The effects of Processing Instruction on the acquisition and processing of grammatical gender in German

Nick Henry, *The University of Texas at Austin*

Pre-Print to Henry, N. (2022). The effects of Processing Instruction on the acquisition and processing of grammatical gender in German. Language Teaching Research. Advanced Online Publication. https://doi.org/10.1177/13621688221096368

Abstract

This study investigates the effects of Processing Instruction (PI) on the acquisition of grammatical gender and gender-marked pronouns in German. PI was compared to a traditional, vocabulary-oriented approach using color cues (TI) and a categorization and memorization task (CM). The results of an immediate posttest showed that the PI group outperformed both TI and CM with respect to gender assignment on both a gender selection task and a writing task. The PI groups also scored higher and responded faster than the TI and CM groups on a comprehension task that required accurate processing of gender-marked pronouns. However, differences between the three groups were not sustained on delayed posttests. These results extend the findings of previous research on PI (e.g., Benati, 2004) by showing that PI can be applied to target forms that are low in communicative value and must be learned item-by-item like grammatical Gender in German. Results also lend support to psycholinguistic research that suggests that L2 learners have difficulty acquiring grammatical gender because they do not process nouns together with gender information (Arnon & Ramscar, 2012).

Introduction

Grammatical gender assignment underlies many functions in languages with robust gender systems. In languages such as German, the ability to assign grammatical gender correctly is necessary for pronominal reference, case assignment, relative-clause attachment, and adjective-noun agreement (Ritterbusch, LaFond, & Agustin, 2008). Although research has shown that late L2 learners can learn to assign gender accurately (Gillon Dowens, Guo, Guo, Barber, & Carreiras, 2011; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010), they have persistent difficulty doing so, especially when gender is not present in their first language (Franceschina, 2001; Hawkins & Chan, 1997; Hopp, 2013). Consequently, L2 learners struggle to process gender information (Hopp, 2016), and they often misunderstand sentences when interpretation rests on accurate gender assignment. Evidence from psycholinguistics suggests that this difficulty stems from a tendency to attend to nouns and genders separately (Arnon & Ramscar, 2012), and to focus on content words while comprehending sentences (VanPatten, 2015b), leaving gender information unprocessed. Thus, learners may benefit from instructional approaches that highlight the psycholinguistic function of gender and promote the processing of gender information.

In order to explore the potential benefits of such an approach, the present study compares Processing Instruction (see VanPatten, 2015a) with two traditional approaches to gender teaching. The study thus evaluates whether psycholinguistically-motivated approaches help learners develop more robust knowledge of gender information and a greater ability to use gender information in multiple contexts. In doing so, the present study further considers the relationship between gender assignment, sentence comprehension, and online processing.

Background and Motivation

Grammatical Gender

While the function of grammatical gender is not immediately clear—either to linguists or to language users—its primary function is likely to aid pronominal reference and guide the interpretation of otherwise ambiguous utterances (Trudgill, 1999). Research also shows that gender serves an important psycholinguistic function, constraining the number of possible lexical items in an utterance and therefore easing processing load and speeding lexical access (Lew-Williams & Fernald, 2007). Indeed, several studies have shown that, along with phonological, semantic, and morphosyntactic information (see Huettig, 2015), L1 speakers use gender to predict upcoming nouns in the input (Dahan, Swingley, Tanenhaus, & Magnuson, 2000; Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen, 2013; Hopp, 2016). The ability to use grammatical gender for prediction develops early on in L1 acquisition. For example, Lew-Williams and Fernald (2007) showed that children 34-42 months old were faster to look towards a target picture when the accompanying distractor was a different gender, than if it had the same gender. Together, these findings show that the parser develops sensitivity to gender as an informative cue for sentence processing early on and continues to use this information throughout adulthood.

In contrast to L1 speakers, L2 learners vary widely in their ability to use gender during online processing. While some research has shown that L2 learners have persistent difficulties with gender in online processing (Grüter, Lew-williams, & Fernald, 2012), others have found that performance varies with the context, learning condition (Lew-Williams & Fernald, 2009), or proficiency (Dussias et al., 2013; Gillon Dowens et al., 2011). There are varying accounts as to why L2 and L1 speakers differ in their processing of grammatical gender, for example, attributing variable performance to maturational constraints (Hawkins & Chan, 1997), limits on

cognitive resources (Keating, 2009), typological similarity to the L1 (Dussias et al., 2013; Foucart & Frenck-Mestre, 2011), and proficiency (Gillon Dowens, Vergara, Barber, & Carreiras, 2010; Hopp, 2013). Other research has found that the variability of gender assignment accuracy and the speed of lexical access are critical factors in the ability to process gender information online. For example, Hopp (2013) only found predictive use of gender information among learners when they had both consistently target-like gender assignment and rapid lexical access. In another study, Lemhöfer, Schriefers, & Indefrey (2014) found that German learners of L2 Dutch were only sensitive to gender violations when stimuli were coded according to the learners' "subjective" gender, rather than objective correctness. Interestingly, Hopp (2016) demonstrated not only that instruction on gender assignment can lead L2 learners to process gender predictively, but also that L2 difficulties with gender agreement can be emulated in L1 processing by introducing gender assignment errors. Hopp thus demonstrates a close link between gender assignment and the tendency to process gender information online.

Lexically-based accounts of gender learning (e.g., Grüter et al., 2012; Arnon & Ramscar, 2012) suggest that a lack of lexical knowledge (e.g., gender assignment), rather than syntactic knowledge (e.g., gender agreement), is the locus of processing difficulties in the L2. These accounts further suggest that L2 learners employ a different type of learning than L1 learners, leading to differences in the representation of gender knowledge. In one such account, Grüter et al. (2012) build on connectionist models of gender (see Schriefers & Jescheniak, 1999), stating that gender is stored as a property of each noun, but that "the associations between nouns and gender class information are unlikely to attain the same strength in L2 as in L1 lexicons" (p. 210). Specifically, they argue that L2 learners form weaker associations than child L1 learners because they do not rely on statistical co-occurrence to learn noun-gender pairings. A similar

proposal from Arnon & Ramscar (2012) suggests that L1 children learn noun-gender pairings by first learning them as lexical chunks, only later learning to discriminate between nouns and genders. L2 learners, by contrast, treat nouns and genders as separate items very early on, perhaps because they are better able to identify which strings carry the most semantic meaning, use knowledge from their L1, and learn individual words through text.

A critical component of Grüter et al. (2012) and Arnon and Ramscar's (2012) hypotheses is that, early in acquisition, children process nouns alongside gender information, while L2 learners do not. This claim is consistent with other psycholinguistic models of L2 acquisition that posit that L2 learners privilege lexical-semantic information over (morpho)syntactic information (see e.g., Clahsen & Felser, 2006). One such model, VanPatten's Input Processing model (e.g., VanPatten, 2015b), consists of a number of principles that describe how L2 learners initially create form-meaning connections. The Primacy of Meaning Principle states that learners are primarily driven to get meaning from the input, while the Availability of Resources Principle states that have limited resources with which to process input. Further, the model states that learners (1) process content words before other elements of the input (the Primacy of Content Words Principle), (2) process meaningful forms before non-meaningful forms (The Meaning-Before-Nonmeaning Principle), and (3) are more likely to process non-redundant forms than redundant forms (The Preference for Nonredundancy Principle). These principles predict that gender information is not likely to be processed unless it carries significant semantic weight and is non-redundant. One might therefore expect that learners acquire forms related to semantic gender before they acquire forms marked for grammatical gender. Indeed, this has been substantiated by empirical studies on gender learning (Lew-Williams & Fernald, 2009).

Note that VanPatten's approach is differentiated from connectionist and usage-based accounts, for example, in how they consider linguistic knowledge (i.e., domain-specific vs. domain-general) and the roles ascribed to frequency and transfer. Despite these differences, this prediction made by VanPatten's approach—that L2 learners would be less likely to process grammatical gender—is also expected under these models, specifically those that emphasize the role of salience (Cintrón-Valentín & Ellis, 2016) and suggest that form-meaning mappings are more difficult to process if they are not one-to-one (i.e., if forms do not map onto single meanings and vice versa; Chan, Lieven, & Tomasello, 2009; Grünloh, Lieven, & Tomasello, 2011). That is, given that gender markings (e.g., definite articles in German) are often low-salience forms that encode other semantic functions (e.g., natural gender or finiteness) these approaches would also predict that L2 learners are less likely to process these forms for grammatical gender information.

Taken together, psycholinguistic research on gender processing suggests that learners need to process grammatical gender information for meaning (i.e., link lexical items to gender information) in order to acquire the gender system and use it for interpretation. In other words, the processing of gender information is not simply the *result* of having robustly represented gender information in the mental lexicon; rather, it is the *source* of that knowledge. Consequently, instructional approaches to gender learning could benefit from comprehension and processing-based approaches that (a) compel learners to process nouns and genders together, and (b) make the functional aspects of gender clear and relevant to learners. I return to this point later, after a brief explanation of grammatical gender in German—the focus of this article—and a discussion of approaches to teaching gender.

Grammatical Gender in German

In German, nouns are marked for one of three grammatical genders: masculine, feminine, or neuter, which correspond to the nominative definite articles *der, die,* and *das.* There are phonological, morphological, and semantic tendencies that guide gender assignment. For example, words ending in /ə/ tend to be marked for feminine, words ending in the suffix –chen are always neuter, and words referring to male/female persons tend to follow semantic gender. However, these tendencies often interact in complex ways, and there are numerous exceptions to individual rules: *Junge* ('boy') ends in /ə/ but is grammatically (and semantically) masculine, and *Mädchen* ('girl') is semantically feminine but is grammatically neuter. Thus, despite the regularities that have been detailed in the linguistic literature (see e.g., Köpke & Zubin, 1984), the system is largely phonologically and semantically opaque (Hopp, 2013), especially for a beginning L2 learner, who does not have enough vocabulary to detect any regularities (see Arzt & Kost, 2016). The acquisition of the gender system therefore poses a distinct challenge to L2 learners in that, to them, it does not appear to be a 'system' at all. Rather, it appears to them that noun-gender pairings are mostly random.

Grammatical gender assignment is related to a host of other grammatical functions, including pronominal reference. In German, the gender of an intended referent is indicated by gender-marked pronouns: masculine, feminine, and neuter nouns are marked by the nominative pronouns *er, sie,* and *es,* respectively (note the phonological correspondence between the pronouns and the definite articles *der, die,* and *das*). As illustrated in (1), this gender marking is important for interpretation because it allows learners to disambiguate between multiple possible interpretations of an utterance.

(1) Auf dem Tisch ist eine Jacke ein Hemd und ein
On the table, there.is a$_{NOM.FEM}$ jacket$_{FEM}$, a$_{NOM.NEUT}$ shirt$_{NEUT}$, and a$_{NOM.MASC}$

Pulli. Er ist wirklich schick.
sweater$_{MASC}$. It$_{NOM.MASC}$ is really fashionable.

Because the pronominal *er* can refer only to the masculine item on the table (the sweater), the phrase *Er ist wirklich schick* ('It is really fashionable') is unambiguous.

Given that German learners often have difficulties with gender assignment (Hopp, 2013), they cannot readily interpret sentences that use pronominal forms in this way. Learners also struggle with pronoun use because they tend to privilege animacy over gender, particularly if their L1 lacks gender as a functional feature (Hawkins & Chan, 1997). Thus, L2 learners often overuse the neuter pronoun *es*, assigning it to any inanimate object.

Approaches to teaching grammatical gender and pronouns

Gender assignment. In the L2 pedagogy and acquisition literatures, researchers have investigated numerous approaches to teaching grammatical gender (Arzt & Kost, 2016; Chew, 1989; Kraiss, 2014; Santos, 2015). However, before describing these approaches, it should be noted that—after the concept of grammatical gender is introduced in the first week of instruction—gender assignment often receives little to no attention through direct, targeted interventions. Past this, some instructors make occasional references to regular tendencies, while some ignore these altogether. Others treat gender assignment as a part of vocabulary learning and may require learners to record vocabulary in a journal along with the word's gender. Those instructors that do address gender through targeted in-class activities generally follow

approaches that can be broadly characterized as rule-based, repetition-based, or mnemonic-based (see Arzt & Kost, 2016 and Bjornstad, 2014).

