquisition and how L1-L2 comparisons can increase L2 teachers' understanding of how to make the classroom input and activities as beneficial as possible to support L2 learning. Such communication can be conducted through teacher in-term training or teacher training programs.

8.6 Conclusion, limitations, and future research

Overall, this investigation assessed the impact of explicit instruction through considering findings on what is easy and problematic in acquiring generic form–meaning mappings by Arabic–speaking learners of English as L2. In addition, it presented an empirical analysis of L1–L2 differences in generic form–meaning mappings by analysing data from native speakers of English and MSA that set out the crosslinguistic differences between the two languages in preparation for this investigation. This study tested hypotheses on the difficulty faced by Arabic–speaking learners in learning form–meaning mappings considering the L1–L2 differences and Lardiere's (2007,2009) FRH. It assessed the impact of explicit instruction that provided positive and negative evidence, explicit explanation of the semantics of genericity and form–meaning mappings, and practice on improving the L2 learners' accuracy in generic form–meaning mapping. To do so, this study used a quasi-experimental design that included pre-test, intervention, post-test, and a twelve-week delayed post-test as well as experimental and comparison L2 learners' groups, in addition to a native control group who acted as the baseline group.

The native control group showed the predicted target–like form–meaning mappings in all three tasks, hence validating the tasks and setting the baseline to which L2 learners' form–meaning mappings can be compared to account for any variability. As predicted, the L2 learner's results showed variability in mapping bare plurals onto characterising and kind generics and indefinite singular NPs onto characterising generics. This difficulty was attributed to the reliance on L1 transfer to meet the reassembly requirements in the learning task, the complexity of the L2, the frequency of the L2 input and the confusion it may cause for low–level learners. It was unexpected that Arabic–speaking L2 learners may face difficulty mapping definite singular NPs onto the kind generic meaning as the L2 is similar to the learners' L1 in this condition. However, the results revealed this difficulty in the three tasks. This unexpected result was explained in light of the L2 complexity as the definite article can serve different functions in English, expressing definite, anaphoric, specific meanings in addition to the generic reading, which might create difficulty at the developmental stage at which the L2 learners might start to restructure their L2 mental grammar away from the L1 system.

The post-test results revealed that the instruction provided in this study supported the L2 learners to improve their knowledge of generic form–meaning mappings in all conditions except for mapping indefinite singular NPs onto characterising generics, which was not supported by instruction in this study. It could be argued that the improvement was noticed in the experimental group in contrast to the comparison group due to instruction. The comparison group did not show improvement in the post-test compared to the pre-test. Hence, the possibility that time is enough for improvement can be ruled out. Drawbacks of the instruction design in this study might provide an account for the lack of improvement in mapping indefinite singular NPs onto characterising generics. However, the learner's L2 proficiency, the L2 complexity and the L1 transfer effect cannot be ignored when interpreting these results. Results on the long–term effect of instruction showed that L2 learners maintain improvement in mapping bare plurals to kind generic meaning in the three tasks, suggesting that the instruction effect might last longer. This result is not surprising considering the silence of this form in the input of the instruction and L2 input in general. However, caution must be applied, as the findings might not generalise to other contexts, and this study showed some limitations.

The conclusions in this research were limited in several ways. Firstly, the results are limited to low-intermediate Arabic–speaking learners. Students at higher proficiency levels might face different difficulties, and therefore a comparison of the difficulties faced by L2 learners from various proficiency levels is worth further investigation. Secondly, the study is limited to the L2 classroom acquisition context; L2 learners in different acquisition contexts might face different difficulties. This highlights the need for future study that compares the difficulty faced by Arabic–speaking learners acquiring generic-form meaning mappings in the L2 classroom context to Arabic–speaking learners in a different context to increase the generalisability of the findings on acquisition difficulty. Thirdly, the study scope was limited to the difficulty in mapping count NPs in the subject position to the characterising and kind generic meanings. Focusing on mass NPs, kind generics other than kind predicates, and genericity in object position could be directions for future research.

Some limitations need to be acknowledged with regard to the research methods and study procedures too. Although this study used a mixture of production, intuition and metalinguistic data collection methods, the data collection was limited to offline methods. Therefore, the methods may not tap into implicit knowledge as hoped. This study should be repeated using offline and online tasks like eye-tracking to give better conclusions. Moreover, this study was limited to explicit teaching that targeted low-intermediate students and lasted eight weeks. Future research may consider experimental investigations that provide longer intervention with continuous revision and practice, consider high intermediate learners, and consider adding an additional year-delayed post-test. Such investigations may give a better understanding of the impact of instruction in L2 learning. Applying the intervention designed in this study to different proficiency levels and comparing the results on how instruction can impact each proficiency level is needed to account for the best starting point to support the acquisition of genericity and whether advanced L2 learners in the EFL context can overcome the associated acquisition difficulty. Additionally, The study's sample was also limited to female participants, meaning the findings may not apply to male learners. Gender differences in experiences and behaviours may affect the results. Future studies should aim to include a more diverse sample to better understand how gender influences learning outcomes. Finally, future research directions could include investigating the intervention with speakers of two different L1s learning English to add to the literature on the impact of explicit instruction by comparing learners from different L1s to see whether instruction is equally valid for both groups.

Despite these limitations, this thesis has provided more profound insights into the difficulty faced by Arabic-speaking learners of English as L2 and how instruction can be operationalised in terms of SLA findings to increase its impact in supporting L2 learners to overcome difficulties. This study contributes to our understanding of what is easy and challenging to acquire in SLA by providing additional evidence for the FRH and the BH claims and supporting the predictive power of these hypotheses.

This study provides important insights into the operationalisation of instruction, and its results contribute to the debate on the relationship between instruction and GenSLA. The current study has extended our understanding of how the GenSLA findings on the L2 learning task and predicted difficulty, input, methods, and knowledge types can inform the L2 instruction, highlighting some strengths and drawbacks which could inform future research. The contribution of this study has been to confirm the need to bridge the gap between GenSLA research and L2 instruction.

This study includes an empirical investigation of how genericity works in MSA and highlights possible variability in the learners' L1. Although this empirical investigation was limited in scope and aimed to support this study in building its prediction on empirical evidence in addition to the semantic literature, it was the first study considering empirical evidence for the expression of genericity in MSA. This language has never been empirically investigated concerning genericity semantic features. Although minor, such an account could provide a broader contribution to the semantic literature by expanding the understanding of how genericity is expressed cross-linguistically.

Appendix A Native speakers' study

A.1 AJT English copy

Instructions:

Each story below is followed by five sentences. Please judge how appropriate each sentence is in the context of the story, and rate it on a scale from 1 to 4. You don't need to rank the sentences with respect to one another: you may give the same rating to two or more of the sentences that follow a single story. If a sentence is **very appropriate**, rate it as a **4**; if it is **completely inappropriate**, rate it as a **1**. If its appropriateness is **somewhere in between**, rate it as a **2 or 3**. Please base your ratings on sentence acceptability in the context of the story. Please circle your response.

Example:

Kevin's daughter Lauren has a favourite toy: her stuffed bear. She plays with her bear all day long and takes it with her everywhere. Even at night...

a)	Kevin knows that Lauren puts her bear next to her.	1	2	3	**4**
b)	Kevin knows that Lauren puts her bear next to herself.	1	2	3	**4**
c)	Kevin knows that Lauren puts her bear next to him.	**1**	2	3	4
d)	Kevin knows that Lauren puts her bear next to himself.	**1**	3	3	4
e)	Kevin knows that Lauren puts her bear next to Lauren.	1	**2**	3	4

Rationale for the ratings: in the context of the story, it makes sense to say that the bear is next to Lauren at night: (a) and (b) are two different ways of expressing this idea, so they are rated with 4. In contrast, (c) and (d) can be interpreted only to mean that the bear is next to Kevin at night, but the context makes this unlikely; additionally, (d) sounds awkward, which is why it is given a lower rating than (c). Finally, (e) expresses an appropriate meaning, but sounds awkward – it's not as good as (a) and (b) at expressing the meaning. Therefore, it might be given a 2 (or alternatively, a 3). Note that these particular ratings reflect only one person's opinion: you may have different intuitions about these sentences.

Standard English AJT (mixed in the native speakers copy)

Context 1: Kind generics

1) The Netherlands is a great country to visit. It has wonderful museums, great food, and excellent public transportation. And, of course, it's a great place to buy flowers. As you probably know...

a)	Tulips are very popular in the Netherlands.	1	2	3	4
b)	The tulip is very popular in the Netherlands.	1	2	3	4
c)	Tulip is very popular in the Netherlands.	1	2	3	4
d)	A tulip is very popular in the Netherlands.	1	2	3	4
e)	The tulips are very popular in the Netherlands.	1	2	3	4

2) I know that you like birds. Well, if you ever visit California, you'll see lots of different kinds of birds there. For example...

a)	Pelican is widespread on the California coast.	1	2	3	4
b)	The pelicans are widespread on the California coast.	1	2	3	4
c)	The pelican is widespread on the California coast.	1	2	3	4
d)	Pelicans are widespread on the California coast.	1	2	3	4
e)	A pelican is widespread on the California coast.	1	2	3	4

3) These woods are really beautiful. And you can do a lot in them: you can hike, pick mushrooms and have picnics. But be very careful – don't leave food around! Otherwise, you might attract animals. You see...

a)	Brown bear is common in these woods.	1	2	3	4
b)	The brown bears are common in these woods.	1	2	3	4
c)	A brown bear is common in these woods.	1	2	3	4
d)	The brown bear is common in these woods.	1	2	3	4
e)	Brown bears are common in these woods.	1	2	3	4

4) I really like going to the zoo. Unfortunately, there are many animals that can't be found in a zoo, or anywhere else. It's very sad. For example...

a)	Dodo birds are extinct.	1	2	3	4

b)	A dodo bird is extinct.	1	2	3	4
c)	The dodo bird is extinct.	1	2	3	4
d)	The dodo birds are extinct.	1	2	3	4
e)	Dodo bird is extinct.	1	2	3	4

5) Sara is a biology teacher. She was teaching a lesson about mammals. Sara discovered students didn't know that some ocean dwellers are not fish. As we all know..

a)	The whale is a mammal.	1	2	3	4
b)	Whales are Mammals.	1	2	3	4
c)	The whales are mammals.	1	2	3	4
d)	Whale is Mammals.	1	2	3	4
e)	A whale is a mammal.	1	2	3	4

6) Mike loves the smell of Oud perfume, which is extracted from the stem of some plants. His passion for Oud leads him to travel to Southeast Asia. As you know,

a)	The Agarwoods are widespread in southeast Asia.	1	2	3	4
b)	An Agarwood is widespread in Southeast Asia.	1	2	3	4
c)	The Agarwood is widespread in Southeast Asia.	1	2	3	4
d)	Agarwood is widespread in Southeast Asia.	1	2	3	4
e)	Agarwoods are widespread in southeast Asia.	1	2	3	4

Context 2: Characterising generics

7) My brother has been in a bad mood lately. And no wonder – his apartment is so uncomfortable; it must be very depressing to live there. And he has a very dim and unpleasant overhead light. I told him he should buy a new lamp – something pleasant. For example, I know that...

a)	A green lamp is very relaxing.	1	2	3	4
b)	Green lamp is very relaxing.	1	2	3	4
c)	The green lamps are very relaxing.	1	2	3	4
d)	The green lamp is very relaxing.	1	2	3	4
e)	Green lamps are very relaxing.	1	2	3	4

8) I would like to give my daughter a pet for her birthday; perhaps I will give her a puppy. My daughter is going to be eight, and she is very responsible. This is really important. As everyone knows...

a)	Little puppies need a lot of time and attention.	1	2	3	4
b)	A little puppy needs a lot of time and attention.	1	2	3	4
c)	Little puppy needs a lot of time and attention.	1	2	3	4
d)	The little puppy needs a lot of time and attention.	1	2	3	4
e)	The little puppies need a lot of time and attention.	1	2	3	4

9) It's my niece's birthday this Saturday – she is going to be three years old. I'm not sure what to get her. Maybe I'll just get her some toys, like a stuffed dog or bear. I can't go wrong with that. We all know that...

a)	The toy animal is a good children's gift.	1	2	3	4
b)	Toy animal is a good children's gift.	1	2	3	4
c)	A toy animal is a good children's gift.	1	2	3	4
d)	Toy animals are good children's gifts.	1	2	3	4
e)	The toy animals are good children's gifts.	1	2	3	4

10) My husband and I are looking for a new car. My husband would like to get a white one because white is such a beautiful colour. But I'm worried about vandalism. I'm worried because...

a)	White car attracts attention.	1	2	3	4
b)	A white car attracts attention.	1	2	3	4
c)	White cars attract attention.	1	2	3	4
d)	The white car attracts attention.	1	2	3	4
e)	The white cars attract attention.	1	2	3	4

11) Animals have different diets. Some of them eat plants. For instance,

a)	The giraffe feeds on the acacia tree leaves.	1	2	3	4
b)	A giraffe feeds on the acacia tree leaves.	1	2	3	4
c)	The giraffes feed on the acacia tree leaves.	1	2	3	4
d)	Giraffes feed on the acacia tree leaves.	1	2	3	4
e)	Giraffe feeds on the acacia tree leaves.	1	2	3	4

12) Food experts report that root vegetables are perfect for your health. It has been said that ...

a)	The carrot is a good source of vitamin K1 and antioxidants.	1	2	3	4
b)	Carrots are good sources of vitamin K1 and antioxidants.	1	2	3	4
c)	A carrot is a good source of vitamin K1 and antioxidants.	1	2	3	4
d)	The carrots are good sources of vitamin K1 and antioxidants.	1	2	3	4
e)	Carrot is a good source of vitamin K1 and antioxidants.	1	2	3	4

