Studies in Second Language Acquisition, 2016, **38**, 265–291.
doi:10.1017/S0272263116000139

THE EFFECTIVENESS OF GUIDED INDUCTION VERSUS DEDUCTIVE INSTRUCTION ON THE DEVELOPMENT OF COMPLEX SPANISH *GUSTAR* STRUCTURES

An Analysis of Learning Outcomes and Processes

Luis Cerezo
American University

Allison Caras and Ronald P. Leow
Georgetown University

Meta-analytic research suggests an edge of explicit over implicit instruction for the development of complex L2 grammatical structures, but the jury is still out as to which type of explicit instruction—*deductive* or *inductive*, where rules are respectively provided or elicited—proves more effective. Avoiding this dichotomy, accumulating research shows superior results for *guided induction*, in which teachers help learners co-construct rules by directing their attention to relevant

We would like to thank Georgetown University's Center for New Designs in Learning and Scholarship (CNDLS) for awarding the Initiative on Technology-Enhanced Learning (ITEL) grant to Ronald P. Leow, which permitted him to create *The Gustar Maze*. We would also like to thank Bill Garr for developing the software and Janire Zabildea, Jorge Mendez Seija, and Celia Chomón Zamora for assisting with data collection and coding. Special thanks to the editors of this volume and the three reviewers for their insightful feedback and comments.

Correspondence concerning this article should be addressed to Ronald P. Leow, Department of Spanish and Portuguese, ICC 411, Georgetown University, 37th and O Streets, NW, Washington, DC, 20057, USA. E-mail: leowr@georgetown.edu

© Cambridge University Press 2016

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

aspects in the input and asking guiding questions. However, no study has jointly investigated the effects of guided induction on both learning outcomes and processes, or whether guided induction can prove effective outside classroom settings where teacher mediation is not possible. In this study, which targeted the complex Spanish *gustar* structures, 70 English-speaking learners of beginning Spanish received either guided induction via a videogame, deductive instruction in a traditional classroom setting, or no instruction. Learning outcomes were measured via one receptive and two controlled production tasks (oral and written) with old and new items. Results revealed that while both instruction groups improved across time, outperforming the control group, the guided induction group achieved higher learning outcomes on all productive posttests (except immediate oral production) and experienced greater retention. Additionally, the think-aloud protocols of the guided induction group revealed high levels of awareness of the L2 structure and a conspicuous activation of recently learned knowledge, which are posited to have contributed to this group's superior performance. These findings thus illustrate, quantitatively and qualitatively, the potential of guided induction for the development of complex L2 grammar in online learning environments.

INTRODUCTION

Recent research meta-analyses concluded that explicit types of instruction (including metalinguistic rules or directions to infer them) proved more effective than implicit types (those that do not) for the development of L2 grammatical structures (Goo, Granena, Yilmaz, & Novella, 2015), both simple and complex (Spada & Tomita, 2010). However, these meta-analyses subsumed both deductive instruction (in which rules are provided) and inductive instruction (in which rules are elicited) as "explicit instruction" and consequently did not clarify which type of explicit instruction yields the best results.

From a theoretical perspective, *deductive instruction* can accelerate second language (L2) development because it allows learners to identify the locus and nature of first language (L1)–L2 contrasts (Carroll, 2001), helping learners to convert declarative into procedural and automatized knowledge through practice (the "strong interface" position; DeKeyser, 1995). Additionally, knowledge of rules can indirectly promote learning because it makes learners more prone to notice the L2 form as it appears in subsequent input (e.g., Ellis, 2002). Many scholars, however, have

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

objected that deductive instruction typically emphasizes form over meaning (Brooks & Donato, 1994), promotes passive participation (Herron & Tomasello, 1992), may create false illusions of understanding (Shaffer, 1989), or may only work for learners who can correctly understand the rules, depending on factors such as their knowledge of metalanguage (e.g., Elder & Manwaring, 2004).