Rule-based approaches focus on explicit rule learning (Bjornstad, 2014; Kraiss, 2014; Schirrmeister, 2015) and generally adapt theoretical work by Köpke and Zubin (1984) and others for use in the classroom. In essence, these approaches attempt to make regularities of the gender system clear to learners. During instruction, learners either receive or construct lists of semantic, morphological, and phonetic rules that apply to given examples. Because of the number of rules involved, the list of rules is typically kept short at first and built up over time. Learners may also conduct targeted practice with these rules to solidify them (Bjornstad, 2014; Kraiss, 2014). Research on rule-based approaches have reported positive effects on vocabulary quizzes and multiple-choice tests, sometimes showing drastic improvement over control groups (e.g., Kraiss, 2014; Schirrmeister, 2015). However, rule-based approaches can only account for a portion of words in the German lexicon, and these approaches are not effective for nouns that do not conform to rules (Schirrmeister, 2015). Further, the rules involved are numerous, complicated, unclear, and likely cannot be used actively in a communicative context (Arzt & Kost, 2016; but see Presson, MacWhinney, & Tokowicz, 2012).

In repetition-based approaches, learners listen to and repeat words together with their gender multiple times in order to build gender-noun mappings directly (Chew, 1989). This approach has received considerably less attention from researchers, but it does follow logically from the oft-repeated suggestion that students should learn the gender with the noun, an idea that is supported by psycholinguistic studies (e.g., Arnon & Ramscar, 2012). It should be noted that gender-noun repetition has been used successfully to train gender and vocabulary in lab-based studies on language processing, (e.g., Henry, 2022), with some studies showing that these

approaches can promote online gender processing (e.g., Hopp, 2016; Schempp, 2017). However, the trainings used in these studies are quite different from the classroom instantiation of repetition-based approaches (see Chew, 1989).

Mnemonic-based approaches have received the most attention in the literature. Despite differences between these approaches, they all seek to help learners create gender-noun pairings indirectly through a gender-specific association. For example, several studies have paired words with an image of a man or woman (Arzt & Kost, 2016; Desrochers, Wieland, & Cote, 1991; Nyikos, 1987). Other studies have coded words using colors, typically presenting masculine nouns in blue, feminine nouns in red/pink, and neuter nouns in green (Arzt & Kost, 2016; Kohler, 2009; Nyikos, 1987). A similar approach employed by Santos (2015) paired vocabulary words with male, female, and genderless (robotic) voices. These different mnemonic devices can also be combined as done in several of the aforementioned studies (Nyikos, 1987; Santos, 2015). Research on these approaches has used fill-in-the-blank tests and vocabulary quizzes and shown that the use of mnemonics can increase accuracy of gender assignment over controls (Desrochers et al., 1991; Kohler, 2009; Nyikos, 1987), though observed effects are not always robust (Arzt & Kost, 2016). The totality of the research does not suggest that any one of the mnemonic approaches is more successful than the others, although individual studies have found differences between approaches, for example, Santos (2015) who found participants made the largest gains when mnemonics included images.

Pronoun Assignment. The acquisition of German nominative pronouns has received very little attention either from teachers or from researchers. This is likely because pronouns correspond directly to a word's grammatical gender, and thus the grammatical rule associated with pronoun assignment is straightforward. It is therefore often assumed that, once learners are

able to assign gender correctly, pronoun assignment will follow. However, as previously noted, errors in pronoun assignment typically persist past the beginning stages of learning, and—as noted previously—learners often overuse the neuter *es* to refer to any inanimate object and misinterpret statements using pronominal reference for inanimate nouns. Thus, a principled approach to teaching pronoun assignment alongside gender assignment could benefit learners.

Processing Instruction

Although the approaches to gender and pronoun acquisition discussed in the previous section are quite different in many respects, they all emphasize decontextualized learning of noun-gender pairings and focus on the learner's ability to supply gender information in isolated contexts. Consequently, they do not provide learners with opportunities to process gender information during sentence comprehension, and they do not address the function of grammatical gender: to constrain interpretations and predict upcoming nouns during sentence comprehension (Dahan et al., 2000; Hopp, 2016; Lew-Williams & Fernald, 2007; Scherag, Demuth, Rösler, Neville, & Röder, 2004). As argued previously, comprehension-based approaches may help learners acquire a more robust representation of grammatical gender and make connections between nouns and pronouns so that they are better able to interpret pronominal forms when they appear in context, and particularly when they refer to inanimate objects.

One comprehension-based approach that has received considerable attention in L2 research is Processing Instruction (PI) (VanPatten & Cadierno, 1993). PI is based on VanPatten's Input Processing model (see VanPatten, 2015b). The core tenets of the approach are that (a) instructional interventions should promote meaningful or functional aspects of targeted grammatical forms by making them essential to the task, and (b) the task should push learners to

overcome default (or nonoptimal) processing strategies. For example, a PI activity focused on the past-tense *–ed* morpheme in English would present learners with sentences like *They listened to music in their rooms*, and then ask learners to determine whether the event took place last week or is taking place right now (see also Benati & Angelovska, 2015). Because the only cue to the past tense is the *–ed* morpheme*,* learners must process the cue for past tense in order to perform the task accurately. Thus, learners are pushed to process morphology that they tend to leave unprocessed, as described in the Lexical Preference Principle (VanPatten, 2015b).

Research on PI has primarily investigated how it influences the acquisition of morphological and syntactic forms. The seminal study on PI, VanPatten and Cadierno (1993), investigated the acquisition of accusative clitic pronouns in Spanish and found that PI increased learners' comprehension of the target form. Moreover, participants' production of the target forms increased in accuracy, despite the fact that learners did not produce the form at all during training. VanPatten and Cadierno concluded that PI was effective because it changed how learners processed input and in turn provided intake for the developing system. Since this initial study, PI research has widened to include a variety of forms and languages, including case in German (AUTHOR et al., XXXX; AUTHOR et al., XXXX), tense in Spanish, (Cadierno, 1995; COAUTHOR & AUTHOR, XXXX, XXXX), past tense in English (Benati, 2005; Benati & Angelovska, 2015; Marsden & Chen, 2011), future tense in Italian (Benati, 2001), copula selection in Spanish (Cheng, 2002), and mood in Spanish (Collentine, 1998; Fernández, 2008). These studies, along with a meta-analysis by Shintani (2014), have generally confirmed the pattern observed by VanPatten & Cadierno (1993) and show especially strong effects for comprehension tasks, indicating that targeted processing of morphosyntactic forms can promote the creation of form-meaning connections. Particularly relevant to the present study was Benati

(2004), which targeted gender-marked adjectives (i.e., noun-adjective agreement) in Italian. Benati found that PI did indeed promote comprehension and production of noun-gender agreement, and that explicit information was not a required component of training.

Despite the evidence that PI promotes morphological learning, including grammatical gender agreement in Italian, it is unclear whether the effects of PI would extend to forms like grammatical gender in German. In some ways, grammatical gender in German is similar to the forms that PI research has investigated in the past. Like those forms, it is not readily processed because learners tend to focus on content words or context instead of functional morphology to make meaning (Jackson, Fowler, Gavin, & Henry, 2018; see VanPatten, 2015b). However, in some ways, grammatical gender in German is quite different from other morphological forms. First, as Benati (2004) noted, unlike tense, case, or mood, grammatical gender is semantically opaque and in many contexts is non-meaningful. That is, in many contexts, gender markings on articles, possessives, or adjectives do not contribute to the meaning of the sentence because they simply agree with nouns in the sentence. Secondly, morphemes that mark gender in German carry other semantic information, such as definiteness. Similarly, learners often assume that gender-marking pronouns carry animacy information, even though they do not. As a result, learners may map form-meaning connections incorrectly. Finally—unlike gender in languages like Italian or Spanish—because the rules governing gender assignment are complex and do not cover most nouns in the language, the acquisition of grammatical gender must be, at least partly, lexical in nature.

To my knowledge, there has only been one study that investigates if PI effectively promotes gender learning and pronoun use among L2 learners of German. In that study, Johnson Fowler and Jackson (in preparation, reported in Jackson et al., 2018) compared the effects of PI

to traditional output-oriented instruction (TI). In this classroom study, the PI group saw three items on a power-point screen, accompanied by either the half-sentence *Hier ist der* 'Here is the$_{MASC}$' or the phrase *Er ist schön* 'It$_{MASC}$ is nice/beautiful.' Students picked the item on the screen that corresponded to the gender in the prompt (in this the masculine item). The TI group focused on identifying the vocabulary items with their genders and producing gender-matched pronouns. The results showed that PI and TI both resulted in similar gains in gender-assignment on an immediate and delayed posttest. While this study points to the possibility that PI can be used as an alternative to more traditional approaches, it did not test the learners' ability to use gender information in context, either in comprehension or in production.

The Present Study

The present study explores whether tasks that push learners to use gender information and comprehend utterances lead to improvement in offline gender assignment and the use of gender-information in context. To address this question, the present study focuses on a comparison of three types of instruction—Processing Instruction (PI), Traditional Instruction (TI), and Categorization/Memorization (CM)—and investigates how these influence the acquisition of gender-marked definite articles and gender-marked pronouns for vocabulary associated with clothing. As previous studies on the acquisition of gender have not investigated either comprehension or pronoun use, the present study utilizes a combination of assessments that include offline gender and pronoun assignment, translation, sentence-level production, and sentence comprehension. In addition, the present study analyzes reaction times during sentence comprehension in order to investigate whether PI promotes online processing of gender information and thus helps learners avoid nonoptimal processing strategies as described by the

Primacy of Content Words, Preference for Nonredundancy, and Preference for Meaning-before-Nonmeaning Principles . This broad approach to testing will provide a more comprehensive picture of how PI influences different skills related to gender use.

Note that, while vocabulary acquisition is not the focus of the present study, the pedagogical use of PI, TI, and CM is also influenced by the degree to which learners can learn word meaning alongside grammatical gender. Thus, the present study also investigates vocabulary learning as an outcome of these trainings.The research question for the study are:

- RQ1: To what extent does PI lead to accurate assignment of gender-marked definite articles, when compared with TI and CM?
- RQ2a: To what extent does PI lead to accurate assignment of gender-marked pronouns, when compared with TI and CM?
- RQ2b: To what extent does PI lead to more accurate and/or faster comprehension of gender-marked pronouns, when compared with TI or CM?
- RQ2c: To what extent does PI lead to more accurate production of pronominal forms in sentence context, when compared with TI and CM?
- RQ3: To what extent do PI, TI, and CM lead to increased vocabulary knowledge?

Given the links between processing and gender assignment described in the review of literature, it was hypothesized that PI would have larger and more durable effects on learners' ability to assign gender accurately on definite articles (RQ1) and pronouns (RQ2a). It was further hypothesized that PI would lead to greater gains for the comprehension (RQ2b) and production (RQ2c) of pronominal forms. Finally, it was hypothesized that all three instructional methods would promote vocabulary knowledge (RQ3).

Methodology

Participants

The participants in the present study were drawn from beginning (i.e., first-semester) German classrooms at a large university in the U.S. Participants received extra credit for their participation in this study. The initial pool consisted of 74 participants, all of whom were native speakers of English. Six participants were removed because they did not complete the study or because computer error resulted in the loss of their data. The final participant pool for data analyses was 68 (32 female, 34 male, 2 non-binary / not specified). These participants were randomly assigned to one of three treatment groups, PI (n = 23), TI (n = 24), and CM (n = 21). These groups completed a language background questionnaire to screen for the above inclusion criteria and to assess their experience and self-rated proficiency in German as seen in Table 1. One-way ANOVAs showed that the three groups were similar with respect to their age, time spent in a German speaking country[1], and self-rated reading, spelling, writing, speaking, and listening ability (all $p > .1$).

Table 1. Means (SDs) for background questionnaire

Variable (Range of Possible Scores)	PI	TI	CM
Age	19.35 (1.85)	21.35 (8.96)	19.9 (1.67)
Time in a German speaking country (months)	4.17 (20.02)	0.13 (0.61)	0.1 (0.44)
Self-Rating: Reading (0-10)	3.04 (1.19)	3.42 (1.21)	3.52 (1.91)
Self-Rating: Spelling (0-10)	3.04 (1.43)	3.29 (1.6)	3.48 (1.97)
Self-Rating: Writing (0-10)	3.17 (1.47)	3.21 (1.22)	3.29 (1.95)

Self-Rating: Speaking (0-10)	2.57 (1.38)	2.83 (1.37)	2.62 (1.75)
Self-Rating: Listening (0-10)	2.96 (1.4)	3.5 (1.38)	2.81 (1.6)

Instructional Treatments

Target Vocabulary Items. All of the materials used in this study, including those used for the instructional treatments and the assessments are found in the IRIS database (Marsden et al., 2016). The target vocabulary items for each of the instructional treatments were 24 words related to clothing and accessories (Table 2). These items were chosen because they are commonly taught in beginning German classes. Importantly, however, the participants' language classes do not introduce these words until the second semester; therefore, it was unlikely that the participants would learn or practice these words in their language classes. Furthermore, because all of the words belonged to one category, it was possible to create trainings with a communicative frame for training—in this case, shopping in a popular clothing store—enhancing the ecological validity of the treatments under investigation. As seen in Table 2, 8 of the target vocabulary items were masculine, 8 were feminine, and 8 were neuter. Among the items were six cognates (in italics), evenly divided among the three genders.