Distractors: (8 sentences)

13) My friend Charles is a teacher. He really loves his job, and the children love Charles. Charles is a very experienced teacher, too:

a)	Charles works as a teacher for fifteen years.	1	2	3	4
b)	Charles had worked as a teacher for fifteen years.	1	2	3	4
c)	Charles has worked as a teacher for fifteen years.	1	2	3	4
d)	Charles is working as a teacher for fifteen years.	1	2	3	4
e)	Charles worked as a teacher for fifteen years.	1	2	3	4

14) I really like going to this museum. It's a small museum. Usually, it's almost empty. But yesterday, I came to the museum, and I heard lots of voices. Then I saw that...

a)	A group of tourists were looking at the pictures.	1	2	3	4
b)	A group of tourists is looking at the pictures.	1	2	3	4
c)	A group of tourists looked at the pictures.	1	2	3	4
d)	A group of tourists was looking at the pictures.	1	2	3	4
e)	A group of tourists are looking at the pictures.	1	2	3	4

15) Last night, Ruth went to a party. She asked her roommate Clara to go with her. But Clara couldn't go because she was in the middle of her history class assignment. When Ruth left...

a)	Clara has written an essay for history class.	1	2	3	4
b)	Clara had written an essay for history class.	1	2	3	4
c)	Clara wrote an essay for history class.	1	2	3	4
d)	Clara was writing an essay for history class.	1	2	3	4
e)	Clara writes an essay for history class.	1	2	3	4

16) A long time ago, I had a neighbour named Robert. We were good friends. But eight years ago, Robert moved to Canada. I am sorry that he is gone. I really miss him. After all...

a)	Robert had lived here for eight years.	1	2	3	4
b)	Robert was living here for eight years.	1	2	3	4

c)	Robert lives here for eight years.	1	2	3	4
d)	Robert lived here for eight years.	1	2	3	4
e)	Robert has lived here for eight years.	1	2	3	4

17) My great aunt Sara had a stroke five years ago. As a result, she does not remember how to play her violin. It's very sad. Sara loves the violin! After all...

a)	Sara had played the violin for five years.	1	2	3	4
b)	Sara was playing the violin for five years.	1	2	3	4
c)	Sara has played the violin for five years.	1	2	3	4
d)	Sara plays the violin for five years.	1	2	3	4
e)	Sara played the violin for five years.	1	2	3	4

18) Yesterday, I went out for a walk in our neighbourhood park. There were lots of people there. At first, I didn't understand why. But then I saw that...

a)	A team of athletes was running through the park.	1	2	3	4
b)	A team of athletes ran through the park.	1	2	3	4
c)	A team of athletes are running through the park.	1	2	3	4
d)	A team of athletes is running through the park.	1	2	3	4
e)	A team of athletes were running through the park.	1	2	3	4

19) I am going to buy a house. James is my real-estate agent. All my friends recommend James. He is really good at his job! He knows the town really well, because...

a)	James has sold houses for ten years.	1	2	3	4
b)	James sells houses for ten years.	1	2	3	4
c)	James is selling houses for ten years.	1	2	3	4
d)	James had sold houses for ten years.	1	2	3	4
e)	James sold houses for ten years.	1	2	3	4

20) Thomas left for work early this morning. But his neighbour Anne was already awake. Anne is a singer, and she practices a lot. When Thomas left...

a)	Anne has sung a song.	1	2	3	4
b)	Anne was singing a song.	1	2	3	4
c)	Anne sang a song.	1	2	3	4
d)	Anne had sung a song.	1	2	3	4
e)	Anne sings a song.	1	2	3	4

A.2 Modern Standard Arabic AJT

التعليمات:

فيما يلي تتبع كل قصة بخمسة جمل، يرجى تقييم مدى ملاءمة كل جملة في سياق القصة وتقييمها على مقياس من 1 إلى 4. لا تحتاج إلى ترتيب الجمل بالنسبة لبعضها البعض: يمكنك إعطاء نفس التقييم لجملتين أو أكثر تتبعان نفس القصة. إذا كانت الجملة مناسبة جدًا، قيمها بـ 4؛ إذا كانت غير مناسبة تمامًا، قيمها بـ 1. إذا كانت ملاءمتها في مكان ما بينهما، قيمها بـ 2 أو 3. يرجى استناد تقييماتك إلى سياق القصة. يرجى وضع دائرة حول تقييمك

لدى ابنة ايمن، منى، لعبة مفضلة: دمية دب محشو و تلعب منى مع دبها طوال اليوم، وتأخذه معها في كل مكان. حتى في الليل ..

أ) يعرف أيمن أن منى تضع دبها بجانبها.	1	2	3	**4**
ب) يعرف أيمن أن منى تضع دبها بجانب نفسها.	1	2	**3**	4
ج) يعرف أيمن أن منى تضع دبها بجانبه.	1	**2**	3	4
د) يعرف أيمن أن منى تضع دبها بجانب نفسه.	1	**2**	3	4
ه) يعرف أيمن أن منى تضع دبها بجانب منى	**1**	2	3	4

التبرير للتقييمات:

في سياق القصة، من المنطقي أن نقول أن الدب بجانب منى في الليل: (أ) و (ب) هما طريقتان مختلفتان للتعبير عن هذه الفكرة، لذا يتم تقييمهما بـ 4 و 3. في المقابل، (ج) و (د) يمكن تفسيرهما فقط بمعنى أن الدب بجانب أيمن في الليل، لكن السياق يجعل هذا غير منطقي؛ ولهذا السبب يتم إعطاؤه تقييمًا أقل من (2). أخيرًا، (هـ) يعبر عن معنى مناسب، لكنه يبدو غير سليم نحويا - ليس جيدًا مثل (أ) و (ب) في التعبير عن المعنى. لذلك، قد يتم إعطاؤه تقييمًا 1. لاحظ أن هذه التقييمات تعكس فقط رأي شخص واحد: قد يكون لديك حدس مختلف تجاه هذه الجمل.

يبداء الاختبار في الصفحة القادمة

Context 1: kind generics (6 items)

(1هولندا بلد رائع يجب زيارته، فهي تحتوي على متاحف رائعة وطعام لذيذ ووسائل نقل عامة ممتازة. وهي مكان رائع لشراء الزهور فكما تعلم..

4	3	2	1	ا) التوليب تحظى بشعبية كبيرة في هولندا.
4	**3**	2	1	ب) توليب تحظى بشعبية كبيرة في هولندا.
4	3	**2**	1	ج) التوليبة تحظى بشعبية كبيرة في هولندا.
4	3	**2**	1	د) يحظى توليب بشعبية كبيرة في هولندا.
4	3	2	**1**	ه) توليب تحظى بشعبية كبيرة في هولندا.

2) أعلم أنك تحب الطيور، إذا زرت كاليفورنيا يومًا ما، فسترى الكثير من أنواع الطيور المختلفة هناك. على سبيل المثال

4	3	2	1	ا) ينتشر بجع على ساحل كاليفورنيا.
4	**3**	2	1	ب) ينتشر البجع على ساحل كاليفورنيا.
4	3	**2**	1	ج) البجعة طائر ينتشر على ساحل كاليفورنيا.
4	3	**2**	1	د) ينتشر بجع على ساحل كاليفورنيا.
4	3	2	**1**	ه) ينتشر بجع على ساحل كاليفورنيا.

3) هذه الغابة جميله حقا ويمكننا فعل الكثير فيها كالمشي وجمع الفطر ولكن لابد ان تكون حذرا ولا تتلرك الطعام مكشوفا بالقرب منك لان ذلك يجلب الحيوانات, فكما تعلم ..

4	3	2		اا) دب بني شائع في هذه الغابات.
4	**3**	2	1	ب) الدب البني شائع في هذه الغابات.
4	3	**2**	1	ج) دببه بنية شائع في هذه الغابات.
4	3	**2**	1	د) دب بني شائع في هذه الغابات.
4	3	2	**1**	ه) الدببه البنية شائع في هذه الغابات.

4) أحمد يحب رائحة عطر العود، الذي يتم استخراجه من جذع بعض النباتات. شغفه بالعود يقوده للسفر إلى جنوب شرق آسيا. كما تعلم،

4	3	2	1	ا) أشجار العود منتشرة على نطاق واسع في جنوب شرق آسيا.
4	**3**	2	1	ب) أشجار عود منتشرة على نطاق واسع في جنوب شرق آسيا.

4	3	**2**	1	ج) شجرة العود منتشرة على نطاق واسع في جنوب شرق آسيا.
4	3	**2**	1	د) شجرة عود منتشرة على نطاق واسع في جنوب شرق آسيا.
4	3	2	**1**	ه) شجرة عود منتشرة على نطاق واسع في جنوب شرق آسيا.

5) سارة معلمة علم الأحياء و كانت تلقي درسًا عن الثدييات. اكتشفت سارة أن الطلاب لا يعرفون أن بعض سكان المحيطات ليسوا أسماكًا, فكما نعلم جميعًا:

4	3	2	1	ا) الحوت ثديي.
4	**3**	2	1	ب) الحيتان ثدييات.
4	3	**2**	1	ج) حيتان ثدييات.
4	3	**2**	1	د) حوت ثديي.
4	3	2	**1**	ه) حوت ثدييٌ.

6) أنا أحب الذهاب إلى حديقة الحيوانات. للأسف، هناك العديد من الحيوانات التي لا يمكن العثور عليها في حديقة الحيوانات، أو في أي مكان آخر. إنه أمر محزن للغاية. على سبيل المثال...

4	3	2	1	ا) طيور الدودو منقرضة
4	**3**	2	1	ب) طائر الدودو منقرض.
4	3	**2**	1	ج) طائر دودو منقرض.
4	3	**2**	1	د) طيور دودو منقرضة.
4	3	2	**1**	ه) طائر دودو منقرض.

Context 2: Characterising generics (6 items)

1) كان أخي في حالة مزاجية سيئة مؤخرًا. شقته غير مريحة للغاية ويبدو أن العيش فيها محبط للغاية. لديه ضوء علوي أخبرته أنه يجب أن يشتري مصباحًا جديدًا زاهيًا. على سبيل المثال.خافت للغاية وغير مبهج

4	3	2	1	ا) المصابيح الخضراء مريحة للغاية.
4	**3**	2	1	ب) مصباح أخضر مريح للغاية.
4	3	**2**	1	ج) مصابيح خضراء مريحة للغاية.

د)	المصباح الأخضر مريح للغاية.	1	**2**	3	4
ه)	مصباح أخضر مريح للغاية.	**1**	2	3	4

2) أود أن أمنح ابنتي حيوانًا أليفًا في عيد ميلادها، ربما سأعطيها جروًا. ستكون في الثامنة من عمرها، وهي مسؤولة للغاية. وهذا ما يهم حقًا. الجميع يعلم أن..

ا)	تحتاج جراء صغيرة إلى الكثير من الوقت والاهتمام.	1	2	3	**4**
ب)	يحتاج جرو صغير إلى الكثير من الوقت والاهتمام.	1	2	**3**	4
ج)	جرو صغير يحتاج إلى الكثير من الوقت والاهتمام.	1	**2**	3	4
د)	يحتاج الجرو الصغير إلى الكثير من الوقت والاهتمام.	1	**2**	3	4
ه)	تحتاج الجراء الصغيرة إلى الكثير من الوقت والاهتمام.	**1**	2	3	4

3) نبحث أنا وزوجي عن سيارة جديدة. هو يرغب في الحصول على سيارة بيضاء لأن اللون الأبيض جميل، لكني قلقة من أعمال التخريب. أنا قلقة لأن..

ا)	سيارة بيضاء تجذب الانتباه.	1	2	3	**4**
ب)	سيارة بيضاء تجذب الانتباه.	1	2	**3**	4
ج)	سيارات بيضاء تجذب الانتباه.	1	**2**	3	4
د)	السيارة البيضاء تجذب الانتباه.	1	**2**	3	4
ه)	السيارات البيضاء تجذب الانتباه	**1**	2	3	4

4) يوم السبت سيكون يوم ميلاد ابنة أخي, و انا محتار فيما ساجلب لها , ستكون في الثالثة من عمرها ولربما اجلب لها بعض الالعاب مثل دب او كلب محشو , لا يمكن ان اكون مخطئا فكلنا يعرف أن ..

ا)	لعبة الحيوان هي هدية جيدة للأطفال.	1	2	3	**4**
ب)	لعبة حيوان هي هدية جيدة للأطفال.	1	2	**3**	4
ج)	لعبة حيوان هي هدية جيدة للأطفال.	1	**2**	3	4
د)	لعبة حيوانات هي هدية جيدة للأطفال.	1	**2**	3	4
ه)	لعبة الحيوانات هي هدية جيدة للأطفال.	**1**	2	3	4

5) لدى الحيوانات أنظمة غذائية مختلفة. فبعضهم يقتات على النباتات. على سبيل المثال:

ا)	الزرافة تتغذى على أوراق شجرة السنط.	1	2	3	**4**
ب)	زرافة تتغذى على أوراق شجرة السنط.	1	2	**3**	4

ج) تتغذى الزرافات على أوراق شجرة السنط.	1	**2**	3	4
د) تتغذى زرافات على أوراق شجرة السنط.	1	**2**	3	4
ه) زرافة تتغذى على أوراق شجرة السنط.	**1**	2	3	4

6) قدم خبراء ا غذية تقارير بأن الخضروات الجذرية مثالية للصحة, فلقد قيل أن

ا) الجزر مصدر جيد لفيتامين ك ومضادات الأكسدة.	1	2	3	**4**
ب) جزر مصدر جيد لفيتامين ك ومضادات الأكسدة.	1	2	**3**	4
ج) جزرة مصدر جيد لفيتامين ك ومضادات الأكسدة.	1	**2**	3	4
د) الجزر مصدر جيد لفيتامين ك ومضادات الأكسدة.	1	**2**	3	4
ه) جزر مصدر جيد لفيتامين ك ومضادات الأكسدة.	**1**	2	3	4

Distractors (8 items)

1) ذهبت إلى هذا المتحف. إنه متحف صغير. عادة ما يكون فارغًا تقريبًا. لكن بالأمس أتيت إلى المتحف وسمعت الكثير من الأصوات. ثم رأيت

ا) مجموعة من السائحين كانوا يشاهدون الصور.	1	2	3	**4**
ب) مجموعة من السائحين يشاهدون الصور.	1	2	**3**	4
ج) شاهدت مجموعة من السائحين الصور.	1	**2**	3	4
د) مجموعة من السائحين كانوا ينظرون إلى الصور.	1	**2**	3	4
ه) مجموعة من السائحين يشاهدون الصور.	**1**	2	3	4

2) صديقي كريم مدرس وهو يحب وظيفته ويحب الاطفال فهو مدرس ذو خبرة كبيرة و ...