On the other hand, the problem-solving nature of *inductive instruction* is posited to induce greater depth of processing that leads to more memorable and meaningful experiences for learners (Craik & Lockhart, 1972; Oded & Walters, 2001). Yet inductive instruction may be slow, requiring extensive exposure to L2 input (Ellis, 1993); it may work only for learners that can successfully infer the underlying rules (e.g., Erlam, 2005) or whose working memory is not overwhelmed by new information (Kirschner, Sweller, & Clark, 2006), and it may even be counterproductive if learners make false assumptions (Hall, 1998).

From an empirical perspective, results are inconclusive. Some studies found an edge for deductive instruction (e.g., Erlam, 2003; Robinson, 1996), others found no differential effects of deductive and inductive instruction (e.g., Rosa & O'Neill, 1999; Shaffer, 1989), and others—a minority—found an edge for inductive instruction (Ayoun, 2001). Arguably, these inconclusive findings may be due to methodological differences (see Vogel, Herron, Cole, & York, 2011), including the typically short nature of the treatments and an insufficient amount of practice, which may have put inductive conditions at a disadvantage (Cerezo, 2016; Ellis, 1993).

Given the pros and cons of each type of instruction and the inconclusive results of the extant research, some have called for a reappraisal of language instruction that avoids the dichotomies of purely deductive versus inductive instruction (e.g., Adair-Hauck, Donato, & Cumo-Johanssen, 2010; Toth, Wagner, & Moranski, 2013). In that vein, a few studies have shown superior results for *guided induction*, in which teachers help learners coconstruct rules by directing their attention to relevant aspects in the input, asking guiding questions, or both (e.g., Haight, Herron, & Cole, 2007; Herron & Tomasello, 1992; Smart, 2014; Vogel et al., 2011). However, none of these studies have jointly investigated the effects of guided induction on both learning outcomes and processes, or whether guided induction can prove effective outside classroom settings in which teacher mediation is not possible. To fill in this gap, the present study investigated the effectiveness of guided induction via an educational video game for the development of complex Spanish *gustar* structures, as compared to a traditional face-to-face deductive instruction group and a control group. Additionally, the cognitive processes triggered by guided induction were tapped via think-aloud protocols.

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

REVIEW OF THE LITERATURE

The Effects of Guided Induction and Deductive Instruction on Learning Outcomes

Very few published studies have compared the effectiveness of guided induction and deductive instruction on L2 development (e.g., Haight et al., 2007; Herron & Tomasello, 1992; Smart, 2014; Vogel et al., 2011). These studies all operationalized deductive instruction as the presentation, practice, production (PPP) approach, but they differed in their operationalizations of guided induction, as explained subsequently.

In Herron and Tomasello's (1992) seminal study, 26 college students of beginning French learned 10 grammatical structures over the course of a semester. Following a counterbalanced design, participants were presented one structure per week, alternatively switching from a deduction to a guided induction condition or vice versa. In the guided induction condition, participants chorally completed (a) a contextualized oral drill and (b) a written fill-in-the-blank activity. The teacher provided right/wrong feedback or models, without explaining grammar rules or inviting learners to formulate them. Results from composite scores on a written fill-in-the-blank task showed an edge for the guided induction condition, both immediately after the treatment and one week later.

Adair-Hauck and Donato expanded Herron and Tomasello's guided induction model into the PACE model (see Adair-Hauck et al., 2010, for a recent account). Drawing on the Vygotskyan notion of *scaffolding*, in the PACE model (a) the teacher *presents* a written or oral narrative that embeds the grammatical structure; (b) learners focus their *attention* on the structure by, for example, locating and underlining tokens; (c) the teacher helps learners *coconstruct* grammar rules by asking guiding questions; and (d) learners complete an *extension* activity in which they need to use the grammatical structure to carry out a task.