Table 2. Target vocabulary items for training and testing

Masculine		Feminine		Neuter	
der Anzug	‘suit'	die Halskette	‘necklace’	das Hemd	‘shirt’
der Stiefel	‘boot’	die Hose	‘pants’	das Sakko	‘sportscoat’
der Gürtel	‘belt’	die Sonnenbrille	‘sunglasses’	das Kleid	‘dress’
der Pulli	‘sweater’	die Handtasche	‘purse’	das Portemonnaie	‘wallet’
der Regenmantel	‘raincoat’	die Krawatte	‘tie’	das Armband	‘bracelet’
der Rock	‘skirt’	die Mütze	‘cap’	das Trikot	‘jersey’
der Schuh	‘shoe’	*die Socke*	‘sock’	*das T-shirt*	‘T-shirt’
der Hut	‘hat’	*die Jacke*	‘jacket’	*das Sweatshirt*	‘sweatshirt’

Categorization / Memorization. Participants in the CM treatment group were given a paper hnadout listing the 24 target vocabulary items along with their English translations and black and white clipart pictures. Participants were told to study the words briefly, and rewrite them by hand, categorizing them by gender. They were provided highlighters and told that they could highlight the words in different colors if they chose. This procedure was designed to mimic a strategy that is often observed among lower-division German learners and sometimes required by German instructors. Most participants completed this training task in about 10-15 minutes.

TI Treatment. The traditional instruction (TI) training was conducted on the computer using the computer program E-Prime. The training consisted of two training blocks: vocabulary training and pronoun training (see Figure 1). In the vocabulary training, participants saw one of the target vocabulary words written (with its gender-marked definite article) along with three pictures. They selected the picture that best matched the word and received simple one-word feedback (‘Correct!’ or ‘Incorrect’). They then proceeded to the next item. The full training

block consisted of the 24 target vocabulary items, each of which was the correct answer one time. Each item's picture also appeared as the incorrect answer twice. Trials that were answered incorrectly were repeated until the participant had answered each question correctly. The words presented to participants appeared in different colors according to their gender (red = feminine, blue = masculine, green = neuter).

In the pronoun training, participants saw a target word along with the gender-marked pronouns *er*, *sie*, and *es*. They selected the pronoun corresponding to the word and received feedback. As in section one, each of the 24 target vocabulary items appeared once in the training, and incorrectly answered trials were repeated.

The TI treatment was administered during both the first and the second research sessions. The training in each session took about 10 minutes to complete. The only difference between these training sessions was the pairing of target and distractor items in section one. In each training session, and in each training block, presentation of the 24 target items was randomized.

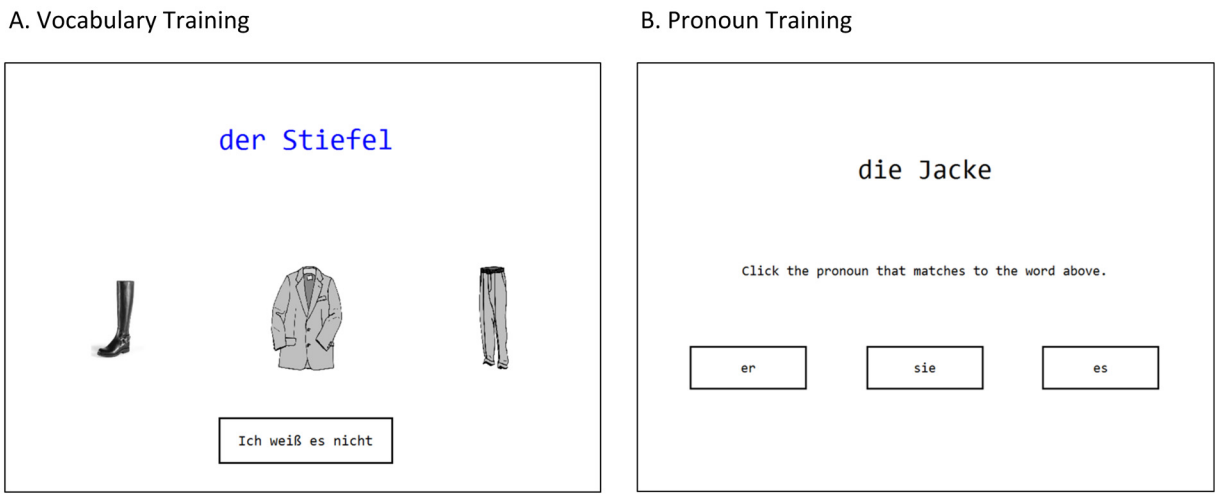

Figure 1. Traditional Instruction training blocks.

The TI treatment was designed to mimic activities that are often seen in college level textbooks when introducing vocabulary words and teaching about pronouns. At no point in the training was it necessary to process the gender together with the meaning of the target item in order to select the correct answer. In the vocabulary section, although participants were told that words would be color coded, and that they should pay attention to gender, they only needed to process the meaning of the word to complete the task. Similarly, in the pronoun training block, participants did not need to process the meaning of the vocabulary word; instead, they could have completed the task by only focusing on the gender-marked definite article presented alongside the word.

PI Treatment. Like the TI treatment, the PI treatment was conducted on the computer using the computer program E-Prime and consisted of vocabulary and pronoun training blocks (Figure 2). In the vocabulary training section, participants saw three target vocabulary items—one for each gender—along with their pictures (Figure 2). At the same time, they heard a half-sentence in German, *Hier ist der / die / das* ('Here is the$_{MASC}$ / the$_{FEM}$ / the$_{NEUT}$'). They selected the word that completed the sentence, received simple one-word feedback ('Correct!' or 'Incorrect'), and then proceeded to the next item. As in the TI treatment, the full training block consisted of the 24 trials, in which each of the 24 items appeared as the correct answer one time and the incorrect answer twice. Trials answered incorrectly were repeated until they were answered correctly.

The pronoun training block was identical to the vocabulary training block, except that, instead of hearing a half-sentence, participants heard a full sentence using the gender-marked pronoun, i.e., *Er / sie / es ist schick* ('It$_{MASC}$ / it$_{FEM}$ / it$_{NEUT}$ is fashionable').

PI training was also administered during both the first and the second research sessions. Presentation of the 24 target items was randomized in each training session, and the only difference between them was the pairing of target and distractor items. The training in each session took about 10 minutes to complete. Note that, while this is a shorter in duration than many of the original studies on PI (e.g., VanPatten & Cadierno, 1993), the total length of training matches or exceeds previous PI studies using computerized training, either in the number of items (e.g., Wong & Ito, 2018), or in duration (Henry, 2022; Henry et al., 2009), or in both (e.g., VanPatten, et al., 2013).

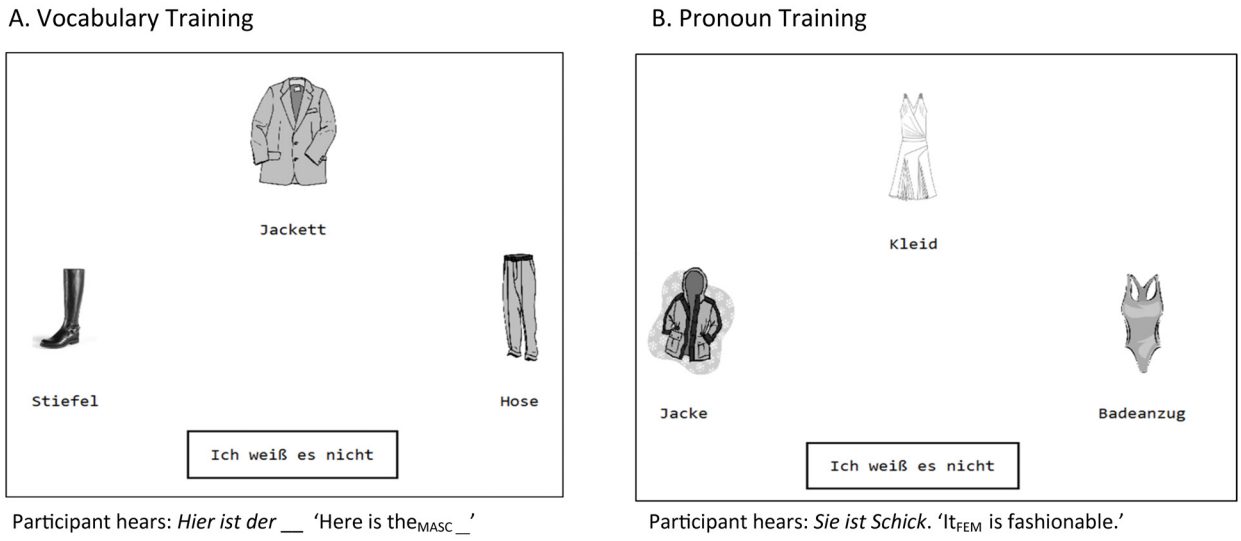

Figure 2. Processing Instruction training blocks.

These tasks may seem similar to the tasks in the TI treatment: they involve matching pictures to a prompt, they provide learners with simple corrective feedback, the number of trials in each treatment are the same, and they both require learners to achieve 100% accuracy. However, participants could not complete the tasks without using gender information. Thus, the tasks are functionally quite different in terms of what information learners must process. In turn,

the PI treatment pushes learners to overcome the tendency to leave gender information unprocessed. While the design of the PI treatment is also somewhat different from other tasks in the PI literature, the task adheres to the primary design features required by PI: it pushes learners to overcome processing difficulties related to the target form, and predicted by the input processing model's principles, in this case, the Primacy of Content Words, Preference for Meaning-before-Nonmeaning, and Preference for Nonredundancy Principles.

Assessment Measures

The participants were assessed prior to training, immediately after training, and four weeks after training. At each time point, they received one of four versions of the test, counterbalanced such that participants did not see the same version of the test more than once.

The test consisted of four primary tasks. (1) First, participants completed a gender selection and translation task. For each of the 24 target vocabulary items, participants selected the matching gender and provided a translation of the word. (2) Then they completed a production task. For this task (and each subsequent task) participants were tested on a subset of the target vocabulary words due to time limitations and to control for learner fatigue. The production task consisted of twelve of the target vocabulary items. Participants saw a word along with its picture and four adjectives. They then wrote two sentences describing the picture following a model sentence: *Hier ist der / die / das* _____. _____ *ist* _____. For example, participants may have written: *Hier ist die Jacke. Sie ist schön* ('Here is the$_{FEM}$ jacket. It$_{FEM}$ is pretty). (3) Next, participants completed a comprehension task, which was identical to the pronoun training block in the PI treatment (see Figure 2b), except that the participants received no feedback on their answers and were tested on twelve of the 24 target vocabulary items. (4)

Finally, participants completed a pronoun matching task that was identical to the pronoun block in the TI treatment (see Figure 1b). However, participants only saw six items from the target list. Six additional novel items were included to test whether participants had generalized the pronoun rule.

Procedure

Participants completed the experiment in three research sessions conducted by the author (or a research assistant) in a laboratory setting. In the first research session, participants first completed the language background questionnaire and the pretest. Each participant then saw a list of the target vocabulary items along with their English translations and the pictures that would be used in the PI and TI trainings. Using this list, the CM group completed their categorization task. The PI and TI groups studied this list for three minutes before completing their respective treatments.

In the second research session, the PI and TI groups completed their trainings for the second time. They then completed the posttest. The CM group completed only the posttest. In the third and final research session, each group completed the delayed posttest, followed by a short debriefing questionnaire that asked whether they liked the training, if it changed their approach to gender or vocabulary learning, and if they had studied the target words outside of the study.