ا) يعمل كريم كمدرس لمدة خمسة أعوام.	1	2	3	**4**
ب) لقد عمل كريم كمدرس لمدة خمسة أعوام.	1	2	**3**	4
ج) عمل كريم كمدرس لمدة خمسة أعوام.	1	**2**	3	4
د) يعمل كريم كمدرس لمدة خمسة أعوام.	1	**2**	3	4
ه) عمل كريم كمدرس لمدة خمسة أعوام.	**1**	2	3	4

3) كان لي جار اسمه أسعد و كنا أصدقاء جيدين ، لكن قبل ثماذ سنوات انتقل إلى كندا وأنا آسف لرحيله و حقا أفتقده ..

4	3	2	1	ا) قد عاش أسعد هنا لمدة ثماني سنوات.
4	**3**	2	1	ب) كان يعيش أسعد هنا لمدة ثماني سنوات.
4	3	**2**	1	ج) يعيش أسعد هنا لمدة ثماني سنوات.
4	3	**2**	1	د) عاش أسعد هنا لمدة ثماني سنوات.
4	3	2	**1**	ه) أتم أسعد هنا ثماني سنوات.

4) خرجت يوم أمس في نزهة لحديقة الحي و كان هناك الكثير من الناس. في البداية ، لم أفهم ماالذي يجري ولكن بعد ذلك رأيت..

4	3	2	1	ا) كان فريق من الرياضيين يركض في المنتزه.
4	**3**	2	1	ب) ركض فريق من الرياضيين في المنتزه.
4	3	**2**	1	ج) يركض فريق من الرياضيين في المنتزه.
4	3	**2**	1	د) يقوم فريق من الرياضيين بالركض في المنتزه.
4	3	2	**1**	ه) كان فريق من الرياضيين يركضون في المنتزه.

5) أصيبت خالتي الكبرى سارة بجلطة دماغية منذ خمس سنوات و نتيجة لذلك، لا تتذكر كيف تعزف على الكمان. إنه أمر محزن للغاية فهي تحب الكمان.

4	3	2	1	ا) لقد عزفت سارة على الكمان لمدة خمس سنوات.
4	**3**	2	1	ب) كانت سارة تعزف على الكمان لمدة خمس سنوات.
4	3	**2**	1	ج) عزفت سارة على الكمان لمدة خمس سنوات.
4	3	**2**	1	د) تعزف سارة على الكمان لمدة خمس سنوات.
4	3	2	**1**	ه) أتمت سارة العزف على الكمان لمدة خمس سنوات.

6) ذهبت اسماء في الليلة الماضية الى حفلة وطلبت من زميلتها في المسكن سمر ان تذهب معها ولكن سمر لم تستطع الذهاب فلقد كانت في منتصف انجاز مهمة لفصل التاريخ, و عندما غادرت أسماء

4	3	2	1	ا) لقد كتبت سمر مقالا لصف التاريخ.
4	**3**	2	1	ب) لقد اتمت سمر كتابة مقال صف التاريخ.
4	3	**2**	1	ج) كتبت سمر مقالا لصف التاريخ.
4	3	**2**	1	د) كانت سمر تكتب مقالا لصف التاريخ.
4	3	2	**1**	ه) تكتب سمر مقالا لصف التاريخ.

7) غادر كريم للعمل مبكرًا هذا الصباح، لكن جارته آية كانت مستيقظة بالفعل هي مغنية، و تتدرب كثيرًا. عندما غادر كريم...

ا) لقد غنت آية أغنية.	1	2	3	**4**
ب) كانت آية تغني أغنية.	1	2	**3**	4
ج) غنت آية أغنية.	1	**2**	3	4
د) أتمت آية غناء أغنية.	1	**2**	3	4
ه) تغني آية أغنية.	**1**	2	3	4

8) سأشتري منزلًا وجيمس هو وكيل العقارات الخاص بي، كل أصدقائي يوصون به فهو يعمل هنا و يعرف المدينة جيدًا. إنه جيد حقًا.

ا) باع جيمس منازل لمدة عشر سنوات.	1	2	3	**4**
ب) يبيع جيمس المنازل لمدة عشر سنوات.	1	2	**3**	4
ج) جيمس يبيع المنازل لمدة عشر سنوات.	1	**2**	3	4
د) لقد باع جيمس منازل لمدة عشر سنوات.	1	**2**	3	4
ه) أتم جيمس بيع المنازل لمدة عشر سنوات.	**1**	2	3	4

A.3 Results

A.3.1 English native speakers' Cumulative Link Mixed-effects Models

Predictors	Odds Ratios (CI)	Estimate	Std. Error	*Z value*	*p* value
Conditions [Bare plurals in kind generics]	0.17 (0.06 – 0.46)	-1.75	0.50	-3.49	< .001*
Conditions [Bare singulars in characterising generics]	0.00 (0.00 – 0.00)	−6.23	0.46	−13.41	< .001*
Conditions [Bare singulars in kind generics]	0.00 (0.00 – 0.01)	−5.90	0.50	−11.83	< .001*
Conditions [Definite plurals in characterising generics]	0.00 (0.00 – 0.01)	-5.38	0.44	-12.13	< .001*
Conditions [Definite plurals in kind generics]	0.01 (0.00 – 0.02)	-4.64	0.48	-9.72	< .001*
Conditions [Definite singulars in characterising generics]	0.01 (0.01 – 0.03)	-4.42	0.44	-10.11	< .001*
Conditions [Definite singulars in kind generics]	0.10 (0.04 – 0.26)	-2.28	0.48	-4.74	< .001*
Conditions [Indefinite singulars in characterising generics]	0.08 (0.03 – 0.19)	-2.54	0.45	-5.69	< .001*
Conditions [Indefinite singulars in kind generics]	0.00 (0.00 – 0.01)	-6.07	0.50	-12.21	< .001*

A.3.2 MSA native speakers' Cumulative Link Mixed-effects Models

Predictors	Odds Ratios (CI)	Estimate	St. Error	*z. value*	*P value*
Conditions [Bare plurals in kind generics]	1.22 (0.64 – 2.32)	0.20	0.33	0.61	.54
Conditions [Bare singulars in characterising generics]	1.61 (0.90 – 2.88)	0.48	0.30	1.61	.12
Conditions [Bare singulars in kind generics]	0.51 (0.25 – 1.05)	-0.67	-0.36	-1.82	.07
Conditions [Definite plurals in characterising generics]	227.64 (109.06 – 475.17)	5.43	0.38	14.46	< .001*
Conditions [Definite plurals in kind generics]	336.22 (146.81 – 769.99)	5.81	0.42	13.76	< .001*
Conditions [Definite singulars in characterising generics]	81.05 (42.79 – 153.53)	4.40	0.33	13.48	< .001*
Conditions [Definite singulars in kind generics]	80.84 (40.80 – 160.18)	4.39	0.35	12.59	< .001*
Conditions [Indefinite singulars in characterising generics]	1.43 (0.80 – 2.56)	0.36	0.30	1.21	.23
Conditions [Indefinite singulars in kind generics]	3.82 (0.32 – 1.26)	-0.46	0.35	-1.31	.19

Appendix B Tasks for Data Collection

B.1 Elicited Written Production Task

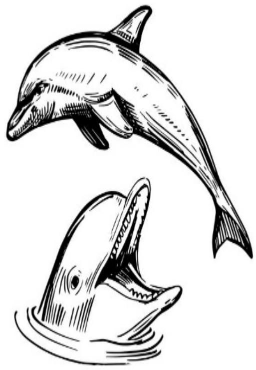
<u>Instruction:</u>

This task involves eight conversations, each accompanied by a picture cue. After reading each conversation and examining the picture cue, you will need to **<u>answer two questions</u>**. Your answers should be in **<u>full sentences</u>** and should incorporate information from both the conversation and the picture cue. Before starting the task, make sure to review the provided example to understand the format and expectations.

<u>Example:</u>

Friends talk: **Sami**: It is a nice drawing. You are talented. Do you consider participating in the art exhibition? **Rami:** Thank you. I have not thought about this before. **Q1: What is Rami doing?** Rami is drawing a cat. **Q2: What is Sami's suggestion?** Sami suggests presenting the drawing in the art exhibition.	

The task starts on the following page. Please, address all the test items. There is no time limit. Answer the test items in the order given. You are not allowed to go back to change your earlier answers.

1- Mother/ daughter talk: **Mother**: Try to eat more vegetables. **Sara**: I hate them. [pointing at the picture]. **Mother:** They are colourful and decrease the risk of heart disease. **Q1: Which vegetables does Sara hate?** **Q2: Which food reduces heart disease risks?**	
2- **Watching a documentary:** **Sam:** Wow! Some ocean dwellers are not fish! **Father**: Sea life is interesting. We will visit SEALIFE Park tomorrow. **Sam**: I can't wait to see these sea dwellers dance. [pointing to the picture] **Q1: Which ocean dwellers are the most widespread marine mammals?** **Q2: Which sea dwellers will Sam see in the SEALIFE Adventure Park?**	
3- Friends' talk: **Mary:** My son cried a lot last night because he lost his preferred toy (pointing at the picture). **Susan**: I can imagine how hard your night was. Kids find emotional comfort in this toy. **Q1: Which toy does Mary's baby lose?** **Q2: Which toy has a comforting effect on kids?**	
4- **A story post-reading discussion:** **Hana**: The bird in this story is wise and kind. I dream about seeing one tomorrow. **Mother**: It is possible because we are not in Antarctica, where it is rare. **Q1: Which bird does Hana dream about seeing?** **Q2: Which bird is rare in Antarctica?**	

5- **Shopping:** **Nora:** This looks warm. I will buy it. (pointing at the items in the photo) **Sara:** Absolutely! **Nora:** Winter is coming. **Q1: Which clothes will Nora buy?** **Q2: Which clothes are useful in Winter?**	
6- **Sisters in the living room:** **Noor**: Wow, what a smell. Is it a new perfume? **Donia:** No, it is these flowers. It is used in making perfumes. **Q1: Which flowers does Noor have in the living room?** **Q2**: **Which flowers are popular in the perfume industry?**	
7- **Friends talk about the pictures:** **Karl:** Plastic is one of the main reasons for pollution. Is what you are using recyclable? **Lulu:** No, it is not. (She has the item in the picture) **Karl**: do you know that it **takes around 450 years to decompose?** **Q1: What drink is Lulu drinking?** **Q2: What takes around 450 years to decompose?**	
8- **Classmates talk in art class:** **Jack:** What to draw for the art group project? Any suggestions? **Tim:** I guess animals, especially the one I am drawing right now. **Jack**: If you visit Australia, Tim, you will see lots of this. [pointing at the picture]. **Q1: Which animal is Tim drawing?** **Q2: Which animal is common in Australia?**	

B.2 Acceptability Judgement Task

Instructions:

Each story below is followed by five sentences. Read each story and judge how appropriate each of the following sentences is to complete the story.

- For **unacceptable sentences** please choose **1**.
- For **acceptable sentences** please choose **4**.
- If you are unsure about the sentence acceptance, please choose either 2 (less acceptable) or 3 (nearly acceptable).

You may give the same rating to two or more of the sentences that follow a single story.

Example:

Kevin's daughter Lauren has a favourite toy: her stuffed bear. She plays with her bear all day long and takes it with her everywhere. Even at night...

a) Kevin knows that Lauren puts her bear next to her.	1	2	3	4
b) Kevin knows that Lauren puts her bear next to herself.	1	2	3	4
c) Kevin knows that Lauren puts her bear next to him.	1	2	3	4
d) Kevin knows that Lauren puts her bear next to himself.	1	2	3	4
e) Kevin knows that Lauren puts her bear next to Lauren.	1	2	3	4

In the context of the story, it makes sense to say that the bear is next to Lauren at night: (a) and (b) are two different ways of expressing this idea, so they are rated with 4. In contrast, (c) and (d) can be interpreted only to mean that the bear is next to Kevin at night, but the context makes this unlikely; additionally, (d) sounds awkward, which is why it is given a lower rating than (c). Finally, (e) expresses an appropriate meaning, but sounds awkward – it's not as good as (a) and (b) at expressing the meaning. Therefore, it might be given a 2 (or alternatively, a 3). Note that these particular ratings reflect only one person's opinion: you may have different intuitions about these sentences. Note that these particular ratings reflect only one person's opinion: you may have different intuitions about these sentences.

Please circle your response in each of the test items in the following. There is no time limit. Answer the test items in the order given. You are not allowed to go back to change your earlier answers.