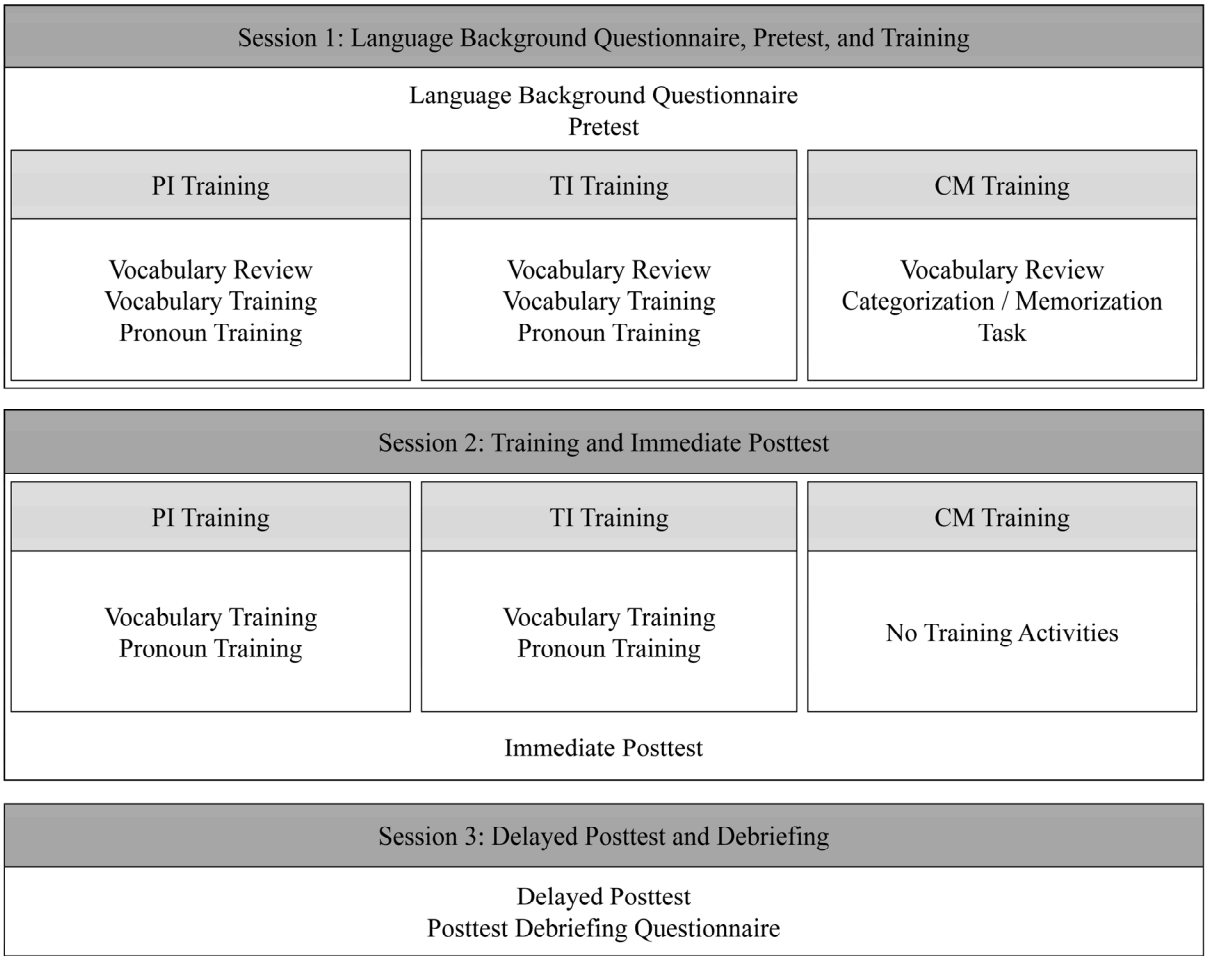

Figure 3. Outline of experimental procedure

Data Scoring

For the gender selection and translation task, participants' responses were scored separately[2]. Participants received one point if the gender selected matched the gender of the word, and no points if it did not. Participants received credit for a correct translation if they provided a translation that closely approximated the translation provided to them in the vocabulary list (e.g., *Regenmantel* 'raincoat', was scored as correct if participants answered 'raincoat' or 'rain jacket'). Spelling and plurality were not considered (e.g., *Socke* 'sock' was

scored as correct if participants answered ‘socks’). One point was given for correct answers, and no points were awarded for incorrect answers.

The production task was first scored with respect to the accuracy of the gender provided for the target noun. Participant responses were then scored according to whether the pronoun matched the gender that they had supplied. Consider the following response: *Hier ist der Jacke. Er ist nett* (‘Here is the$_{MASC}$ Jacket$_{FEM}$. It$_{MASC}$ is nice’). This sentence received a score of zero for gender accuracy, because the gender (masculine) does not match that of the noun (feminine). However, it received one point for pronoun accuracy, because the pronoun *er* (masculine) matched the gender that the participant had assigned to the noun.

The comprehension task was scored for accuracy, with one point awarded for a correct answer. Reaction times (RTs) were also collected for each trial. RTs represented the time participants needed to respond after presentation of the audio stimulus. Following Angelovska & Roehm (2020), data were trimmed such that, for each subject, RTs more than 2.5 standard deviations (SDs) above the mean were replaced with the value 2.5 SD above the mean RT.[3] The pronoun matching task was scored for accuracy, with one point awarded for a correct answer.

Data Analysis

The data were not normally distributed and thus it was inappropriate to analyze the data using traditional parametric tests. The analyses were therefore performed using linear mixed-effects models, which allow for repeat testing of participants and are robust to violations of normality, homoscedasticity, and sphericity (see McManus & Marsden, 2019). The mixed-effects models were computed in R version 1.4.1103 (R Studio Team, 2021) using the *lmer*

function in the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). Separate models were created for analyses of each measure as described in *Data Scoring*. Each model contained the fixed-factors Group, Time, and the interaction Group x Time, and the random factor Participant. The maximal model was the best fit model for all but one measure (noted in the results), as determined by AIC and BIC values. For the mixed effects models, visual inspection of the model data indicated that the models all met the assumptions of linearity, homogeneity of variance, and normality. For the purposes of interpretation, simple effects were converted to main effects, using the *mixed* function in the *afex* package (Singmann, Bolker, Westfall, Aust, & Ben-Schachar, 2021). When a Group x Time interaction was obtained, pairwise comparison for both Group and Time were computed, and are presented in Tables. For ease of presentation, the narrative description of results focuses on Group comparisons, as these are more pertinent to the research questions and more difficult to discern from the graphical representation of the data.

In order to control for effects of cognates, a separate analysis was conducted in which all of the cognate items (italicized items in Table 2) were removed from the data. The full results of this analysis are presented in the online supplemental materials, and relevant findings are noted alongside the main analyses.

Results

The means and standard deviations for each of the seven measures are displayed in Table 3. Statistical results for each measure are detailed by task in the following sections.

Table 3. Means (Standard Deviations) for Experimental Tasks

	Gender Selection	Translation	Gender Production	Pronoun Production	Comprehension Accuracy	Reaction Time (ms)	Pronoun Matching
Pre							
CM	0.54 (0.18)	0.29 (0.12)	0.46 (0.22)	0.64 (0.37)	0.40 (0.23)	5658 (2229)	0.97 (0.1)
TI	0.55 (0.15)	0.29 (0.15)	0.52 (0.15)	0.75 (0.15)	0.40 (0.15)	6248 (2758)	0.82 (0.15)
PI	0.47 (0.1)	0.27 (0.1)	0.45 (0.1)	0.78 (0.1)	0.26 (0.1)	5247 (2083)	0.96 (0.1)
Post							
CM	0.57 (0.22)	0.55 (0.21)	0.61 (0.21)	0.76 (0.3)	0.49 (0.26)	4518 (1968)	0.96 (0.11)
TI	0.69 (0.16)	0.87 (0.16)	0.72 (0.16)	0.95 (0.16)	0.63 (0.16)	4529 (1928)	0.99 (0.16)
PI	0.82 (0.13)	0.82 (0.13)	0.85 (0.13)	0.98 (0.13)	0.92 (0.13)	2638 (1413)	1 (0.13)
Delayed							
CM	0.59 (0.18)	0.59 (0.17)	0.62 (0.21)	0.77 (0.27)	0.46 (0.28)	3816 (1823)	0.90 (0.23)
TI	0.66 (0.18)	0.74 (0.18)	0.64 (0.18)	0.89 (0.18)	0.62 (0.18)	4710 (2079)	0.98 (0.18)
PI	0.66 (0.16)	0.65 (0.16)	0.62 (0.16)	0.94 (0.16)	0.74 (0.16)	4140 (2058)	0.98 (0.16)

Gender Selection and Translation Task

Figure 4 displays the estimated marginal means and summary statistics for Gender Selection (i.e., accuracy).[4] The model results for Gender Selection revealed a main effect for Time ($F(2, 130) = 33.5$, $p < .001$), which was qualified by a Group x Time interaction ($F(4, 130) = 9.47$, $p < .001$). Pairwise comparisons for Group and Time were conducted to investigate this interaction and are presented in Table 4. At pretest, there were no differences between any of the groups. On the Posttest, the PI group outperformed the TI and CM groups, and the TI group had marginally higher scores than the CM group. On the Delayed Posttest, there were no differences between any of the groups.

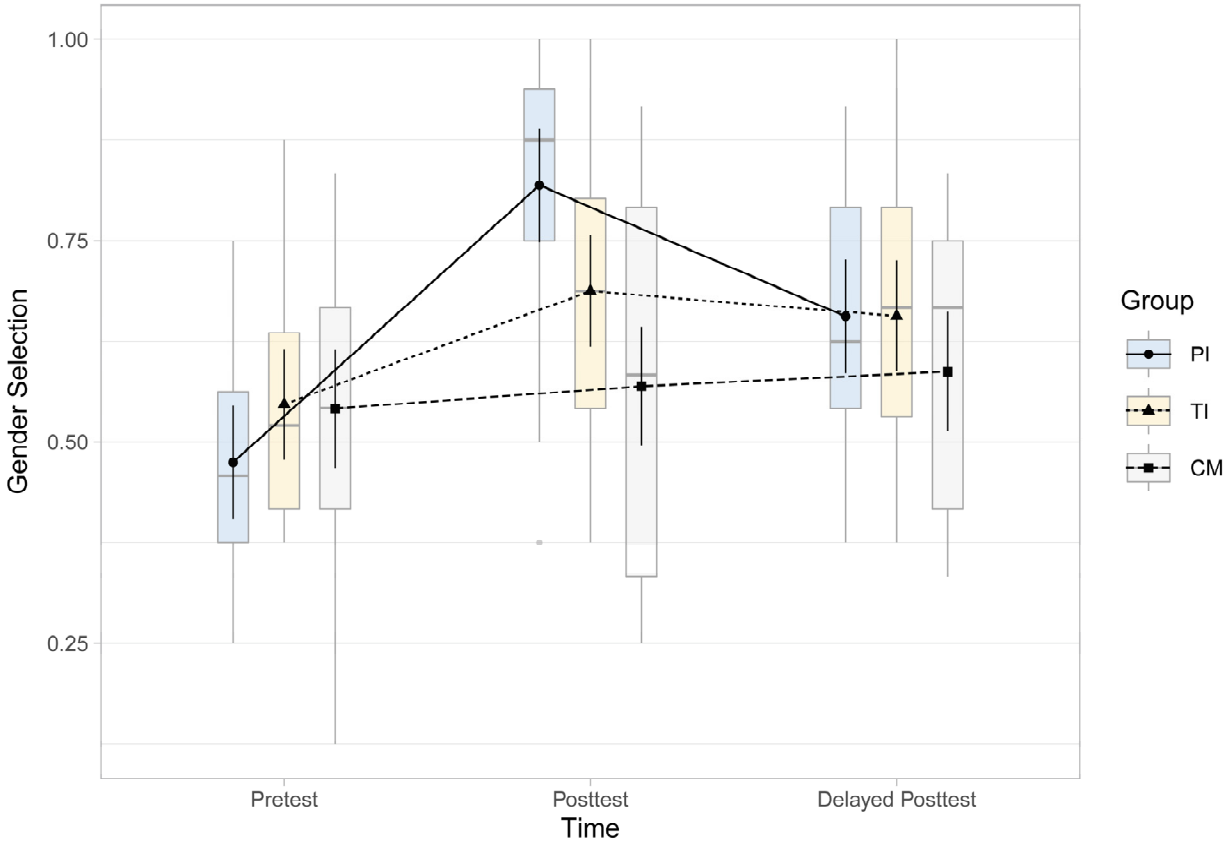

Figure 4. Estimated marginal means for Gender Selection by Group and Time

Table 4. Pairwise Comparisons for Gender Selection by Time and Group

	Pairwise Comparisons by Time			Pairwise Comparisons by Group			
CM	df	*t*	*p*	Pre	df	*t*	*p*
Pre - Post	130	-0.73	0.75	CM - PI	133.77	1.3	0.40
Pre - Delayed	130	-1.20	0.46	CM - TI	133.77	-0.1	0.99
Post - Delayed	130	-0.47	0.89	PI - TI	133.77	-1.45	0.32
PI				Post			
Pre - Post	130	-9.45	< .001***	CM - PI	133.77	-4.83	< .001***
Pre - Delayed	130	-4.97	< .001***	CM - TI	133.77	-2.31	0.06†
Post - Delayed	130	4.47	< .001***	PI - TI	133.77	2.63	0.03*
TI				Delayed			
Pre - Post	130	-3.94	< .001***	CM - PI	133.77	-1.33	0.38
Pre - Delayed	130	-3.07	0.01*	CM - TI	133.77	-1.35	0.37
Post - Delayed	130	0.88	0.66	PI - TI	133.77	-0.01	1.00

Figure 5 displays the estimated marginal means and summary statistics for Translation. The model results for Translation revealed a main effect for Time ($F(2, 130) = 457.76$, $p < .001$), and Group ($F(2, 65) = 7.57$, $p = .001$). These main effects were qualified by a Group x Time interaction ($F(4, 130) = 20.22$, $p < .001$). Pairwise comparisons for Group and Time are in Table 5. Results revealed that there were no differences between any of the groups at pretest. On the Posttest, both the PI and TI groups outperformed the CM group. There was no difference between the PI and TI groups. On the Delayed Posttest, the TI group had higher scores than the CM group. The PI group was similar to both the TI and CM groups. When cognates were removed from the analysis, the TI group also had significantly higher scores than the PI group on the delayed posttest ($M = .67$ vs. $M = .54$, $t(115.76) = -2.4$, $p = .047$).

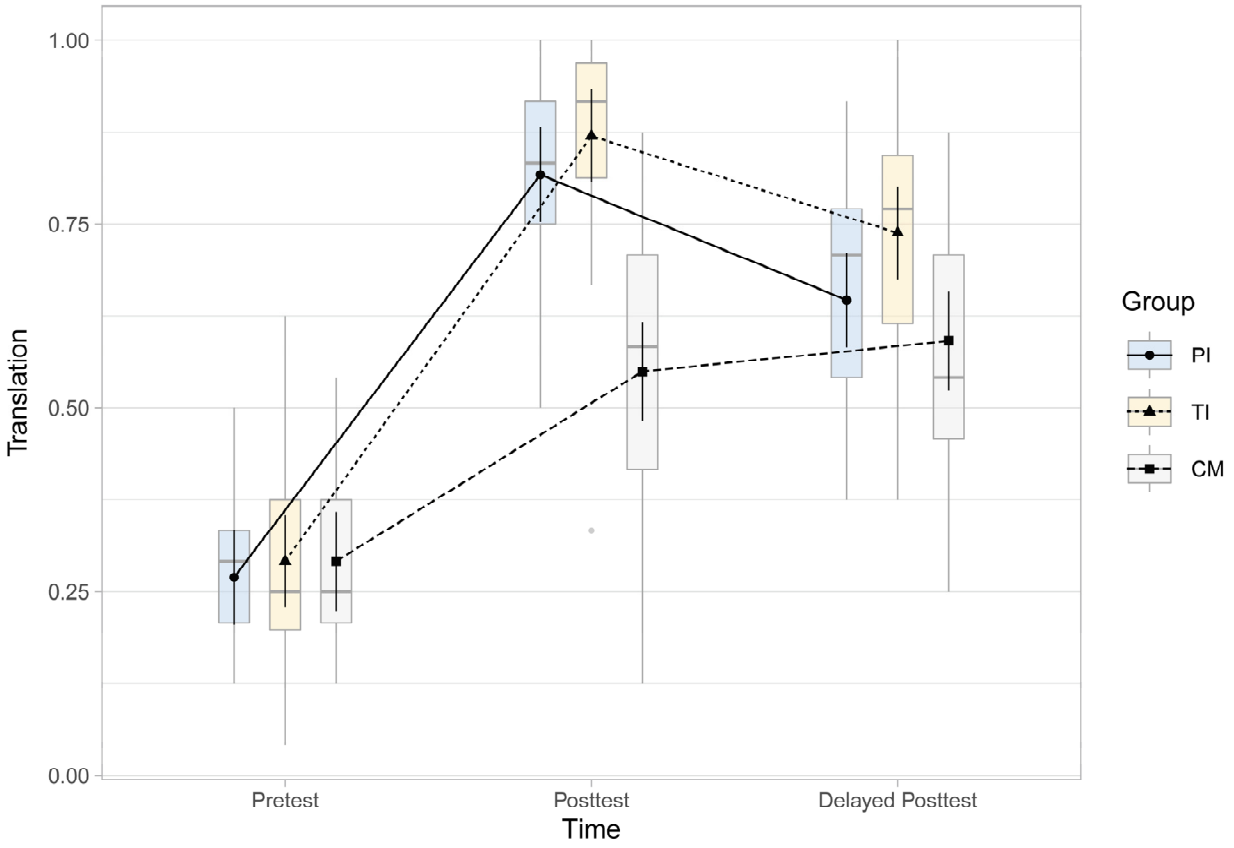

Figure 5. Estimated marginal means for Translation by Group and Time

Table 5. Pairwise Comparisons for Translation by Time and Group

	Pairwise Comparisons by Time			Pairwise Comparisons by Group			
CM	df	*t*	*p*	Pre	df	*t*	*p*
Pre - Post	130	-8.87	< .001***	CM - PI	108.44	0.46	0.89
Pre - Delayed	130	-10.3	< .001***	CM - TI	108.44	0	1
Post - Delayed	130	-1.43	0.33	PI - TI	108.44	-0.48	0.88
PI				Post			
Pre - Post	130	-19.68	< .001***	CM - PI	108.44	-5.7	< .001***
Pre - Delayed	130	-13.55	< .001***	CM - TI	108.44	-6.9	< .001***
Post - Delayed	130	6.13	< .001***	PI - TI	108.44	-1.16	0.48
TI				Delayed			
Pre - Post	130	-21.24	< .001***	CM - PI	108.44	-1.18	0.47
Pre - Delayed	130	-16.39	< .001***	CM - TI	108.44	-3.16	0.01*
Post - Delayed	130	4.85	< .001***	PI - TI	108.44	-2.01	0.11

Production Task

Figure 6 displays the estimated marginal means and summary statistics for (written) Gender Production. The model results for Gender Production revealed a main effect for Time ($F(2, 130) = 33.45$, $p < .001$), which was qualified by a Group x Time interaction ($F(4, 130) = 9.47$, $p < .001$). Pairwise comparisons for Group and Time are in Table 6. Results revealed that there were no differences between any of the groups at pretest. On the Posttest, the PI group had higher scores than the CM group and marginally higher scores than the TI group. On the Delayed Posttest, there were no differences between the groups.

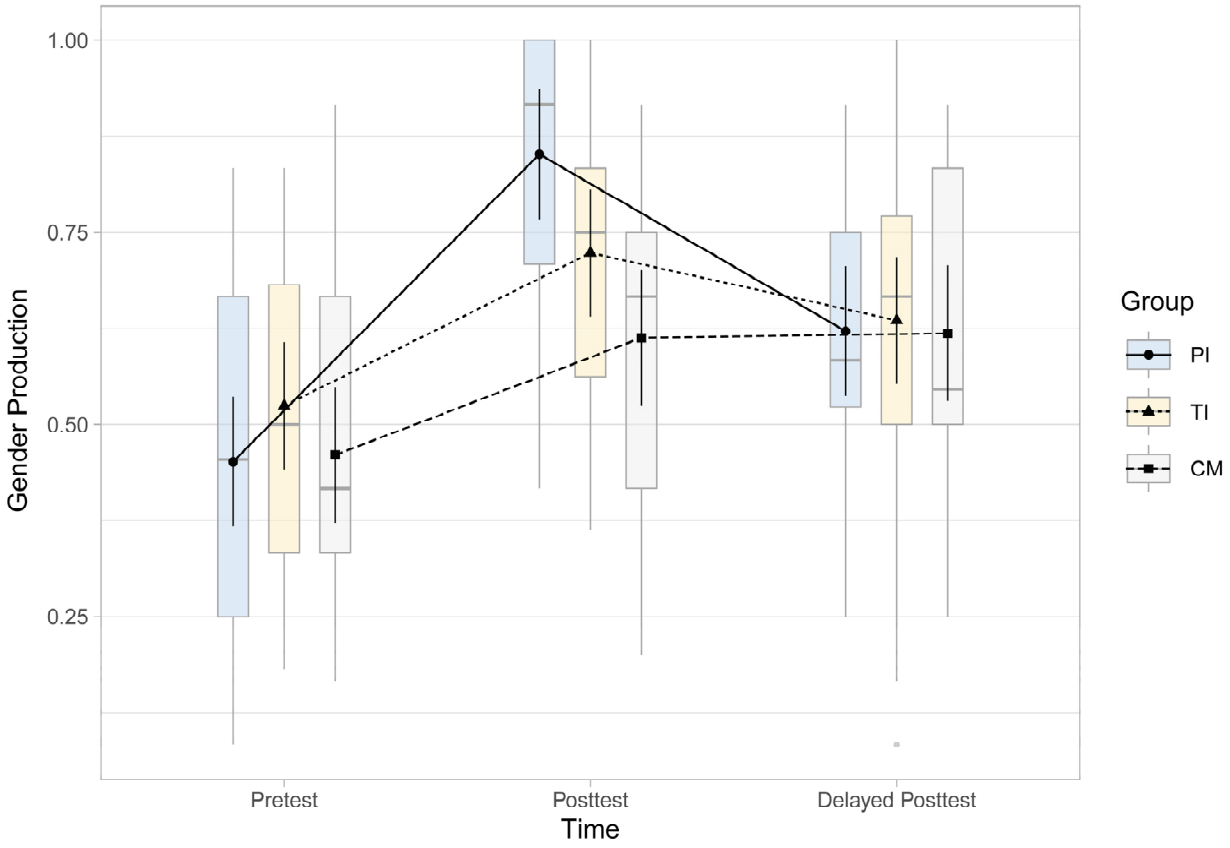

Figure 6. Estimated marginal means for Gender Production by Group and Time

Table 6. Pairwise Comparisons for Gender Production by Time and Group

	Pairwise Comparisons by Time			Pairwise Comparisons by Group			
CM	df	*t*	*p*	Pre	df	*t*	*p*
Pre - Post	130	-3.17	0.01*	CM - PI	144.27	0.14	0.99
Pre - Delayed	130	-3.29	< .001***	CM - TI	144.27	-1.05	0.55
Post - Delayed	130	-0.12	0.99	PI - TI	144.27	-1.22	0.45
PI				Post			
Pre - Post	130	-8.7	< .001***	CM - PI	144.27	-3.87	< .001***
Pre - Delayed	130	-3.7	< .001***	CM - TI	144.27	-1.81	0.17
Post - Delayed	130	5	< .001***	PI - TI	144.27	2.15	0.08†
TI				Delayed			
Pre - Post	130	-4.42	< .001***	CM - PI	144.27	-0.05	1.00
Pre - Delayed	130	-2.47	0.04*	CM - TI	144.27	-0.27	0.96
Post - Delayed	130	1.95	0.13	PI - TI	144.27	-0.23	0.97

The best fit model for Pronoun Production included only the factors Group and Time, but not the interaction between them. Figure 7 displays the estimated marginal means and summary statistics for Pronoun Production. The model results revealed a main effect for Time ($F(2, 125.60) = 12.16$, $p < .001$), and Group ($F(2, 62.84) = 6.12$, $p = .004$). These main effects indicated that the CM group was lower than the PI and TI groups at all test times, and all three groups improved over time.