1) My friend Charles is a teacher. He really loves his job, and the children love Charles. Charles is a very experienced teacher, too:

a) Charles works as a teacher for fifteen years.	1	2	3	4
b) Charles had worked as a teacher for fifteen years.	1	2	3	4
c) Charles has worked as a teacher for fifteen years.	1	2	3	4
d) Charles is working as a teacher for fifteen years.	1	2	3	4
e) Charles worked as a teacher for fifteen years.	1	2	3	4

2) The Netherlands is a great country to visit. It has wonderful museums, great food, and excellent public transportation. And, of course, it's a great place to buy flowers. As you probably know...

a) Tulips are very popular in the Netherlands.	1	2	3	4
b) The tulip is very popular in the Netherlands.	1	2	3	4
c) Tulip is very popular in the Netherlands.	1	2	3	4
d) A tulip is very popular in the Netherlands.	1	2	3	4
e) The tulips are very popular in the Netherlands.	1	2	3	4

3) My brother has been in a bad mood lately. And no wonder – his apartment is so uncomfortable; it must be very depressing to live there. And he has a very dim and unpleasant overhead light. I told him he should buy a new lamp – something pleasant. For example, I know that...

a) A green lamp is very relaxing.	1	2	3	4
b) Green lamp is very relaxing.	1	2	3	4
c) The green lamps are very relaxing.	1	2	3	4
d) The green lamp is very relaxing.	1	2	3	4
e) Green lamps are very relaxing.	1	2	3	4

4) I am going to buy a house. James is my real-estate agent. All my friends recommend James. He is really good at his job! He knows the town really well, because...

a) James has sold houses for ten years.	1	2	3	4
b) James sells houses for ten years.	1	2	3	4
c) James is selling houses for ten years.	1	2	3	4
d) James had sold houses for ten years.	1	2	3	4
e) James sold houses for ten years.	1	2	3	4

5) I know that you like birds. Well, if you ever visit California, you'll see lots of different kinds of birds there. For example...

a) Pelican is widespread on the California coast.	1	2	3	4
b) The pelicans are widespread on the California coast.	1	2	3	4
c) The pelican is widespread on the California coast.	1	2	3	4
d) Pelicans are widespread on the California coast.	1	2	3	4
e) A pelican is widespread on the California coast.	1	2	3	4

6) I would like to give my daughter a pet for her birthday; perhaps I will give her a puppy. My daughter is going to be eight, and she is very responsible. This is really important. As everyone knows...

a) Little puppies need a lot of time and attention.	1	2	3	4
b) A little puppy needs a lot of time and attention.	1	2	3	4
c) Little puppy needs a lot of time and attention.	1	2	3	4
d) The little puppy needs a lot of time and attention.	1	2	3	4
e) The little puppies need a lot of time and attention.	1	2	3	4

7) I really like going to this museum. It's a small museum. Usually, it's almost empty. But yesterday, I came to the museum, and I heard lots of voices. Then I saw that...

a) A group of tourists were looking at the pictures.	1	2	3	4
b) A group of tourists is looking at the pictures.	1	2	3	4
c) A group of tourists looked at the pictures.	1	2	3	4
d) A group of tourists was looking at the pictures.	1	2	3	4
e) A group of tourists are looking at the pictures.	1	2	3	4

8) These woods are really beautiful. And you can do a lot in them: you can hike, pick mushrooms, and have picnics. But be very careful – don't leave food around! Otherwise, you might attract animals. You see...

a) Brown bear is common in these woods.	1	2	3	4
b) The brown bears are common in these woods.	1	2	3	4
c) A brown bear is common in these woods.	1	2	3	4
d) The brown bear is common in these woods	1	2	3	4
e) Brown bears are common in these woods.	1	2	3	4

9) It's my niece's birthday this Saturday – she is going to be three years old. I'm not sure what to get her. Maybe I'll just get her some toys, like a stuffed dog or bear. I can't go wrong with that. We all know that...

a) The toy animal is a good children's gift.	1	2	3	4
b) Toy animal is a good children's gift.	1	2	3	4
c) A toy animal is a good children's gift.	1	2	3	4
d) Toy animals are good children's gifts.	1	2	3	4
e) The toy animals are good children's gifts.	1	2	3	4

10) Last night, Ruth went to a party. She asked her roommate Clara to go with her. But Clara couldn't go because she was in the middle of her history class assignment. When Ruth left...

a) Clara has written an essay for history class.	1	2	3	4
b) Clara had written an essay for history class.	1	2	3	4
c) Clara wrote an essay for history class.	1	2	3	4
d) Clara was writing an essay for history class.	1	2	3	4
e) Clara writes an essay for history class.	1	2	3	4

11) I really like going to the zoo. Unfortunately, there are many animals that can't be found in a zoo, or anywhere else. It's very sad. For example...

a) Dodo birds are extinct.	1	2	3	4
b) A dodo bird is extinct.	1	2	3	4
c) The dodo bird is extinct.	1	2	3	4
d) The dodo birds are extinct.	1	2	3	4
e) Dodo bird is extinct.	1	2	3	4

12) My husband and I are looking for a new car. My husband would like to get a white one because white is such a beautiful color. But I'm worried about the damage. I'm worried because...

a) White car attracts attention.	1	2	3	4
b) A white car attracts attention.	1	2	3	4
c) White cars attract attention.	1	2	3	4
d) The white car attracts attention.	1	2	3	4
e) The white cars attract attention.	1	2	3	4

13) A long time ago, I had a neighbour named Robert. We were good friends. But eight years ago, Robert moved to Canada. I am sorry that he is gone. I really miss him. After all...

a) Robert had lived here for eight years.	1	2	3	4
b) Robert was living here for eight years.	1	2	3	4
c) Robert lives here for eight years.	1	2	3	4
d) Robert lived here for eight years.	1	2	3	4
e) Robert has lived here for eight years.	1	2	3	4

14) Sara is a biology teacher. She was teaching a lesson about mammals. Sara discovered students didn't know that some ocean dwellers are not fish. As we all know..

a) The whale is a mammal.	1	2	3	4
b) The whales are mammals.	1	2	3	4
c) Whales are mammals.	1	2	3	4
d) A whale is a mammal.	1	2	3	4
e) Whale is a mammal.	1	2	3	4

15) Lots of people enjoy skiing in December. Winter is very cold. Everyone knows that ...

a) A warm coat is necessary for winter.	1	2	3	4
b) Warm coats are necessary for winter.	1	2	3	4
c) Warm coat is necessary for winter.	1	2	3	4
d) The warm coat is necessary for winter.	1	2	3	4
e) The warm coats are necessary for winter.	1	2	3	4

16) My great-aunt Sara had a stroke five years ago. As a result, she does not remember how to play her violin. It's very sad. Sara loves the violin! After all...

a) Sara had played the violin for five years.	1	2	3	4
b) Sara was playing the violin for five years.	1	2	3	4
c) Sara has played the violin for five years.	1	2	3	4
d) Sara has played the violin for five years.	1	2	3	4
e) Sara played the violin for five years.	1	2	3	4

17) My youngest brother is finally convinced that being fat is associated with high carbs diets and drinks. To be healthy, he avoids high-carb food and drinks. We all know that...

a) A fizzy drink contains a lot of sugar.	1	2	3	4
b) Fizzy drinks contain a lot of sugar.	1	2	3	4
c) Fizzy drink contains a lot of sugar.	1	2	3	4
d) The fizzy drink contains a lot of sugar.	1	2	3	4
e) The fizzy drinks contain a lot of sugar.	1	2	3	4

18) Many struggling Australian species are being pushed even further toward the brink of extinction, as you know:

a)	Koalas may be extinct soon	1	2	3	4
b)	The koalas may be extinct soon	1	2	3	4
c)	A koala may be extinct soon	1	2	3	4
d)	The koala may be extinct soon	1	2	3	4
e)	Koala may be extinct soon	1	2	3	4

19) Yesterday, I went out for a walk in our neighbourhood park. There were lots of people there. At first, I didn't understand why. But then I saw that

a)	A team of athletes was running through the park.	1	2	3	4
b)	A team of athletes ran through the park.	1	2	3	4
c)	A team of athletes are running through the park.	1	2	3	4
d)	A team of athletes is running through the park.	1	2	3	4
e)	A team of athletes were running through the park.	1	2	3	4

20) Thomas left for work early this morning. But his neighbour Anne was already awake. Anne is a singer, and she practices a lot. When Thomas left...

a)	Anne has sung a song.	1	2	3	4
b)	Anne was singing a song.	1	2	3	4
c)	Anne sang a song.	1	2	3	4
d)	Anne had sung a song.	1	2	3	4
e)	Anne sings a song.	1	2	3	4

B.3 Forced-choice Task

Instruction:

Each conversation below has a **<u>blank to fill in</u>** and four options. Read each conversation carefully and **<u>choose the appropriate answer</u>** based on the context of the conversation. Circle your chosen answer. Study the provided example before starting.

Example:

Conversation: Friends' talk

Fay: Did you see Sarah yesterday?

Malak: Yes, she ____ to the store yesterday.

goes	**<u>went</u>**	going	gone

There is no time limit. Please, answer the questions in the order they are presented, and you are not allowed to go back and change your previous answers.

1.

A: I left my wallet behind this morning.

B: That's terrible! What ___ you do?

A: I returned home to get the wallet.

Options: does / do / did / had

2.

A: I think that climate change is leading some species to leave their home.

B: Seriously!

A: __ vulture is widespread in South America. In the last decade, they have been moving north as a result of warmer weather.

Options: the / a / an / Ø

3.

A: You look really tired.

B: Yes. I have been working hard for so long without any breaks.

A: But you know, ____ holiday refreshes our body and soul!

Options: the / a / an / Ø

4.

A: Hurry up or we'll miss our train. What are you doing?

B: I ___ for my keys.

A: You're so absent-minded. You just put the keys in your bag.

Options: looks / am looking / looked / is looking

5.

A: Many animals still live in the mountains while others have disappeared.

B: Give an example, please.

A: ___Atlas lions died out long ago.

Options: the/ a/ an/ Ø

6.

A: Many food experts say that people should consider eating vegetables.

B: Do you think it is important?

A: I am not sure, but __ green vegetables are highly digestible.

Options: the / a / an / Ø

7.

A: I have just seen some new girls come in.

B: Really? Where __ they?

A: That waitress is asking the girls what they want to order.

Options: is / am / are / Ø

8.

A: Do you know who Willis Haviland Carrier is?

B: No, I don't. Who is he?

A: He is someone who invented air conditioners. ___ air conditioner was invented in 1911.

Options: the / a / an / Ø

9.

A: Why are they criticizing Prof. Brown?

B: Well, ___ scientist should be able to produce evidence in support of his theory.

A: I agree.

Options: the / a / an / Ø

10.

A: My wife has just come back from Germany.

B: What was she doing there?

A: She ____ friends we hadn't seen for 20 years.

Options: is visiting / visits / was visiting / visit

11.

A: I am planning to visit North America this summer.

B: It is a great choice. You will enjoy discovering the history of some flowers.

A: For example, ____ Sunflowers were originally planted in North America.

Options: the / a / an / Ø

12.

A: I am planning to buy a new car.

B: You have to avoid buying a white car to avoid vandalism.

A: You are right. ___ white cars attract attention.

Options: the / a / an / Ø

13.

A: Excuse me.

B: How ____ I help?

A: I would like to buy a CD that I have been trying to find for ages.

Options: can / could / should / must

14.

A: This book gives interesting facts about South America.

B: Like what?

A: For example, ____ potato was first cultivated in South America.

Options: the / a / an / Ø

15.

A: Different foods have different nutritional values.

B: For example?

A: ___ red fruit contains important minerals and fibre.

Options: the / a / an / Ø

16.

A: Philippa has been shopping.

B: What did she get?

A: She ____ a book that is one of my favourites.

Options: had bought / buys / buy / bought

17.

A: The Island of Mauritius is home to some of the world's rarest plants and animals. Unfortunately, some are already extinct.

B: That is true. ___ Dodo birds disappeared 300 years ago.

Options: the / a / an / Ø

18.

A: I think about changing the wallpaper in my room.

B: Why?

A: As you know, __ white walls make spaces look bigger. I will put up some white wallpaper.

Options: the / a / an / Ø

19.

A: I took introductory linguistics courses in my first term.

B: I see you ______ several books.

A: Yes, my sister is going to do the courses next year.

Options: have kept / keep / kept / had kept

20.

A: The conservationists are making news again.

B: What are they doing now?

A: They are trying to encourage ____ oyster-catcher to come back to urban rivers.

Options: the / a / an / Ø

21.

A: I've read two interesting articles on pregnancy nutrition.

B: Oh, what do they say?

A: Both studies say ___ developing baby needs a lot of calcium, which is taken from the mother's bones.

Options: the / a / an / Ø

22.

A: Where is the bus? It was supposed to come five minutes ago!

B: The schedule ___ changed. The bus will be late.

Options: is / has / have / Ø

23.

A: As many people own cars, pollution is increasing.

B: Many car owners are worried about the environment. As a result, __ electronic cars are becoming popular.

Options: the / a / an / Ø

24.

A: Sometimes at this time of year, small garden animals need our help. For example:

B: ____ hungry hedgehogs love milk.

Options: the / a / an / Ø

Appendix C Examples of instructional activities

Activity : Error Correction with Justification

Focus: Identifying and correcting subject-related grammatical errors .

Objective:
Learners will identify and correct grammatical errors involving noun phrases in the **subject position**, then **justify** the correction with grammatical reasoning based on generic form-meaning mapping.

Instructions for Students:

Each of the sentences below contains a **mistake**.

Your tasks:

1. **Correct the sentence.**
2. **Justify your correction** by explaining the grammatical rule.

Sentences to Correct:

1. **The educational games support learning.**
   - **Correction:** Educational games support learning.
   - **Justification:** "Educational games" is a general plural subject; the sentence is referring to a regulation about educational games and no article is needed.
2. **The word puzzle improves spelling and activate the brain.**
   - **Correction:** ....
   - **Justification:** ....
3. **The sweets increase the risk of tooth decay.**
   - **Correction:** ....
   - **Justification:** ....
4. **The strawberries are a good source of antioxidants.**
   - **Correction:**

- **Justification:**

5. **a robot was invented in 1957.**
    - **Correction:**..
    - **Justification:** ....

Activity 3: Picture-Based Sentence Production

Objective:

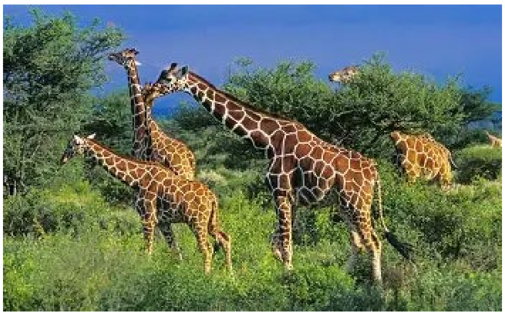
- To help learners produce generic sentences (characterizing and kind plurals/singulars) based on images.
- To test their understanding of the correct use of **articles** in both characterising and kind generics

---

Instructions for Students:

1. Look at the picture carefully.
2. Write two sentences that describe each picture.
3. Your sentences should make a general statement about the subject in the picture, similar to the examples below.

Example Sentences for Guidance:



- *Giraffes eat leaves.*
- *The giraffe is widespread in Africa.*

1.  2.  3. 

4.  5. 

Discussion Questions for Learners:

- Why did you choose “a” or no article in your sentence?
- Is your sentence about a specific object or a whole category? Is it generic or specific?
- Could you change your sentence to the other form (from plural to singular or vice versa)?

Appendix D Results

D.1 Pe-intervention statistics

D.1.1 Elicited written production task

Model 1: L2 learners to the native baseline

Predictors:

P_S_P Chracterising generics plural form.

P_S_S Characterising generics singular form.

P_NP_S Kind generics singular form

Reference level: bare plural in plural kind generics

	Bare singulars (1)	**Definite plurals (2)**	**Definite singulars (3)**	**Indefinite singulars (4)**	**Others (5)**
GroupLearners	5.459***	3.971***	2.369***	0.877**	2.496***
	(1.028)	(1.019)	(0.452)	(0.440)	(0.495)
ConditionP_NPS	2.874***	-0.471	3.706***	13.353	0.012
	(0.486)	(0.654)	(1.061)	(263.653)	(0.492)
ConditionP_S_P	0.017	0.146	1.662	11.810	-1.002**
	(0.490)	(0.416)	(1.067)	(263.653)	(0.432)
ConditionP_S_S	1.645***	0.084	2.967***	13.858	0.088
	(0.471)	(0.481)	(1.055)	(263.653)	(0.419)
Constant	-6.611***	-4.729***	-5.663***	-15.524	-2.950***
	(1.100)	(1.057)	(1.099)	(263.653)	(0.545)
Akaike Inf. Crit.	1,492.237	1,492.237	1,492.237	1,492.237	1,492.237

Note: *p<0.1; **p<0.05; ***p<0.01

Model 2: Comparison to experimental groups' pre-test EWPT

	Bare singulars (1)	Definite plurals (2)	Definite singulars (3)	Indefinite singulars (4)	Others (5)
GPExperimental	0.243 (0.847)	0.754 (0.719)	25.058*** (0.629)	-3.456*** (0.563)	3.462*** (1.118)
ConditionP_NPS	4.985*** (1.132)	-11.143 (442.365)	29.217*** (0.979)	26.319*** (0.871)	2.833 (1.749)
ConditionP_S_P	-0.131 (0.645)	-0.196 (0.607)	25.313*** (0.715)	21.276*** (0.821)	0.498 (1.193)
ConditionP_S_S	2.690*** (0.681)	0.636 (0.801)	27.137*** (0.750)	24.864*** (0.560)	2.987** (1.170)
GPExperimental: ConditionP_NPS	-2.873** (1.368)	10.515 (442.366)	-27.138*** (1.052)	-0.455 (1.211)	-3.685** (1.864)
GPExperimental: ConditionP_S_P	0.173 (1.014)	0.292 (0.870)	-4.578*** (0.818)	3.754*** (1.164)	-2.065 (1.331)
GPExperimental: ConditionP_S_S	-2.066** (1.022)	-1.146 (1.046)	-25.489*** (0.811)	1.405* (0.814)	-4.309*** (1.308)
Constant	-1.224** (0.509)	-1.041** (0.475)	-27.138*** (0.629)	-24.710*** (0.399)	-2.833*** (1.029)
Akaike Inf. Crit.	1,274.43	1,274.43	1,274.432	1,274.432	1,274.432

Note: *p<0.1; **p<0.05; ***p<0.01

Model 2 post hoc pairwise comparisons

Contrast	df	df	F.ratio	P value
Comparison P_NPS - Experimental P_NPS Bare plural	1	40	8.214	.007*
Comparison P_S_P - Experimental P_S_P Bare plural	1	40	5.213	.03
Comparison P_S_S - Experimental P_S_S Bare plural	1	40	7.115	.01*
Comparison P_NPP - Experimental P_NPP Bare singular	1	40	0.894	.35
Comparison P_NPS - Experimental P_NPS Bare singular	1	40	9.162	.004*
Comparison P_S_P - Experimental P_S_P Bare singular	1	40	0.009	.92

Contrast	df	df	F.ratio	P value
Comparison P_S_S - Experimental P_S_S Bare singular	1	40	10.069	.003*
Comparison P_NPP - Experimental P_NPP Definite plural	1	40	0.062	.80
Comparison P_NPS - Experimental P_NPS Definite plural	1	40	4.267	.05
Comparison P_S_P - Experimental P_S_P Definite plural	1	40	2.945	.09
Comparison P_S_S - Experimental P_S_S Definite plural	1	40	1.232	.27
Comparison P_NPP - Experimental P_NPP Definite singular	1	40	1.032	.31
Comparison P_NPS - Experimental P_NPS Definite singular	1	40	0.082	.78
Comparison P_S_P - Experimental P_S_P Definite singular	1	40	0.001	.97
Comparison P_S_S - Experimental P_S_S Definite singular	1	40	1.683	.20
Comparison P_NPP - Experimental P_NPP indefinite singular	1	40	5.317	.03
Comparison P_NPS - Experimental P_NPS indefinite singular	1	40	3.116	.09
Comparison P_S_P - Experimental P_S_P indefinite singular	1	40	0.013	.91
Comparison P_S_S - Experimental P_S_S indefinite singular	1	40	2.600	.11

Note: significant at .01 level

D.1.2 Acceptability Judgement Task

Tukey-corrected Pairwise comparisons for the native baseline

NPL = kind generics, SL = Characterising generics, BP = bare plural, BS = bare singular, DefP = definite plural, DefS = definite singular, INDF = indefinite singular.

Contrast	Estimate	SE	df	z.ratio	p.value
NPL_BP - NPL_BS	6.0719	0.365	Inf	16.624	<.001
NPL_BP - NPL_DefP	4.7524	0.343	Inf	13.859	<.001
NPL_BP - NPL_DefS	2.5702	0.326	Inf	7.873	<.001
NPL_BP - NPL_INDF	4.9689	0.355	Inf	13.997	<.001
NPL_BP - SL_BP	0.0494	0.391	Inf	0.127	1
NPL_BP - SL_BS	5.9650	0.377	Inf	15.820	<.001
NPL_BP - SL_DefP	5.6525	0.368	Inf	15.355	<.001

Contrast	Estimate	SE	df	z.ratio	p.value
NPL_BP - SL_DefS	5.3550	0.367	Inf	14.587	<.001
NPL_BP - SL_INDF	1.3536	0.349	Inf	3.874	0.004
NPL_BS - NPL_DefP	-1.3194	0.255	Inf	-5.176	<.001
NPL_BS - NPL_DefS	-3.5017	0.277	Inf	-12.646	<.001
NPL_BS - NPL_INDF	-1.1030	0.270	Inf	-4.079	0.002
NPL_BS - SL_BP	-6.0224	0.373	Inf	-16.140	<.001
NPL_BS - SL_BS	-0.1069	0.283	Inf	-0.378	1
NPL_BS - SL_DefP	-0.4193	0.273	Inf	-1.535	0.88
NPL_BS - SL_DefS	-0.7169	0.279	Inf	-2.574	0.23
NPL_BS - SL_INDF	-4.7183	0.318	Inf	-14.818	<.001
NPL_DefP - NPL_DefS	-2.1822	0.252	Inf	-8.676	<.001
NPL_DefP - NPL_INDF	0.2164	0.254	Inf	0.852	0.99
NPL_DefP - SL_BP	-4.7030	0.351	Inf	-13.387	<.001
NPL_DefP - SL_BS	1.2126	0.272	Inf	4.454	0.004
NPL_DefP - SL_DefP	0.9001	0.261	Inf	3.448	0.02
NPL_DefP - SL_DefS	0.6026	0.265	Inf	2.276	0.41
NPL_DefP - SL_INDF	-3.3989	0.295	Inf	-11.536	<.001
NPL_DefS - NPL_INDF	2.3987	0.268	Inf	8.945	<.001
NPL_DefS - SL_BP	-2.5208	0.335	Inf	-7.516	<.001
NPL_DefS - SL_BS	3.3948	0.292	Inf	11.612	<.001
NPL_DefS - SL_DefP	3.0823	0.281	Inf	10.965	<.001
NPL_DefS - SL_DefS	2.7848	0.282	Inf	9.878	<.001
NPL_DefS - SL_INDF	-1.2166	0.280	Inf	-4.345	0.006
NPL_INDF - SL_BP	-4.9195	0.363	Inf	-13.545	<.001
NPL_INDF - SL_BS	0.9961	0.287	Inf	3.474	0.02
NPL_INDF - SL_DefP	0.6836	0.276	Inf	2.473	0.28
NPL_INDF - SL_DefS	0.3861	0.280	Inf	1.378	0.93

Contrast	Estimate	SE	df	z.ratio	p.value
NPL_INDF - SL_INDF	-3.6153	0.309	Inf	-11.706	<.001
SL_BP - SL_BS	5.9156	0.359	Inf	16.458	<.001
SL_BP - SL_DefP	5.6031	0.350	Inf	16.004	<.001
SL_BP - SL_DefS	5.3056	0.349	Inf	15.201	<.001
SL_BP - SL_INDF	1.3041	0.330	Inf	3.949	0.003
SL_BS - SL_DefP	-0.3125	0.255	Inf	-1.225	0.96
SL_BS - SL_DefS	-0.6100	0.261	Inf	-2.340	0.36
SL_BS - SL_INDF	-4.6114	0.302	Inf	-15.255	<.001
SL_DefP - SL_DefS	-0.2975	0.250	Inf	-1.192	0.97
SL_DefP - SL_INDF	-4.2990	0.291	Inf	-14.753	<.001
SL_DefS - SL_INDF	-4.0014	0.291	Inf	-13.750	<.001

Model 1: Native baseline to L2 learners.

formula: Ratings ~ Group * Conditions + (1 | id) + (1 | item)

data: Pretest_AJT

Random effects:

Groups Name Variance Std.Dev.

id (Intercept) 0.282 0.53

item (Intercept) 0.002 0.05

Number of groups: id 84, item 12

	Estimate	Std. Error	z value	Pr(>|z|)
GroupNS	2.245	0.31	7.198	.001 ***
ConditionsNPL_BS	-0.311	0.13	-2.306	.02 *
ConditionsNPL_DefP	-0.157	0.13	-1.159	.24
ConditionsNPL_DefS	-0.089	0.13	-0.673	.50
ConditionsNPL_INDF	-0.938	0.13	-6.925	.001 ***
ConditionsSL_BP	-0.242	0.13	-1.745	.08.
ConditionsSL_BS	-0.471	0.13	-3.454	.001 ***

	Estimate	Std. Error	z value	Pr(>|z|)
ConditionsSL_DefP	-0.280	0.13	-2.02	.04 *
ConditionsSL_DefS	-0.150	0.13	-1.085	.27
ConditionsSL_INDF	-0.185	0.13	-1.332	.18
GroupNS:ConditionsNPL_BS	-4.637	0.34	-13.352	< .001 ***
GroupNS:ConditionsNPL_DefP	-3.798	0.33	-11.365	< .001 ***
GroupNS:ConditionsNPL_DefS	-2.201	0.33	-6.524	.001 ***
GroupNS:ConditionsNPL_INDF	-3.187	0.34	-9.273	< .001 ***
GroupNS:ConditionsSL_BP	0.193	0.39	0.495	.62
GroupNS:ConditionsSL_BS	-4.390	0.34	-12.693	< .001 ***
GroupNS:ConditionsSL_DefP	-4.320	0.33	-12.74	< .001 ***
GroupNS:ConditionsSL_DefS	-4.254	0.34	-12.455	< .001 ***
GroupNS:ConditionsSL_INDF	-1.047	0.35	-2.987	.003**

Note*:** Signif. codes: 0 '' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Model 1 post hoc Tukey-corrected pairwise comparison

contrast	estimate	SE	df	z.ratio	p.value
L2 learners NPL_BP - NS NPL_BP	-2.245	0.31	Inf	-7.198	< .001'***
L2 learners NPL_BS - NS NPL_BS	2.392	0.24	Inf	9.835	< .001'***
L2 learners NPL_DefP - NS NPL_DefP	1.553	0.22	Inf	6.902	< .001'***
L2 learners NPL_DefS - NS NPL_DefS	-0.043	0.23	Inf	-0.188	.99
L2 learners NPL_INDF - NS NPL_INDF	0.942	0.23	Inf	3.936	.02**
L2 learners SL_BP - NS SL_BP	-2.438	0.30	Inf	-7.957	< .001***
L2 learners SL_BS - NS SL_BS	2.145	0.24	Inf	8.891	< .001***
L2 learners SL_DefP - NS SL_DefP	2.075	0.23	Inf	8.965	< .001***
L2 learners SL_DefS - NS SL_DefS	2.009	0.23	Inf	8.538	< .001***
L2 learners SL_INDF - NS SL_INDF	-1.197	0.25	Inf	-4.750	.001**