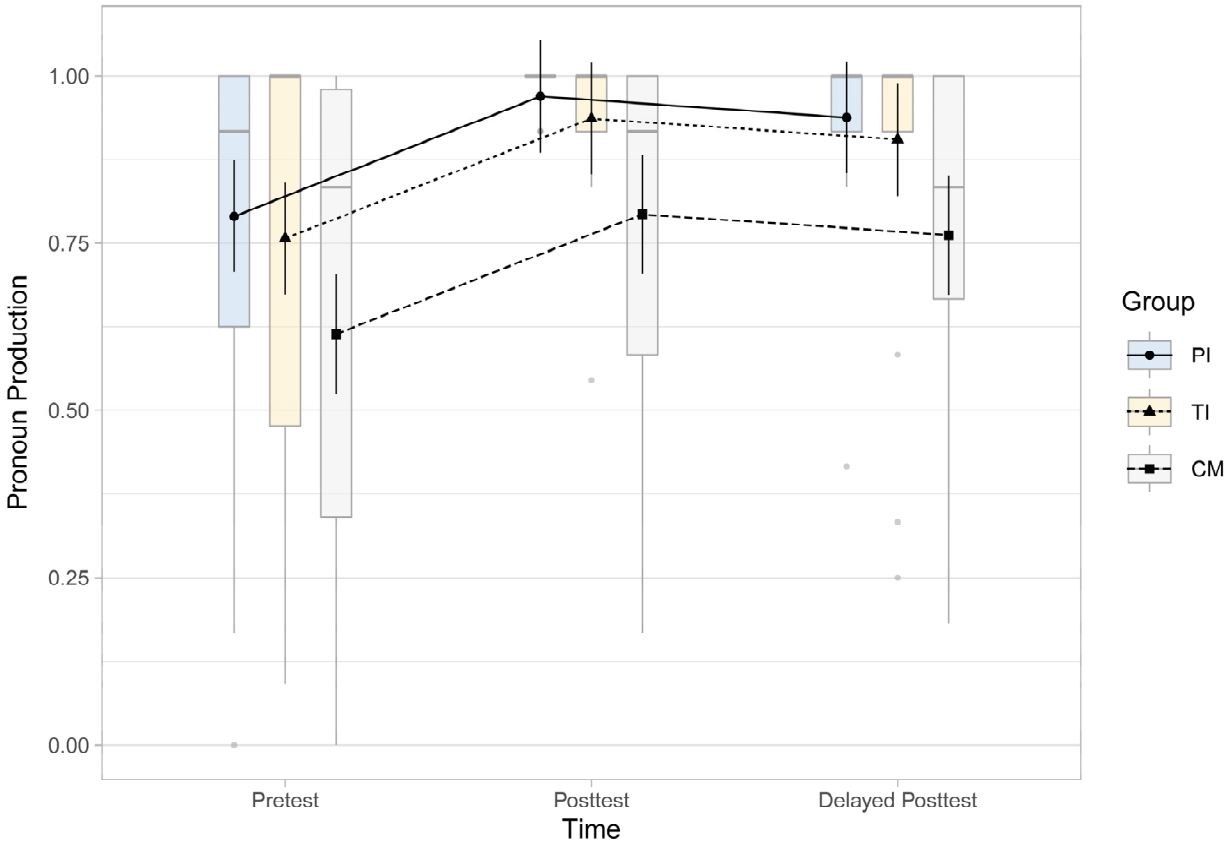

Figure 7. Estimated marginal means for Pronoun Production by Group and Time

Comprehension Task

Figure 8 displays the estimated marginal means and summary statistics for accuracy on the Comprehension task. The model results revealed a main effect for Time ($F(2, 130) = 44.14$, $p < .001$), which was qualified by a Group x Time interaction ($F(4, 130) = 4.97$, $p < .001$). Pairwise comparisons for Group and Time are in Table 7. There were no differences between any of the groups at pretest. On the Posttest, the PI group outperformed both the TI and CM groups. On the Delayed Posttest, the PI group had higher scores than the CM group, but not the TI group. The TI group had marginally higher scores than the CM group.

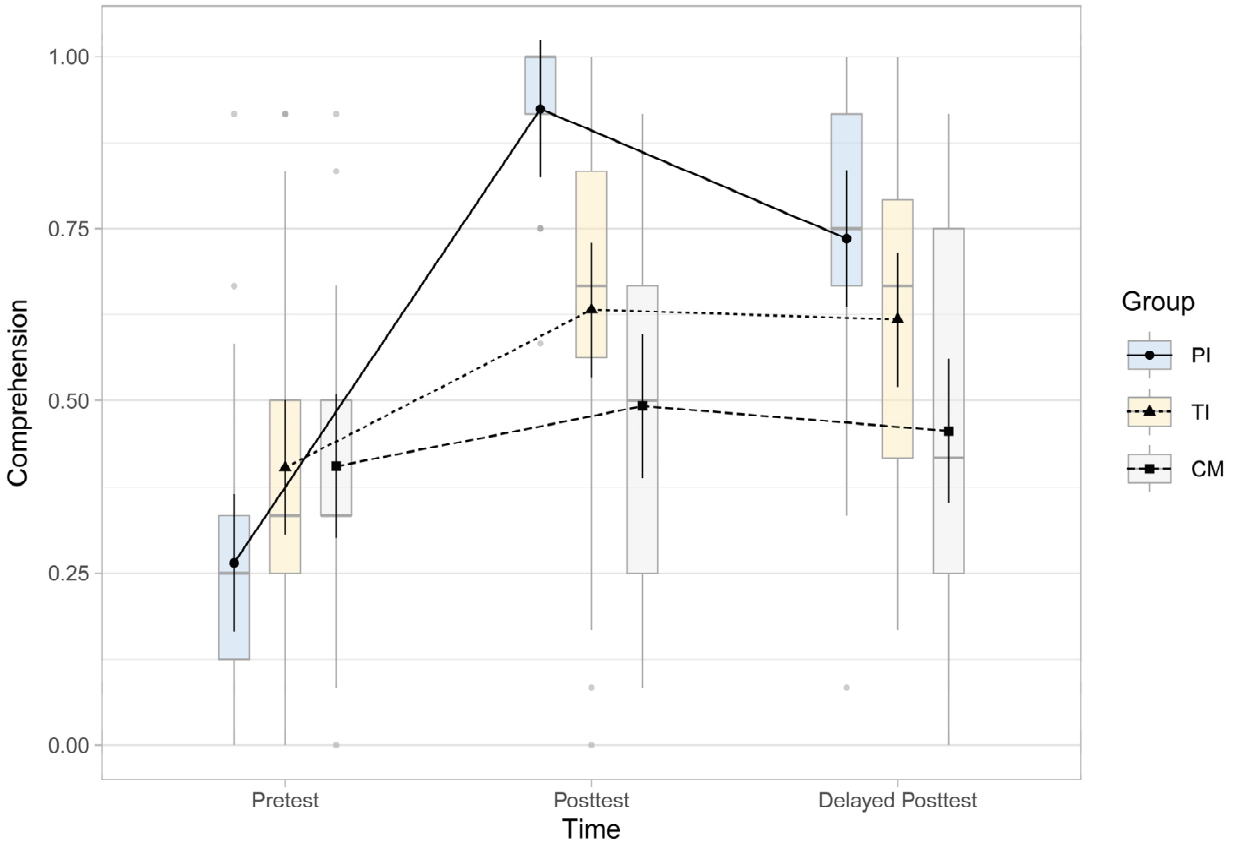

Figure 8. Estimated marginal means for Comprehension accuracy by Group and Time

Table 7. Pairwise Comparisons for Comprehension Accuracy by Time and Group

	Pairwise Comparisons by Time			Pairwise Comparisons by Group			
CM	df	*t*	*p*	Pre	df	*t*	*p*
Pre - Post	130	-1.74	0.19	CM - PI	121.95	1.92	0.14
Pre - Delayed	130	-1.03	0.56	CM - TI	121.95	0.03	1
Post - Delayed	130	0.71	0.76	PI - TI	121.95	-1.96	0.13
PI				Post			
Pre - Post	130	-13.76	< .001***	CM - PI	121.95	-5.93	< .001***
Pre - Delayed	130	-9.83	< .001***	CM - TI	121.95	-1.94	0.13
Post - Delayed	130	3.93	< .001***	PI - TI	121.95	4.14	< .001***
TI				Delayed			
Pre - Post	130	-4.89	< .001***	CM - PI	121.95	-3.83	< .001***
Pre - Delayed	130	-4.59	< .001***	CM - TI	121.95	-2.24	0.07†
Post - Delayed	130	0.3	0.95	PI - TI	121.95	1.67	0.22

Figure 9 displays the estimated marginal means and summary statistics for Reaction Times on the Comprehension Task. The model results revealed a main effect for Time ($F(2, 129.24) = 29.62$, $p < .001$), and a marginal effect for Group ($F(2, 65.01) = 2.75$, $p = .071$). These main effects were qualified by a Group x Time interaction ($F(4, 129.23) = 3.16$, $p = .016$). Pairwise comparisons for Group and Time are in Table 8. There were no differences between any of the groups at pretest. On the Posttest, the PI group was faster than both the TI and the CM groups. On the Delayed Posttest, there were no differences between any of the groups.

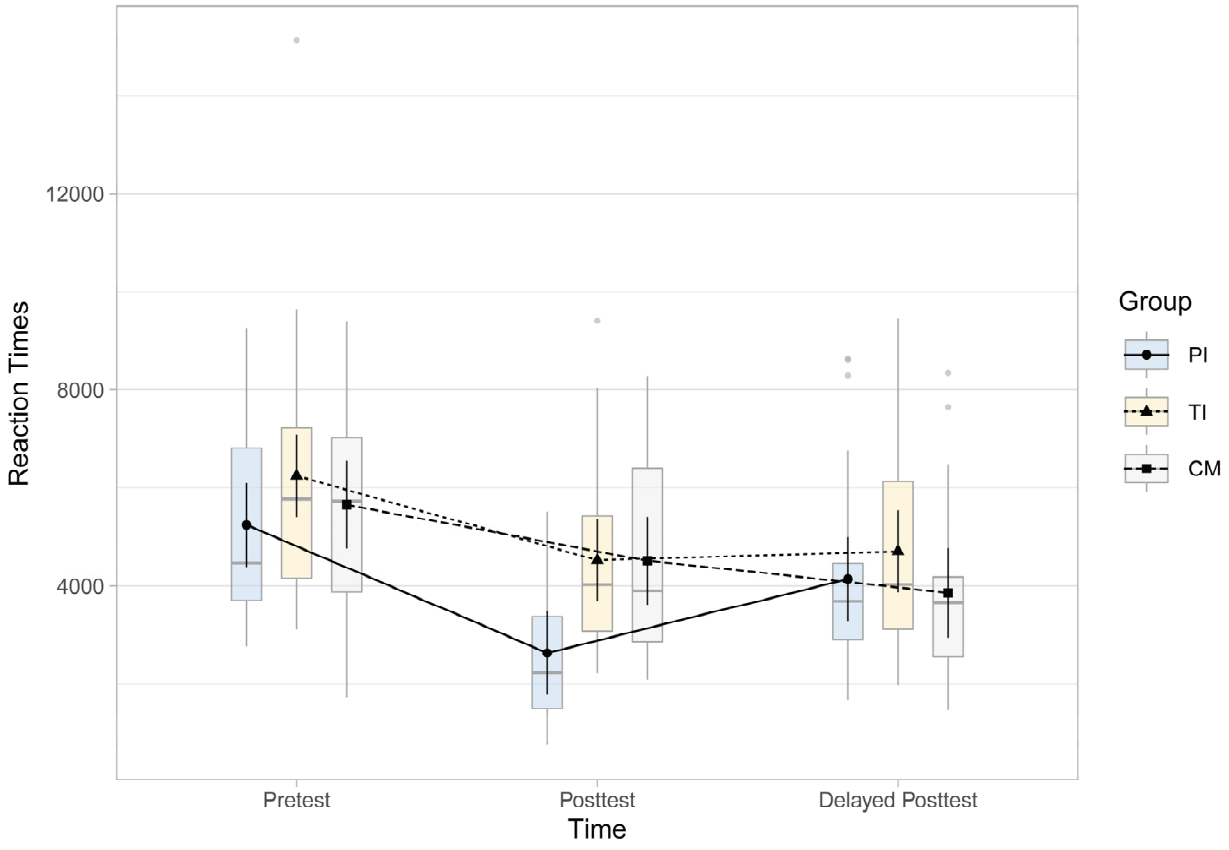

Figure 9. Estimated marginal means for Reaction Times by Group and Time

Table 8. Pairwise Comparisons for Reaction Times by Time and Group

	Pairwise Comparisons by Time			Pairwise Comparisons by Group			
CM	df	*t*	*p*	Pre	df	*t*	*p*
Pre - Post	129.01	2.52	0.03*	CM - PI	129.6	0.66	0.79
Pre - Delayed	129.94	3.91	< .001***	CM - TI	129.6	-0.95	0.61
Post - Delayed	129.94	1.43	0.33	PI - TI	129.6	-1.66	0.23
PI				Post1			
Pre - Post	129.01	6.04	< .001***	CM - PI	129.6	3.01	0.01*
Pre - Delayed	129.01	2.56	0.03*	CM - TI	129.6	-0.02	1
Post - Delayed	129.01	-3.48	< .001***	PI - TI	129.6	-3.13	0.01*
TI				Post2			
Pre - Post	129.01	4.07	< .001***	CM - PI	132.07	-0.44	0.9
Pre - Delayed	129.01	3.64	< .001***	CM - TI	132.12	-1.36	0.37
Post - Delayed	129.01	-0.43	0.9	PI - TI	129.6	-0.94	0.61

Pronoun Matching Task

Figure 10 displays the estimated marginal means and summary statistics for Pronoun Matching. The model results revealed a main effect for Time ($F(2, 130) = 5.11$, $p = .007$), which was qualified by a Group x Time interaction ($F(4, 130) = 5.44$, $p < .001$). Pairwise comparisons for Group and Time are in Table 9. The CM and PI groups outperformed the TI group on the pretest. There were no differences between the groups at Posttest or Delayed Posttest.