Note: Significance. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1, NS= native baseline

Model 2: L2 learners pretest Characterising generics, a separate model.

formula: clmm(Ratings ~ Conditions + (1 | id) + (1 | item))

Reference level: bare plurals

Random effects:

Groups Name Variance Std.Dev.

id (Intercept) 0.12 0.35

item (Intercept) 0.002 0.048

Number of groups: id 64, item 6

Condition	**Estimate**	**Std. Error**	**z value**	***P***
ConditionsSL_BS	-0.218	0.13	-1.685	0.09
ConditionsSL_DefP	-0.028	0.13	-0.218	0.82
ConditionsSL_DefS	0.089	0.13	0.676	0.49
ConditionsSL_INDF	0.060	0.13	0.452	0.65

Note: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Model 3: Pre-test between L2 learners' groups.
Clmm (Ratings ~ Conditions * GP + GP + (1 | id) + (1 | item)
reference NPL_Bare plural
Random effects:
Groups Name Variance Std.Dev.
id (Intercept) 0.18 0.42
item (Intercept) 0.003 0.02
Number of groups: id 64, item 12

Coefficients	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
ConditionsNPL_BS	-0.25325	0.200029	-1.266	0.2055
ConditionsNPL_DefP	-0.20647	0.200054	-1.032	0.302
ConditionsNPL_DefS	0.010642	0.19792	0.054	0.9571
ConditionsNPL_INDF	-0.85633	0.196271	-4.363	1.28e-05 ***
ConditionsSL_BP	-0.28271	0.20019	-1.412	0.1579
ConditionsSL_BS	-0.24659	0.199619	-1.235	0.2167
ConditionsSL_DefP	0.002455	0.201195	0.012	0.9903

Coefficients	Estimate	Std. Error	z value	Pr(>|z|)
ConditionsSL_DefS	-0.18705	0.19881	-0.941	0.3468
ConditionsSL_INDF	-0.25463	0.199178	-1.278	0.2011
GPExperimental	0.169087	0.220398	0.767	0.443
ConditionsNPL_BS:GPExperimental	-0.08342	0.268358	-0.311	0.7559
ConditionsNPL_DefP:GPExperimental	0.095975	0.269879	0.356	0.7221
ConditionsNPL_DefS:GPExperimental	-0.18351	0.265716	-0.691	0.4898
ConditionsNPL_INDF:GPExperimental	-0.066	0.267004	-0.247	0.8048
ConditionsSL_BP:GPExperimental	0.092982	0.269715	0.345	0.7303
ConditionsSL_BS:GPExperimental	-0.36723	0.265637	-1.382	0.1668
ConditionsSL_DefP:GPExperimental	-0.49152	0.269927	-1.821	0.0686 .
ConditionsSL_DefS:GPExperimental	0.077061	0.26906	0.286	0.7746
ConditionsSL_INDF:GPExperimental	0.150283	0.270958	0.555	0.5791

Significance. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

D.1.3 Forced choice task

Model 1: Native baseline to L2 learners model

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']

Family: binomial (logit)

Formula: Accuracy ~ GP + (1 | item) + (1 | id)

Data: FCT

Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 20000))

AIC   BIC   logLik deviance df.resid

1531.6   1552.4   -761.8   1523.6   1340

Scaled residuals:

Min   1Q Median   3Q   Max

-3.4794 -0.6465 -0.5112 0.8335 2.1389

Random effects:

Groups Name   Variance Std.Dev.

id (Intercept) 0.05983 0.2446

item (Intercept) 0.29018 0.5387

Number of obs: 1344, groups: id, 84; item, 16

Fixed effects:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.6861 0.1546 -4.438 9.07e-06 ***

GPNS 2.6980 0.1982 13.615 < 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

BTFCTnl2: Accuracy ~ GP + (1 | item) + (1 | id)

BTFCTnl1: Accuracy ~ Condition * GP + (1 | item) + (1 | id)

npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)

BTFCTnl2 4 1531.6 1552.4 -761.80 1523.6

BTFCTnl1 10 1534.1 1586.1 -757.04 1514.1 9.5205 6 0.1464

Model 2: Comparison vs. Experimental group pre-test

Formula: glmer(Accuracy ~ Group + Condition + Condition * Group + (1 | item) + (1 | id)

Random effects:
Groups Name Variance Std.Dev.
id (Intercept) 0.06 0.25
item (Intercept) 0.14 0.38
Number of obs: 1024, groups: id, 64; item, 16

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	-1.383	0.3053	-4.530	< .001***
Group Experimental	0.2889	0.3086	.936	.35
Condition Bare plural in Kind generics	0.414	0.4	0.99	.32
Condition Definite singular NP in kind generics	1.084	0.40	2.66	.007**
Condition Indefinite singular NP in characterising generics	1.167	0.40	2.88	.003 **

	Estimate	Std. Error	z value	Pr(>|z|)
Group Experimental: Condition Bare plural in Kind generics	-0.140	0.41	-0.34	.73
Group Experimental: Condition Definite singular NP in kind generics	-0.331	0.40	-0.83	.41
Group Experimental: Condition Indefinite singular NP in characterising generics	-0.494	0.39	-1.24	.21

Note: Significance. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Model 2 post hoc pairwise comparisons

Contrast	**Estimate**	**SE**	**DF**	**Z value**	**p**
Comparison to experimental Indefinite singular NP in characterising generics	0.21	.26	Inf	0.77	.99
Comparison to experimental Bare plural NP in characterising generics	-0.28	.31	Inf	- 094	.98
Comparison to experimental definite singular NP in kind generics	0.04	0.27	Inf	0.157	1
Comparison to experimental Bare plural NP in kind generics	-0.14	0.29	Inf	-0.52	.99

Between L2 Ex vs Comparison groups (group only model)

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']

Family: binomial (logit)

Formula: Accuracy ~ Group + (1 | item) + (1 | id)

Data: L2FCT

Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 20000))

```
   AIC      BIC   logLik deviance df.resid
 1289.4  1309.1  -640.7   1281.4     1020
```

Scaled residuals:

```
    Min      1Q  Median      3Q     Max
-1.3066 -0.6571 -0.5763  1.0330  1.9980
```

Random effects:

Groups Name Variance Std.Dev.

id (Intercept) 0.06365 0.2523

item (Intercept) 0.29126 0.5397

Number of obs: 1024, groups: id, 64; item, 16

Fixed effects:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.70121 0.17572 -3.991 6.59e-05 ***

GroupExperimenatl 0.02603 0.15085 0.173 0.863

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

D.2 Post-intervention results

D.2.1 Elicited Written Production Task

Model 1: Between groups post-test results

predictor	**Bare singular** (2)	**Definite plural** (3)	**Definite singular** (4)	**indefinite singular** (5)	**others** (6)
GroupComparison	3.587*** (1.108)	0.106 (0.923)	21.519*** (0.485)	22.671*** (0.523)	-16.421*** (0.551)
ConditionP_NPS	4.605*** (1.290)	-16.726*** (0.00000)	26.281*** (0.704)	25.156*** (0.766)	-14.669*** (0.00000)
ConditionP_S_P	0.041 (1.249)	0.446 (0.572)	20.250*** (0.642)	19.899*** (0.886)	-2.3335
ConditionP_S_S	4.030*** (1.184)	-12.721*** (0.651)	25.004*** (0.615)	25.562*** (0.648)	1.545* (0.935)
GroupComparison: ConditionP_NPS	-2.999** (1.442)	-9.338*** (0.000)	-23.613*** (0.832)	-25.667*** (1.219)	2.863*** (0.000)
GroupComparison: ConditionP_S_P	-0.772 (1.362)	0.320 (1.036)	-20.272*** (0.771)	-21.936*** (1.252)	17.849*** (0.908)

GroupComparison: ConditionP_S_S	-3.664*** (1.309)	11.740*** (0.651)	-23.276*** (0.700)	-25.850*** (0.856)	14.917*** (0.832)
Constant	-3.219*** (1.020)	-1.610*** (0.490)	-23.023*** (0.485)	-23.770*** (0.523)	-1.833*** (0.539)
Akaike Inf. Crit.	1,078.43	1,078.43	1,078.43	1,078.43	1,078.43

Note: **p*<0.1; ***p<0.05; ***p<0.01*;

Model 2: Comparison group post-test to pre-test

predictor	**Bare singular (2)**	**Definite plural (3)**	**Definite singular (4)**	**indefinite singular (5)**	**others (6)**
ConditionP_NPS	1.741*** (0.568)	-17.226*** (0.00000)	3.415*** (0.903)	0.236 (1.259)	-5.374*** (0.796)
ConditionP_S_P	-0.597 (0.448)	-0.096 (0.536)	0.725 (0.892)	-1.290 (1.196)	13.082*** (0.677)
ConditionP_S_S	0.500 (0.467)	-1.844* (1.112)	2.475*** (0.838)	0.459 (0.896)	13.732*** (0.687)
TestPretest	-1.029** (0.445)	-0.595 (0.544)	-1.182 (1.259)	-1.589 (1.191)	12.783*** (0.526)
ConditionP_NPS:TestPretest	2.815** (1.204)	3.643*** (0.00000)	2.098 (1.724)	4.807** (1.954)	8.115*** (0.796)
ConditionP_S_P:TestPretest	0.037 (0.679)	0.096 (0.757)	0.884 (1.435)	1.290 (1.869)	-12.676*** (0.838)
ConditionP_S_S:TestPretest	1.761** (0.726)	2.675** (1.340)	0.959 (1.438)	3.129** (1.465)	-10.837*** (0.823)
Constant	0.234 (0.307)	-0.642 (0.391)	-2.251*** (0.743)	-1.846*** (0.621)	-15.524*** (0.526)
Akaike Inf. Crit.	1,231.01	1,231.01	1,231.01	1,231.01	1,231.01

*Note: *p<0.1; **p<0.05; ***p<0.01;*

Appendix D

Model 3: Experimental group post-test to pre-test

predictor	**Bare singular (2)**	**Definite plural (3)**	**Definite singular (4)**	**indefinite singular (5)**	**others (6)**
TestPretest	2.383*** (0.822)	1.216** (0.486)	1.573 (1.251)	21.689*** (0.650)	2.326*** (0.611)
ConditionP_NPS	4.663*** (1.070)	-12.128*** (0.342)	7.229*** (1.240)	26.166*** (0.857)	-11.169*** (0.287)
ConditionP_S_P	0.099 (1.020)	0.505 (0.454)	1.198 (1.172)	20.908*** (0.966)	-1.287 (1.136)
ConditionP_S_S	4.088*** (0.938)	- 13.601*** (0.264)	5.952*** (1.142)	26.571*** (0.753)	2.296** (0.923)
TestPretest:ConditionP_NPS	-2.638** (0.342)	11.664*** (1.197)	- 4.831*** (1.511)	-25.378*** (1.249)	11.204*** (0.287)
TestPretest:ConditionP_S_P	-0.143 (1.162)	-0.244 (0.648)	-0.144 (1.459)	-20.952*** (1.310)	0.606 (1.244)
TestPretest:ConditionP_S_S	-3.551*** (1.076)	13.255*** (0.264)	-3.984*** (1.406)	-25.378*** (0.925)	-2.732*** (1.052)
Constant	-3.277*** (0.720)	-1.668*** (0.345)	-3.971*** (1.009)	-24.780*** (0.650)	-2.584*** (0.519)
Akaike Inf. Crit.	1,435.834	1,435.834	1,435.834	1,435.834	1,435.834

Note: **p<0.1; **p<0.05; ***p<0.01;*

Model 4: Delayed post-test to post-test for the experimental group

predictor	Bare singular (2)	Definite plural (3)	Definite singular (4)	indefinite singular (5)	others (6)
TestPost-test2	0.446 (1.450)	-1.163 (1.141)	-3.848*** (0.334)	-3.505*** (0.468)	0.159 (0.828)
ConditionP_NPS	4.605*** (1.290)	-12.592*** (0.584)	22.662*** (0.576)	22.514*** (0.658)	-11.798 (644.311)
ConditionP_S_P	0.041 (1.249)	0.446 (0.572)	16.632*** (0.499)	17.257*** (0.795)	- 2.039* (1.145)
ConditionP_S_S	4.030*** (1.184)	-14.994*** (0.00001)	21.386*** (0.463)	22.920*** (0.517)	1.545* (0.935)
TestPost-test2:ConditionP_NPS	-1.427 (1.701)	14.672*** (0.584)	0.485 (0.690)	1.426* (0.826)	-8.845*** (0.544)
TestPost-test2:ConditionP_S_P	1.046 (1.691)	-0.969 (1.558)	5.117*** (0.625)	4.080*** (1.117)	1.110 (1.498)
TestPost-test2:ConditionP_S_S	-0.816 (1.627)	-3.245*** (0.00000)	1.972*** (0.599)	1.595** (0.695)	-0.970 (1.309)
Constant	-3.219*** (1.020)	-1.609*** (0.490)	-19.404*** (0.269)	-21.128*** (0.348)	-1.833*** (0.539)
Akaike Inf. Crit.	962.868	962.868	962.868	962.868	962.868

Note: *p<0.1; **$p<0.05$; ***$p<0.01$;

D.2.2 Acceptability Judgement task

Model 1: Between experimental and comparison groups post-test AJT characterising generic .