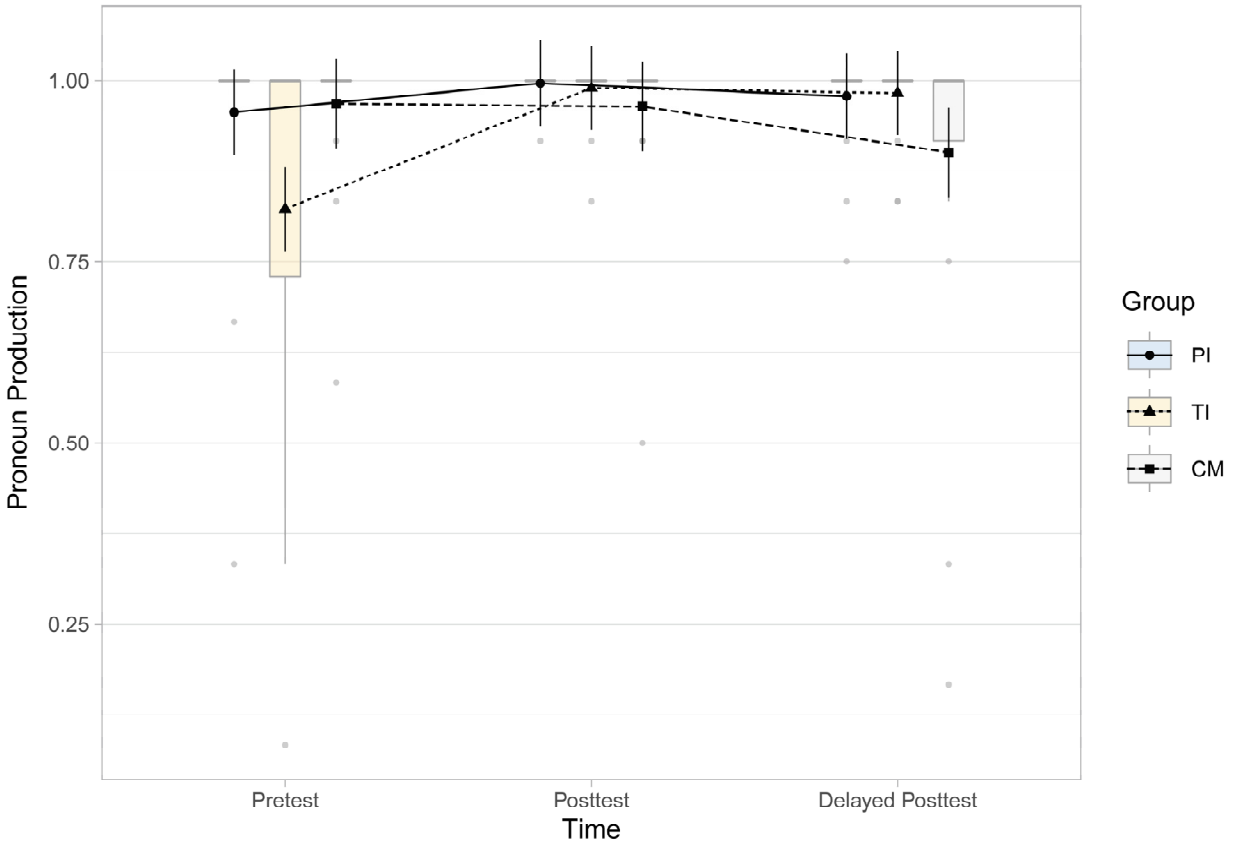

Figure 10. Estimated marginal means for Pronoun Matching by Group and Time

Table 9. Pairwise Comparisons for Pronoun Matching Task by Time and Group

	Pairwise Comparisons by Time			Pairwise Comparisons by Group			
CM	df	*t*	*p*	Pre	df	*t*	*p*
Pre - Post	130	0.1	0.99	CM - PI	171.21	0.27	0.96
Pre - Delayed	130	1.77	0.18	CM - TI	171.21	3.39	< .001***
Post - Delayed	130	1.67	0.22	PI - TI	171.21	3.19	< .001***
PI				Post			
Pre - Post	130	-1.1	0.52	CM - PI	171.21	-0.74	0.74
Pre - Delayed	130	-0.6	0.82	CM - TI	171.21	-0.59	0.83
Post - Delayed	130	0.5	0.87	PI - TI	171.21	0.16	0.99
TI				Delayed			
Pre - Post	130	-4.68	< .001***	CM - PI	171.21	-1.79	0.18
Pre - Delayed	130	-4.49	< .001***	CM - TI	171.21	-1.91	0.14
Post – Delayed	130	0.2	0.98	PI - TI	171.21	-0.1	0.99

Discussion

The results of the present study show a relatively consistent pattern across tasks, with both the PI and TI groups showing improvement over the course of the experiment. While the CM group showed improvement in some tasks, it did not in others. In addition, comparisons between groups showed that the PI group consistently equaled or outperformed both the TI and the CM groups on the Posttest, though they did not retain this advantage on the Delayed Posttest four weeks later. The following sections discuss these findings in relation to gender assignment (RQ1), and pronoun use (RQ2a, 2b, and 2c). I then turn to vocabulary learning (RQ3), pedagogical implications and limitations of the study.

Gender Assignment

Research question 1 asked to what extent PI leads to accurate assignment of gender-marked definite articles, when compared with TI and CM. It was hypothesized that PI would have larger and more durable effects on learners' ability to assign gender accurately on definite articles. Taken together, the results of the Gender Selection and Gender Production measures substantiate this hypothesis: the PI training led to an improved ability to assign gender to the target words. Further, the PI group had higher scores than both the TI and CM groups at posttest. These results clearly indicate that PI is an effective means by which L2 learners can acquire noun-gender pairings. At the same time, the results show that the gains made by the PI group were not fully durable, and, though the PI group's Delayed Posttest scores still showed improvement relative to the pretest, there were no differences between any of the groups on the Delayed Posttest.

These findings are significant because they show that PI is effective for forms with low communicative value, adding support for Benati's (2004) study, which found that PI benefits Italian gender learning. More importantly, however, this finding extends previous research by showing that PI leads to gains even when the target form involves learning gender item-by-item (i.e., lexically-based), not according to a regularized pattern (as in Italian). These findings also support evidence from the psycholinguistic literature that difficulties in L2 gender assignment and gender processing stem from differences between L1/L2 learning processes, specifically, that the associations built between L2 genders and nouns are weaker (Arnon & Ramscar, 2012; Grüter et al., 2012; Hopp, 2016). Because the PI training focused on creating meaningful gender-noun pairings, learners were forced to process this association rather than treating nouns and gender as separate. The subsequent drop in accuracy on the Delayed Posttest could also indicate that noun-gender pairings, such as those created during the PI training, need continual reinforcement. Especially because grammatical gender in German does not follows systematic regularities that are widely available in the input (as in Italian), learners may need consistent exposure to specific lexical items before codifying noun-gender associations in memory.

It is also noteworthy that the TI and CM group improved on gender accuracy during the experiment (though only statistically significant for the CM group in the production task), demonstrating that these instructional methods can benefit learners. While there were no statistical differences between the groups on these measures, it appears that repeated practice with color-coded gender cues did benefit learners to a greater extent than categorization and memorization did. This replicates previous findings that explicit training with color-coded gender cues is perhaps useful but does not generally promote the creation of strong noun-gender associations (Arzt & Kost, 2016, but see Kohler, 2009).

Pronoun Use

Research question 2a, 2b, and 2c asked to what extent PI leads to accurate assignment, comprehension, and production of gender-marked pronouns, compared with TI and CM. It was hypothesized that PI would lead to larger gains for all of these measures.

The hypothesis that the PI group would outperform the TI and CM groups on pronoun assignment and production cannot be substantiated, given that results of the pronoun production measure and pronoun matching tasks are inconclusive. With regard to the former, all three groups improved at roughly the same rate. This is noteworthy particularly because the CM group did not receive any training that should have affected pronoun usage, and gains must therefore be attributed to the participants' outside (i.e., classroom) activities. Thus, it is difficult to determine whether and to what extent the PI and TI trainings may have contributed to the observed gains in pronoun production. Results of the pronoun matching task were similarly inconclusive given that the PI and CM groups were at ceiling at Pretest. These results are perhaps not surprising given the explicitness of the tasks and the ease with which the rule can be applied in explicit matching tasks such as this one. While these two measures produced inconclusive results, the differences between these tasks do underscore the general tendency that learners have more difficulty using gender information in less explicit contexts like the production task.

The hypothesis that PI would lead to greater gains in comprehension was upheld: the comprehension Task clearly showed that the PI group achieved higher accuracy than both the TI and CM groups at Posttest. Although accuracy subsequently dropped, this result points to PI's primary advantage, namely, that training helps learners use pronouns in context to interpret utterances. Not only was the PI group more accurate at Posttest than the CM and TI groups, they

were also significantly faster to respond on the posttest, indicating that the participants were faster to process the gender-marked pronoun for meaning. Taken together, these results suggest that PI is likely to have a larger real-world benefit than either TI or CM, which focus on decontextualized knowledge and do not help learners translate knowledge of grammatical gender into meaningful use. In contrast, the more contextualized practice delivered by PI helps learners form associations between nouns and pronouns, understand that gender information is an important cue for interpreting language in contextualized communicative situations, and process this information more quickly. These results again support learning-based accounts of L1/L2 differences and suggest that L2 learning is more effective when learners process nouns and gender-marked morphology together.

Vocabulary and Implications for Pedagogy

While the focus of the present study was on the acquisition of gender and pronoun use, it was also hypothesized that all three instructional methods would promote vocabulary knowledge. It is therefore noteworthy that the results of the translation task showed that all three methods also led to successful learning of the target vocabulary. Moreover, the PI and TI trainings were more effective than simple memorization, at least in the short term, likely due to the repeated practice offered by training. It is particularly interesting that—despite TI's explicit focus on word learning—there were no differences between the PI and TI trainings at posttest, even though there were some small differences on the delayed posttest when cognates were removed from the data. Given that the PI training was more effective than the TI and CM trainings with respect to gender assignment and pronoun use, the present study suggests that PI would provide the largest opportunity for learning vocabulary, including gender, in the classroom. Importantly, the

decrease in scores across tasks on the delayed Posttest suggests that training—no matter the approach taken—would be most effective if it is made a consistent part of classroom activities, and if learners are continually pushed to process noun and gender pairings for meaning.

In practice, PI trainings such as the one in this study can be implemented in several different ways. One approach, taken by Johnson-Fowler and Jackson (in preparation), is to present short training sessions at the beginning of class. These trainings could be used several times a week as warm-up or cool-down exercises and would be particularly useful to introduce or review words that are relevant to that day's activities. This approach would also help teachers avoid developing lengthy instructional trainings, which can be time consuming and tedious. In addition to in-classroom use, the present study—which used a self-guided computerized training—shows that PI of this type would also be effective for self-study, as homework, or for use in lab classrooms (see also Angelovska & Roehm, 2020; Fowler & Jackson, in preparation). With the rise in virtual language learning platforms, many of which allow instructor-created assignments, teachers could easily assign this type of work to be done at home, replacing textbook activities that do not promote meaningful processing.

Limitations and Future Directions

The present study has several important limitations that could be explored in future research. First, the training and the testing were quite targeted and required relatively simple language use from participants, which may have invited the use of explicit knowledge during some of the tasks. This was most apparent in the Pronoun Matching Task and in the learners' pronoun production, which both had very high scores on the pretest, even though learners at this level rarely use pronouns accurately in connected discourse. Research using discourse-level tasks

and assessments, including oral production would therefore be a welcome addition to this research. Secondly, the PI differed from the TI and the CM trainings in that the PI training used both auditory and written input, whereas the TI and CM trainings used only written input. It is unclear what effect this may have had on the results, but it is possible that this would have advantaged the PI group in the Comprehension Task specifically. Third, it is important to note that, although the TI training in this study used color-coded cues to gender, this study cannot speak to the effect of color-coding itself because there was no equivalent training without color-coded cues; rather, results speak to the entirety of the TI training. While other studies (Arzt & Kost, 2016; Kohler, 2009; Nyikos, 1987) do address color-coding specifically, further research in this area is needed to draw definitive conclusions. Fourth, the present study used relatively short trainings with a relatively long gap between immediate and delayed posttests. It would be useful, particularly given the recommendation for consistent practice, to investigate the level of instruction needed to effect durable noun-gender associations. Finally, it is unclear the extent to which the results are generalizable to other populations (e.g., intermediate or advanced learners), other types of stimuli (e.g., abstract nouns), or other languages with different gender systems (i.e., transparent vs. opaque; two-gender vs. three-gender vs. complex gender systems like Swahili).