formula: clmm(Ratings ~ GP + GP * Conditions + (1 | id) + (1 | item)

Random effects:

Groups	Name	Variance	Std.Dev.
id	(Intercept)	0.247	0.497

item (Intercept) 0.000 0.000

Number of groups: id 64, item 6

	Estimate	Std. Error	z value	Pr(>|z|)
GPExperimental	0.79	0.22	3.51	< .001 ***
ConditionsBare Singular	-.20	0.19	-1.05	.29
ConditionsDefinite plural	0.14	0.19	0.75	.45
ConditionsDefinite singular	-0.0001	0.19	0.00	.99
ConditionsIndefinite singular	-0.25	0.19	-1.29	.20
GPExperimental:ConditionsBare Singular	-1.59	0.26	-5.99	<.001***
GPExperimental:ConditionsDefinite plural	-1.92	0.27	-7.10	<.001***
GPExperimental:ConditionsDefinite singular	-0.88	0.27	-3.31	<.001***
GPExperimental:ConditionsIndefinite singular	0.03	0.27	0.11	.92

Note: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Model 2: Between experimental and comparison groups post-test AJT characterising generic.

formula: clmm(Ratings ~ GP * Conditions + (1 | id) + (1 | item)

Random effects:

Groups Name Variance Std.Dev.

id (Intercept) 0.14 0.38

item (Intercept) 0.00 0.0002

Number of groups: id 64, item 6

	Estimate	Std. Error	z value	Pr(>|z|)
PExperimental	0.68	0.22	3.99	< .001 ***
ConditionsBare Singular	-.05	0.19	-0.27	.78
ConditionsDefinite plural	- 0.03	0.19	-0.16	.87
ConditionsDefinite singular	-0.07	0.19	-0.41	.68
ConditionsIndefinite singular	-0.20	0.19	-1.07	.28
GPExperimental:ConditionsBare Singular	-1.72	0.27	-6.39	<.001***
GPExperimental:ConditionsDefinite plural	-1.98	0.27	-7.26	<.001***
GPExperimental:ConditionsDefinite singular	-0.18	0.27	-0.68	.50
GPExperimental:ConditionsIndefinite singular	-0.88	0.27	-3.35	.001*

Note: Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Model 3: Comparison group between pre-test and post-test AJT for characterising generics

formula: Ratings ~ Test * Conditions + (1 | id) + (1 | item)

Random effects:

Groups	Name	Variance	Std.Dev.
id	(Intercept)	0.14	0.36
item	(Intercept)	0.00	0.00

Number of groups: id 29, item 6

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
Test, Pre-test	0.19	0.20	0.95	.34
ConditionsBare Singular	-0.22	0.19	-1.19	.23
ConditionsDefinite plural	0.12	0.20	0.62	.53
ConditionsDefinite singular	-0.03	0.19	-0.14	.88
ConditionsIndefinite singular	-0.26	0.20	-1.25	.17
Test, Pre-test:ConditionsBare Singular	0.27	0.27	0.98	.32
Test, Pre-test:ConditionsDefinite plural	0.14	0.27	0.52	.60
Test, Pre-test: ConditionsDefinite singular	0.12	0.27	0.45	.56
Test, Pre-test: Conditions Indefinite singular	0.31	0.27	1.11	.27

Note: Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Model 4: Comparison group between pre-test and post-test AJT for kind generics

formula: clmm(Ratings ~ Test * Conditions + (1 | id) + (1 | item))

Random effects:

Groups	Name	Variance	Std.Dev.
id	(Intercept)	0.11	0.33
item	(Intercept)	0.00	0.00

Number of groups: id 29, item 6

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
Test, Pre-test	0.14	0.20	2.09	.03*
ConditionsBare Singular	-0.05	0.20	-0.30	.67

Appendix D

	Estimate	Std. Error	z value	Pr(>|z|)
ConditionsDefinite plural	-0.03	0.20	-0.18	.86
ConditionsDefinite singular	-0.08	0.20	-0.45	.65
ConditionsIndefinite singular	-0.22	0.20	-1.17	.24
Test, Pre-test: Conditions Bare Singular	-0.17	0.27	-0.61	.53
Test, Pre-test: Conditions Definite plural	-0.17	0.27	-0.60	.55
Test, Pre-test: Conditions Definite singular	0.11	0.27	0.41	.68
Test, Pre-test: Conditions Indefinite singular	-0.58	0.27	-2.13	.03*

Note: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Appendix E Ethical Approval

Outlook

Approved by Faculty Ethics Committee - ERGO II 71758

From <ergo2@soton.ac.uk>
Date Fri 2022-04-08 22:04
To M.A.A.Jallalah@soton.ac.uk <M.A.A.Jallalah@soton.ac.uk>

Approved by Faculty Ethics Committee - ERGO II 71758

ERGO II – Ethics and Research Governance Online https://www.ergo2.soton.ac.uk

Submission ID: 71758
Submission Title: The impact of instruction on the acquisition of English genericity by Arabic-speaking learners of English: An experimental study
Submitter Name: Manal Jallalah

Your submission has now been approved by the Faculty Ethics Committee. You can begin your research unless you are still awaiting any other reviews or conditions of your approval.

Comments:

-

Click here to view the submission

TId: 23011_Email_to_submitter___Approval_from_Faculty_Ethics_committee__cat_B___C_ Id: 461303

M.A.A.Jallalah@soton.ac.uk coordinator

References

Aboras, A. (2020). The role of first language transfer in the acquisition of definiteness in specific and generic contexts by Saudi-Arabic learners of English (PhD thesis). Swansea University.

Abudalbuh, M. (2016). The acquisition of English articles by Arabic L2-English learners: A semantic approach. *Arab World English Journal*, 7(2), 104-117. https://doi.org/10.24093/awej/vol7no2.7

Abumelha, M. (2019). The role of the input in the acquisition of English generics by L1 Najdi Arabic speakers. In R. Slabakova, J. Corbet, L. Domínguez, A. Dudley, & A. Wallington (Eds.), *Explorations in second language acquisition and processing*, pp. 138-157. Cambridge Scholars Publishing.

Alhaysony, M. (2012). An analysis of article errors among Saudi female EFL students: A case study. *Asian Social Science, 8*(12), 55-72. https://doi.org/10.5539/ass.v8n12p55

Al-Qadi, M. (2017). English article system errors among Saudi Arab EFL learners: A case of the preparatory year English program learners of King Saud University. *International Journal of English Language & Translation Studies, 5*(1), 69-77.

Alzamil, A. (2019). Sentence –level vs. NP –level genericity: Are arabic learners of English sensitive to genericity type? *Advances in Language and Literary Studies, 10*(2), 95 –101. doi:http://dx.doi.org/10.7575/aiac.alls.v.10n.2p.95

Baker, M. (2008). The macroparameter in a microparametric world. In T. Biberauer (ed.) *The limits of syntactic variation*. Amsterdam: John Benjamins, pp. 351-373. doi: 10.1075/la.132.16bak

Bataineh, R. (2005). Jordanian undergraduate EFL students' errors in the use of the indefinite article. *The Asian EFL Journal Quarterly, 7*(1), 56-76.

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi:10.18637/jss.v067.i01

Biber, D., et al. (1999). *Longman grammar of spoken and written English*. Edinburgh: Pearson Education Ltd.

Borer, H. (1984). *Parametric Syntax: Case studies in Semitic and Romance languages*. Dordrecht: Foris.

Bowles, M. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition*, 33(2), 247-271. https://doi.org/10.1017/S0272263110000756

Carlson, G. (1977). *Reference to kinds in English*. Doctoral Dissertation, University of Massachusets, Amherst.

Carlson, G. (2011). Genericity. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 2, pp. 1153-1185). De Gruyter Mouton.

Chierchia, G. (1998). Reference to kinds across language. *Natural language semantics*, 6(4), 339 –405.

Cho, J., & Slabakova, R. (2014). Interpreting definiteness in a second language without articles: The case of L2 Russian. *Second Language Research*, 30(2), 159-190.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Mass: Cambridge.

Chomsky, N. (1981). *Lectures on government and binding: The Pisa Lectures*. Dordrecht: Foris.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N. (2000). Minimalist Inquiries: The Framework. In R. Martin et al. (Eds.), Step by Step. *Essays on Minimalist Syntax in Honor of Howard Lasnik* (pp. 89-155). Cambridge, MA: MIT Press.

Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 1-52). MIT Press.

Chomsky, N. (2005). Three Factors in Language Design, *Linguistic Inquiry*, 36(1), pp. 1–28. doi: 10.1162/0024389052993655.

Christensen, R. (2023). Ordinal: Regression models for ordinal data (Version 2023.12-4.1) [R package]. https://CRAN.R-project.org/package=ordinal.

Crompton, P. (2011). Immediate postoperative recovery of the surgical patient – part two. *Vet Times,* 11(10), 30-32.

Dahl, Ö. (1995). The marking of the episodic/generic distinction in tense-aspect systems. In G. N. Carlson & F. J. Pelletier (Eds.), *The Generic Book* (pp. 412-425). University of Chicago Press.

Dayal, V. (2004). Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy*, 27, 393 –450.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313-348). Oxford: Blackwell Publishing. https://doi.org/10.1002/9780470756492.

DeKeyser, R. M. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 15-32). Routledge. https://doi.org/10.4324/9781315676968-2

Domínguez, L. (2014). How suitable is the macro-micro parametric distinction in acquisition? *Linguistic Approaches to Bilingualism*, 4, 10.1075/lab.4.3.04dom.

Domínguez, L., & Arche, M. J. (2021). The 'Comparative Logic' and why we need to explain interlanguage grammars. *Frontiers in Psychology*, 12, 717635. https://doi.org/10.3389/fpsyg.2021.717635

Ellis, N. C. (1994). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review, 11*, 37-56. https://doi.org/10.1075/aila.11.04ell

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27(*2), 305–352. https://doi.org/10.1017/S027226310505014X

Ellis, N. C. (2015). Implicit and explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 3-23). John Benjamins. https://doi.org/10.1075/sibil.48.01ell

Ellis, N. C., & Röver, C. (2021). The measurement of implicit and explicit knowledge. The *Language Learning Journal, 49*, 160-175.

Ellis, R. (2001). Introduction: Investigating form-focused instruction. *Language Learning, 51*(s1), 1–46. https://doi.org/10.1111/0023-8333.00152

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*(2), 141-172.

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, 40, 83-107.

Ellis, R. (2008). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3-25). Multilingual Matters. https://doi.org/10.21832/9781847691767-0031

Esteki, B. (2014). The relationship between implicit and explicit knowledge and second language proficiency. *Theory and Practice in Language Studies*, 4(7), 1520-1525. https://doi.org/10.4304/tpls.4.7.1520-1525

Fassi Fehri, A. (2004). Nominal classes, reference, and functional parameters, with particular reference to Arabic. *Linguistic Variation Yearbook*, 4(1), 41–108.

Fassi Fehri, A. (2012). *Key Features and Parameters in Arabic Grammar*. John Benjamins Publishing Company.

Gass, S. M., & Mackey, A. (Eds.). (2012). *The Routledge Handbook of Second Language Acquisition*. NY: Routledge. https://doi.org/10.4324/9780203808184

Gass, S. M., Behney, J., & Plonsky, L. (2020). S*econd Language Acquisition: An Introductory Course* (5th ed.). NY: Routledge. https://doi.org/10.4324/9781315181752

Gelman, S., & Bloom, P. (2007). Developmental changes in the understanding of generics. *Cognition, 105,* 166–183. https://doi.org/10.1016/j.cognition.2006.09.009

Gelman, S., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development, 74*, 310–325.

Gelman, S., & Tardif, T. (1998). Generic noun phrases in English and Mandarin: An examination of child-directed speech. *Cognition, 66*, 215-248.

Godfroid, A., Loewen, S., Jung, S., Park, J., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition, 37*(2), 269–297. https://doi.org/10.1017/S0272263114000850

Golzar, J., Tajik, O., & Noor, S. (2022). Convenience sampling. *International Journal of Education and Literacy Studies, 1*(2), 72-77. https://doi.org/10.22034/ijels.2022.162981

Gross, S. (2021) Linguistic Judgments as Evidence. In Allot, N., Lohndal, T., and Rey, G. (eds) *A Companion to Chomsky*. John Wiley & Sons, pp. 529–543.

Hackl, M. (2013). The syntax–semantics interface. *Lingua*, 130, 66-87. https://doi.org/10.1016/j.lingua.2013.01.010.

Hawkins, R., & Chan, C. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. *Second Language Research, 13* (3), 187-226. https://doi.org/10.1191/026765897671476153

Hermas, A. (2020). Genericity in third language English: Acquisition pattern and transfer in ultimate attainment. International Journal of Bilingualism, 24(2), 266-288. https://doi.org/10.1177/1367006919826865.

Hirakawa, M., Shibuya, M., & Endo, M. (2018). Explicit instruction, input food or study abroad: which help Japanese learners of English acquire adjective ordering? *Language Teaching Research, 23*, 158-178.

Hua, F., Ogawa, T., Fujita, S., Sone, R., & Kawahara, A. (2021). Efficient statistical modeling using the multinom function: A practical approach. *Journal of Statistical Methods, 12*(3), 345-367.

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition, 27* (2), 129–140. https://doi.org/10.1017/S0272263105050084

Ionin, T. (2012). Formal theory-based methodologies. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 30–52). Oxford: Wiley-Blackwell.

Ionin, T. (2012). Judgment and interpretation tasks in second language research. *Annual Review of Applied Linguistics, 32*, 31-50. https://doi.org/10.1017/S026719051200001X

Ionin, T., & Montrul, S. (2010). The article of L1 transfer in the interpretation with articles in plural definite in L2 English. *Language Learning, 60*, 877-925.

Ionin, T., & Montrul, S. (2023). *Second language acquisition: Introducing intervention research. Cambridge*, United Kingdom: Cambridge University Press.

Ionin, T., & Zyzik, E. (2014). Judgment and interpretation tasks in second language research. *Annual Review of Applied Linguistics, 34*, 37–64. https://doi.org/10.1017/S0267190514000026

Ionin, T., Ko, H., & Wexler, K. (2004). Article Semantics in L2 Acquisition: The Role of Specificity. *Language Acquisition*, *12*(1), 3–69. https://doi.org/10.1207/s15327817la1201_2

Ionin, T., Montrul, S., & Scutti Santos, H. (2011a). An experimental investigation of the expression of genericity in English, Spanish and Brazilian Portuguese. *Lingua, 121*, 963-985. https://doi.org/10.1016/j.lingua.2010.12.008

Ionin, T., Montrul, S., & Slabakova, R. (Eds.) (2024).*The Routledge Handbook of Second Language Acquisition, Morphosyntax, and Semantics (*pp: 195-209). NY: Routledge.

Ionin, T., Montrul, S., Kim, J., & Philippov, V. (2011b). Genericity distributions and the interpretation of determiners in second language acquisition. *Language Acquisition, 18*, 242-280.

Kang, E., Sok, S., & Han, Z. (2019). Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research, 23*(4), 428-453. https://doi.org/10.1177/1362168818776671

Khemlani, S., Leslie, S., & Glucksberg, S. (2012). Inferences about members of kinds: The generics hypothesis. *Language and Cognitive Processes, 27*(6), 887-900. https://doi.org/10.1080/01690965.2011.601900

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

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

Krifka, M., Pelletier, F.J., Carlson, G., ter Meulen, A., Chierchia, G. and Link. G., (1995). Genericity: An Introduction. In G. Carlson, and F. Pelletier, (eds.), *The generic book*. Chicago: University of Chicago.Press, pp.1-124.

Kupisch, T., & Snape, N. (2024). Genericity. In T. Ionin, S. Montrul, & R. Slabakova (Eds.). *The Routledge Handbook of Second Language Acquisition, Morphosyntax, and Semantics (*pp: 195-209). NY: Routledge.

Langsford, S., Perfors, A., Hendrickson, A., Kennedy, L., & Navarro, D. (2018). Quantifying sentence acceptability measures: Reliability, bias, and variability. *Glossa: A Journal of General Linguistics, 3*(1), 37. https://doi.org/10.5334/gjgl.396

Lardiere, D. (2000) Mapping features to forms in second language acquisition, in Archibald, J. (ed.) *Second Language Acquisition and Linguistic Theory*. Oxford: Blackwell, pp. 103–129.

Lardiere, D. (2007) *Ultimate Attainment in Second Language Acquisition: A Case Study*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lardiere, D. (2008). Feature assembly in second language acquisition. In J. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 106-140). Abingdon: Lawrence Erlbaum Associates.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research, 25* (2), 173–227. https://doi.org/10.1177/0267658308100283

Lazaridou-Chatzigoga, D., Katsos, N., & Stockall, L. (2019). Generalising about striking properties: Do glippets love to play with fire? *Frontiers in Psychology, 10*, 1971. https://doi.org/10.3389/fpsyg.2019.01971

Lazaridou-Chatzigoga, D. (2019). Genericity. In C. Cummins & N. Katsos (Eds.), *The Oxford Handbook of Experimental Semantics and Pragmatics*. Oxford University Press.

Lenth, R. (2023) Emmeans: estimated marginal means, aka Least-Squares Means. *R package version 1.8.7, p. https://cran.r-project.org/package=emmeans*. Available at: https://cran.r-project.org/package=emmeans

Linck, J. and Cunnings, I. (2015). The Utility and Application of Mixed-Effects Models in Second Language Research. *Second Language Research, 65*(S1), pp. 185–207. doi: 10.1111/lang.12117.

Loewen, S. (2015). *Introduction to Instructed Second Language Acquisition*. New York: Routledge.

Loewen, S., & Sato, M. (Eds.). (2017). *The Routledge Handbook of Instructed Second Language Acquisition.* Routledge. https://doi.org/10.4324/9781315676968

Long, M. (1988). Focus on Form: A design feature in language teaching methodology. Paper presented at the National Foreign Language Centre European Cultural Foundation Conference on empirical research on Second Language Learning in institutional settings, Bellagio, Italy.

Long, M. (1990). Focus on Form as a design feature in language teaching methodology. In K. De Bot, R. Ginsberg & C. Kramsch (Eds.). Foreign Language Research in Cross-Cultural Perspective (pp. 39-52). Amsterdam: John Benjamin.

Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511605789

Mackey, A., & Gass, S.M. (2021). *Second Language Research: Methodology and Design* (3rd ed.). Routledge. https://doi.org/10.4324/9781003188414.

Mari, A. (2015). Overt and covert modality in generic sentences. *Cahiers Chronos, 27*, 265-288. https://hal.science/ijn_02161876

Mari, A., Beyssade, C. and Del Prete, F. (2013). *Genericity.* Oxford: OUP

Marsden, H. (2018). Generative second language acquisition and language teaching: Advancing the dialogue. In J. Cho, M. Iverson, T. Judy, T. Leal, & E. Shimanskaya (Eds.), *Meaning and structure in second language acquisition: In honor of Roumyana Slabakova* (pp. 283-308). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.55

Marsden, H., & Slabakova, R. (2018). Grammatical meaning and the second language classroom: Introduction. *Language Teaching Research, 23*(2), 147-157.

Master, P. (1990). Teaching the English article system. *Theory into Practice, 29* (4), 245-251.

Master, P. (1994). The effect of systematic instruction on learning the English article system. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 229-252). Cambridge University Press.

Master, P. (2002). Information structure and English article pedagogy. *System, 30* (3), 331-348.

Master, P.(1997). The English article system: acquisition. function, and pedagogy. *SYSTEM*, 25(2),. 215 –232.

Mitchell, R., Myles, F., & Marsden, E. (2019). *Second Language Learning Theories: Fourth Edition* (4th ed.). NY: Routledge. https://doi.org/10.4324/9781315617046

Norris, J., & Ortega, L. (2003). Defining and measuring SLA. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 717-761). Oxford: Blackwell Publishing Ltd.

Norris, . & Ortega, L. (2000), Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-analysis. *Language Learning*, 50, 417-528. https://doi.org/10.1111/0023-8333.00136

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. John Benjamins Publishing Company. https://doi.org/10.1075/sibil.40

Pérez-Leroux, A. (2016). The expression of genericity in child language. In J. Lidz, W. Snyder, & J. Pater (eds.), *Oxford Handbook of Developmental Linguistics*, (pp. 565-586). Oxford, Oxford University Press.

R Core Team (2023). A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. R Foundation for Statistical Computing. Available at: http://www.r-project.org

Ramchand, G., & Svenonius, P. (2008). Mapping a parochial lexicon onto a universal semantics. *In* Biberauer (ed.) *The limits of syntactic variation.* Amsterdam: John Benjamins, 219-245.

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781009024532

Roberts, I.(2016). Introduction. In I. Roberts (Ed.),*The Oxford Handbook of Universal Grammar*(pp.1–35). Oxford: Oxford University Press

Roberts, I. (2019). *Parameter hierarchies and universal grammar*. Oxford University Press.

Rogers, J., & Révész, A. (2020). Experimental and quasi-experimental designs. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 133-143). Routledge. https://doi.org/10.4324/9780367824471-12

Rothman, J. (2008). Aspect selection in adult L2 Spanish and the competing systems hypothesis when pedagogical and linguistic rules conflict. *Languages in Contrast*, 8(1), 74-106.

Rothman, J., & Slabakova, R. (2018). The generative approach to SLA and its place in modern second language studies. *Studies in Second Language Acquisition*, *40*, 417–442. https://doi.org/10.1017/S0272263117000134

Sabir, M. (2015). *Explicit instruction and translation: A generative view of the acquisition of English articles*. Ph.D thesis, University of Leeds.

Lopez, E., & Sabir, M. (2019). Article Pedagogy: Encouraging Links between Linguistic Theory and Teaching Practice. *RELC Journal, 50*(1), 188-201. https://doi.org/10.1177/0033688217730143

Sarko, G. (2009) L2 English article production by Arab and French speakers. In M. d. P. García Mayo and R. Hawkins (eds), *Second Language Acquisition of Articles* 37–66. Amsterdam: John Benjamins.

Schwartz, B. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition*, 15, 147-163.

Schwartz, B., & Sprouse, R. (1994). Word order and Nominative Case in nonnative language acquisition: A longitudinal study of (L1 Turkish) German Interlanguage. In T. Hoekstra & B. Schwartz (Eds.), *Language acquisition studies in generative grammar* (pp. 317–368). John Benjamins.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209- 231.

Sharwood Smith, M. (1981). Consciousness-raising and the second language learner. *Applied Linguistics, 2*(2), 159-168. https://doi.org/10.1093/applin/II.2.159

Sharwood Smith, M. (1991). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition, 15*(2), 165-179. https://doi.org/10.1017/S0272263100010743

Sinclair, J. (1991). *Corpus, concordance and collocation*. Oxford: Oxford University Press.

Slabakova, R. (2009). Features or parameters: which one makes second language acquisition easier, and more interesting to study? *Second Language Research*, 25(2), 313-324.

Slabakova, R. (2013). What is easy and what is hard to acquire in a second language. *Contemporary approaches to second language acquisition*, 9, 5.

Slabakova, R. (2014). The bottleneck of second language acquisition. *Foreign Language Teaching and Research*, 46(4), 1–29.

Slabakova, R. (2016) *Second Language Acquisition*. Oxford: Oxford University Press.

Slabakova, R. (2019). The Bottleneck Hypothesis updated. In *Three streams of generative language acquisition research* (pp. 319–345). John Benjamins Publishing Company. https://doi.org/10.1075/lald.63.16sla

Slabakova, R., & García Mayo, M. P. (2015). Testing the Current Models of Third Language Acquisition. In *Cross-Linguistic Influence: From Empirical Evidence to Classroom Practice* (pp. 101-120). Springer.

Slabakova, R., Leal, T., Dudley, A., & Stack, M. (2020). *Generative Second Language Acquisition (Elements in Second Language Acquisition)*. Cambridge: Cambridge University Press. doi:10.1017/9781108762380a

Snape, N. (2008). Resetting the Nominal Mapping Parameter: definite article use and the count – mass distinction in L2 English. *Bilingualism: Language and Cognition, 11*(1): 63-79.

Snape, N. (2013). Japanese and Spanish adult learners of English: L2 acquisition of generic reference. *Journal of the Japanese Society for Language Sciences, 12*, 70–94.

Snape, N. (2018). Definite generic vs. definite unique in L2 acquisition. *Journal of the European Second Language Association*, *2*(1), 83–95. DOI: http://doi.org/10.22599/jesla.46

Snape, N. and & Yusa, N. (2013) Explicit article instruction in definiteness, specificity, genericity and perception. In M. Whong, K. –H. Gil and H. Marsden (eds) *Universal Grammar and the Second Language Classroom*, Volume 16, 161–83. London: Springer. https://doi.org/10.1007/978 –94 –007 –6362 –3_9

Snape, N., García –Mayo, M. D. P., & Gurel, A. (2013). L1 transfer in article selection for generic reference by Spanish, Turkish and Japanese L2 learners. *International Journal of Studies*, 13, 1 –28. http://revistas.um.es/ijes.

Spada, N, & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: a meta-analysis, *Language Learning*, 60(2), p. 263-308.

Suzuki, Y., DeKeyser, R., & Vafaee, P. (2023). Automatized explicit knowledge in second language acquisition: Insights from individual differences in cognitive aptitudes. *Language Learning, 73*(2), 345-367. https://doi.org/10.1111/lang.12624

Umeda, M., Snape, N., Yusa, N., & Wiltshier, J. (2019). The long-term effect of explicit instruction on learners' knowledge on English articles. *Language Teaching Research*, *23*(2), 179-199. https://doi.org/10.1177/1362168817739648

VanPatten, B. (2002). Processing instruction: an update. *Language Learning*, 52, 755-803.

VanPatten, B. (2009). Processing instruction: An update. *Language Teaching, 42*(3), 327-336. https://doi.org/10.1017/S0261444808005569

VanPatten, B. (2020). Input Processing in Adult L2 Acquisition. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (3rd ed., pp. 21-42). Taylor & Francis Group.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S (4th ed.)*. Springer. ISBN 0-387-95457-0

White, L. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research, 7*(2), 133-161. https://doi.org/10.1177/026765839100700204

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815065

White, L. (2011). *Second language acquisition and universal grammar*. Cambridge University Press.

White, L. (2011). Second language acquisition at the interface. *Lingua, 121*(4), 577–590. https://doi.org/10.1016/j.lingua.2010.05.006

White, L. (2020). Linguistic theory, universal grammar, and second language acquisition. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (3rd ed., pp. 21-42). Taylor & Francis Group. https://doi.org/10.4324/9780429503986-2

Whong, M., Gil, K.-H., & Marsden, H. (2014). Beyond paradigm: The 'what' and the 'how' of classroom research. *Second Language Research, 30*(4), 551-568. https://doi.org/10.1177/0267658314526749

Whong, M., Marsden, H., & Gil, K.-H. (2013). How we can learn from acquisition: The acquisition-learning debate revisited. In *Proceedings of the 12th Generative Approaches to Second Language Acquisition Conference* (pp. 203-210). Cascadilla Press. https://eprints.whiterose.ac.uk/77351/

Willcott, P. (1978). Problems of definiteness in the written English of Arabic speakers. *ELT Journal, 33*(1), 67–73. https://doi.org/10.1093/elt/XXXIII.1.67