Conclusions

The present study demonstrates that Processing Instruction was overall effective in promoting accurate gender assignment, pronoun use, and word learning. While PI did not result in durable changes over the long-term, results suggest that PI could be effective if implemented and used consistently in L2 classrooms. Importantly, these results are consistent with and lend

support to psycholinguistic theories that state that L2 difficulties in gender assignment and processing stem from differences in L1 and L2 learning. Results further suggest that PI could influence learners' ability to process input for gender information more automatically and, perhaps, use gender information online (e.g., in prediction). Future research should continue to investigate whether and how the effects of psycholinguistically motivated trainings like PI translate into broader effects on L2 processing, as approaches such as the one taken here can shed light both on pedagogical practices and psycholinguistic theory.

Notes

1. While the statistical results indicated no differences between the groups in terms of time spent in a German speaking country, the PI group had a comparatively high mean for this measure. This was caused by a single participant. None of the other participants in the PI group had any experience abroad (mean = 0). In order to ensure that this did not affect the study's findings, additional analyses were conducted with this participant removed from the data. Analyses were virtually identical regardless of this participant's inclusion in the data, and thus, this participant was included in the analyses reported in the remainder of the study.
2. This was done to ensure that results could differentiate between gender and word learning. Results were also tabulated with a combined gender and translation score, in which participants received one point, only if both the gender and translation were correct. These results mirrored the results of the translation task and will not be presented due to space limitations.
3. In order to investigate whether other trims may have changed results, three additional analyses were conducted and are available in the supplemental materials published online. None of these analyses changed the primary pattern of results presented here.
4. This figure and similar figures throughout the results were generated using the afexplot function in the afex package (Singman, et al., 2021). Shapes and lines in the foreground show estimated marginal means and model-based standard errors. Box plots in the background represent the raw data, including the median line and outlier points.

Funding / Conflicts of Interest

There is no funding information to report. The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

I would like to thank Courtney Johnson Fowler for her work on an earlier iteration of this project, which had a substantial impact on the present research. I would also like to thank Melisa Dracos, Amanda Hua, Carrie Jackson, and Preston Norris for their numerous contributions. Portions of this research were presented at the 2019 Germanic Linguistics Annual Conference in Iowa City, Iowa, USA.

References

Angelovska, T., & Roehm, D. (2020). Processing instruction effects regardless of input modality and developmental processing constraints? A school lab classroom study on the morphosyntactic acquisition of L2 English. Instructed Second Language Acquisition, 4(2), 180–202.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*(3), 292–305.

Arzt, J., & Kost, C. (2016). Effect of Different Teaching Techniques on the Acquisition of Grammatical Gender by Beginning German Second Language Learners. *Die Unterrichtspraxis/Teaching German*, *49*(1), 1–23.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Benati, A. (2001). A comparative study of the effects of processing instruction and output-based instruction on the acquisition of the Italian future tense. *Language Teaching Research*, *5*(2), 95–127.

Benati, A. (2004). The effects of processing instruction and its components on the acquisition of gender agreement in Italian. *Language Awareness*, *13*(2), 67–80.

Benati, A. (2005). The effects of processing instruction, traditional instruction and meaning-output instruction on the acquisition of the English past simple tense. *Language Teaching Research*, *9*(1), 67–93.

Benati, A. & T. Angelovska (2015). The Effects of Processing Instruction on the Acquisition of English Past Simple Tense: Age and Cognitive Tasks Demands. In: International Review of Applied Linguistics in Language Teaching (IRAL) 53 (2), 249-269

Bjornstad, J. (2014). Der? Die? Das? Easing the struggle to learn definite articles. *Die Unterrichtspraxis/Teaching German*, *47*(1), 69–78.

Cadierno, T. (1995). Formal instruction from a processing perspective: An investigation into the Spanish past tense. *The Modern Language Journal*, *79*(2), 179–193.

Chan, A., Lieven, E., & Tomasello, M. (2009). Children's understanding of the agent-patient relations in the transitive construction: Cross-linguistic comparisons between Cantonese, German, and English. *Cognitive Linguistics*, *20*(2), 267–300.

Cheng, A. C. (2002). The effects of Processing Instruction on the acquisition of ser and estar. *Hispania*, *85*(2), 308–323.

Chew, J. S. (1989). Learning german nouns: bridging the gender gap. *Die Unterrichtspraxis/Teaching German*, *22*(1), 52-54.

Cintrón-Valentín, M., & Ellis, N. C. (2016). Salience in Second Language Acquistion: Physical form, learner attention, and instructional focus. *Frontiers in Psychology, section Language Science*, *7*, 1–21.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(01), 3–42.

Collentine, J. (1998). Processing Instruction and the Subjunctive. *Hispania*, *81*(3), 576–587.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic Gender and Spoken-Word Recognition in French. *Journal of Memory and Language*, *42*(4), 465–480.

Desrochers, A., Wieland, L. D., & Cote, M. (1991). Instructional effects in the use of the mnemonic keyword method for learning German nouns and their grammatical gender. *Applied Cognitive Psychology*, *5*, 19–36.

Dracos, M. & Henry, N. (2018). The effects of task-essential training on L2

processing strategies and the development of Spanish verbal morphology. Foreign Language Annals, 51, 344-368. https://doi.org/10.1111/flan.12341

Dracos, M. & Henry, N. (2021). Task essential training and L2 Spanish verbal morphology: An investigation of online processing and working memory capacity. Languages, 6(1), 24. https://doi.org/10.3390/languages6010024

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand. *Studies in Second Language Acquisition*, *35*(2), 353–387.

Fernández, C. (2008). Reexamining the role of explicit information in Processing Instruction. *Studies in Second Language Acquisition*, *30*, 277–305.

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1-L2 syntactic similarity. *Bilingualism*, *14*(3), 379–399.

Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, *17*(3), 213–247.

Gillon Dowens, M., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish - Evidence from Event Related Potentials. *Neuropsychologia*, *49*(7), 1651–1659.

Gillon Dowens, M., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, *22*(8), 1870–1887.

Grünloh, T., Lieven, E., & Tomasello, M. (2011). German children use prosody to identify participant roles in transitive sentences. *Cognitive Linguistics*, *22*(2), 393–419.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or

a real-time processing problem? Second Language Research, 28(2), 191-215.

Hawkins, R., & Chan, C. Y. H. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis.' *Second Language Research*, *13*(3), 187–226.

Henry, N. (2022). The effects of prosody on the processing of morphosyntax in L2 German. Studies in Second Language Acquisition. Advanced Online Publication, 2022. https://doi.org/10.1017/S0272263122000092

Henry, N., Culman, H. & VanPatten, B. (2009). More on the effects of explicit information in instructed SLA. Studies in Second Language Acquisition, 31, 559-575. https://doi.org/10.1017/S0272263109990027

Henry, N., Jackson, C. N., DiMidio, J. (2017). The role of explicit instruction and prosodic cues in Processing Instruction. The Modern Language Journal, 101, 294-314. http://dx.doi.org/10.1111/modl.12397

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56.

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain research*, *1626*(11), 118–135.

Jackson, C. N., Fowler, C. J., Gavin, B., & Henry, N. (2018). Zusammenhänge zwischen der Sprachverarbeitung und dem Erwerb neuer Strukturen unter erwachsenen Lernern [Connections between sentence processing and the acquisition of new structures among adult L2 learners]. In S. Schimke & H. Hopp (Eds.), Sprachverarbeitung im

Zweitspracherwerb [Sentence Processing and Second Language Acquisition]. Amsterdam: De Gruyter. https://doi.org/10.1515/9783110456356-013

Johnson Fowler, C., & Jackson, C.N. (in preparation). Learning grammatical gender through processing instruction methods: making L2 learners aware of gender's role in the language.

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, *59*(3), 503–535.

Kohler, T. (2009). *Using neurocognitive theory to develop a new approach for teaching German grammar the effect of color-coding German grammar on language acquisition*. (Unpublished Doctoral Dissertation). University of Kansas. Lawrence, KS.

Köpke, K., & Zubin, D. (1984). Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation, *Linguistische Berichte, 93,* 26-50.

Kraiss, A. (2014). Taking the awful out of the German language: A study of a new way to teach German gender and plural assignment. *Die Unterrichtspraxis / Teaching German*, *47*(1), 23–41.

Lemhöfer, K., Schriefers, H., & Indefrey, P. (2014). Idiosyncratic Grammars: Syntactic Processing in Second Language Comprehension Uses Subjective Feature Representations, *Journal of Cognitive Neuroscience, 26*(7)*,* 1428–1444.

Lew-Williams, C., & Fernald, A. (2009). Fluency in Using Morphosyntactic Cues to Establish Reference: How Do Native and Non-Native Speakers Differ? In J. Chandlee, M. Franchini, S. Lord, & G. Rheiner (Eds.), *Proceedings of the 33rd annual Boston University conference on language development, Vol. 1*, pp. 290-301. Cascadilla Press.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198.

Marsden, E., & Chen, H. Y. (2011). The roles of structured input activities in processing instruction and the kinds of knowledge they promote. *Language Learning*, *61*(4), 1058–1098.

Marsden, E., Mackey A., & Plonsky, L. (2016). The IRIS Repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS Repository of Instruments for Research into Second Languages* (pp. 1-21). New York: Routledge.

McManus, K., & Marsden, E. (2019). Signatures of automaticity during practice: Explicit instruction about L1 processing routines can improve L2 grammatical processing. *Applied Psycholinguistics*, *40*(1), 205–234.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: an event-related potential study. *Language learning*, *60*(1), 154–193.

Nyikos, M. (1987). *The effect of color and imagery as mnemonic strategies on learning and retention of lexical items in German*. (Unpublished Doctoral Dissertation). Purdue University. West Lafayette, IN.

Presson, N., MacWhinney, B., & Tokowicz, N. (2012). Learning grammatical gender: The use of rules by novice learners. *Applied Psycholinguistics*, *35*(4), 709–737.

R Studio Team. (2021). RStudio: Integreted development environment for R. Boston, MA: RStudio, PBC.

Ritterbusch, R., LaFond, L., & Agustin, M. (2008). Learner Difficulties with German Case: Implications from an Action Research Study. *Die Unterrichtspraxis/Teaching German*, *39*(1–2), 30–45.

Santos, V. D. O. (2015). Can colors, voices, and images help learners acquire the grammatical gender of German nouns? *Language Teaching Research*, *19*(4), 473–498.

Schempp, P. L. (2017). *L2 learners' processing of grammatical gender varies according to cognate status and proficiency: an ERP study*. (Unpublished Doctoral Dissertation). The Pennsylvania State University. University Park, PA.

Scherag, A., Demuth, L., Rösler, F., Neville, H. J., & Röder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition*, *93*(3), 97–108.

Schirrmeister, L. (2015). Möglichkeiten der regelgeleiteten Genusvermittlung im Deutschen als Fremdsprache. In B. Handwerker, R. Bäuerle, I. Doval, B. Lübke (Eds.). Zwischenräume: Lexikon und Grammatik im Deutschen als Fremdsprache, 137–154.

Schriefers, H., & Jescheniak, J. D. (1999). Representation and processing of grammatical gender in language production: A review. *Journal of Psycholinguistic Research*, *28*(6), 575–600.

Shintani, N. (2014). The Effectiveness of Processing Instruction and Production-based Instruction on L2 Grammar Acquisition: A Meta-Analysis. *Applied Linguistics*, *1*(November), 1–21.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Schachar, M. (2021). afex: Analysis of Factorial Experiments. Retrieved from https://cran.r-project.org/package=afex

Trudgill, P. (1999). Language contact and the function of linguistic gender. *Poznan Studies in Contemporary Linguistics, 35, 133-152.*

VanPatten, B. (2015a). Foundations of processing instruction. *International Review of Applied Linguistics in Language Teaching*, *53*(2), 91–109.

VanPatten, B. (2015b). Input processing in adult SLA. In B. VanPatten & J. Williams (Eds.),

Theories in Second Language Acquisition (2nd ed., pp. 113–134). Mahwah, NJ: Lawrence Erlbaum.

VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, *15*, 225.

Wong, W., & Ito, K. (2018). The effects of processing instruction and traditional instruction on L2 online processing of the causative construction in French: An eye-tracking study. Studies in Second Language Acquisition, 40, 241–268. https://doi.org/10.1017/S0272263117000274