Herron and colleagues (Haight et al. 2007; Vogel et al., 2011) conducted two partial replications of Herron and Tomasello (1992), adding the coconstruction phase of the PACE model at the end of the guided induction treatments and targeting different structures and participants. Unlike the PACE model, however, the participants in these studies could only answer prescripted guiding questions, rather than also asking their own, and the teacher did not disclose the correct rules. The results of Haight et al. (2007), which tested 47 college students of elementary French on eight grammatical structures, showed an edge for guided induction, both after the treatment on a written fill-in-the-blank task and 14 weeks later on a written multiple-choice task. Vogel et al. (2011), testing 40 college students of intermediate French on 10 grammatical structures, partially confirmed these findings. Guided induction

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

yielded higher achievement immediately after the treatment, as measured on a written sentence-reconstruction task in which constituents had been parenthesized. However, it performed comparably to deductive instruction at the end of the semester on a written multiple-choice task.

More recently, Smart (2014) implemented a corpus linguistics approach to guided induction. Forty-nine college students of advanced English grammar completed two weeks of instruction on the passive voice under three conditions—data-driven learning through guided induction, deductive-corpus informed instruction, and traditional grammar instruction. Based on Flowerdew's (2009) four *I*s model, in the guided induction group (a) learners received preselected concordance lines that *illustrated* differences in form, meaning, and use of the grammatical structure; (b) learners *interacted* in groups to discuss their observations; (c) the teacher optionally *intervened* to provide learners with hints for induction; and (d) learners completed productive writing activities in which they applied the rules of their own *induction* process. Results from three written assessment tasks (grammaticality judgment with error correction, active/passive sentence rewriting, and register awareness) showed that only the guided induction group improved significantly, both immediately after the treatment and two weeks later, and for both composite and individual task scores, except register awareness.

Overall, then, existing studies suggest an edge for guided induction over deductive instruction on L2 development, as measured on a variety of written assessment tasks, including multiple-choice recognition, fill-in-the-blank production, sentence reconstruction, and grammaticality judgment with error correction. This edge held both immediately after the treatment and after a short period of time (1 to 2 weeks); however, findings are contradictory after longer periods (cf. Haight et al., 2007; Vogel et al., 2011). These results, however, must be taken with caution due to the paucity of studies, methodological differences and caveats (e.g., the studies by Herron and colleagues included only two to four test items per grammatical structure), and different operationalizations of guided induction. Moreover, no study documented *how* guided induction propelled superior learning by analyzing learners' cognitive processes; thus the topic warrants further research.

Cognitive Processes in Guided Induction

To investigate how guided induction may promote L2 development, studies need to triangulate measures of L2 learning outcomes with learning processes. Focusing on the latter, Toth et al. (2013) analyzed how learners coconstructed rules for the Spanish pronominal clitic *se*, first in groups of two/three students and later with the teacher as a

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

class. Cognitive processes were coded under four categories: *labeling* (observations about morphosyntactic tokens or phrases, such as naming a form or its meaning), *categorizing* (commenting on properties shared by several tokens), *patterning* (commenting on links between two categories of form, such as subject-verb agreement, or between form and meaning, such as the use of *se* without an explicit agent), and *rule formulation* (extrapolation of patterns to the Spanish language). Results showed that analytic talk was more prevalent in whole-class discussions with the teacher (57%) than in small-group discussions among learners (17%), with great variability in both contexts (see Wagner & Toth's 2013 follow-up study). Patterning was the predominant cognitive process, taking up around half the time of analytic talk (50% and 44% whole-class and small-group discussions, respectively). The teacher and learners played different roles during coconstruction. They both spent a similar amount of time labeling, but the teacher spent more time categorizing, which seemed to help learners produce higher rates of patterning and rule formulation.

The studies by Toth and colleagues illustrated the learning processes during *dialogic* coconstruction of rules in guided induction, but they did not establish a link with L2 development. In contrast, other studies (see Leow, 2012, for a review) have both documented learning processes and measured learning outcomes in guided induction-like conditions while learners *individually* completed "consciousness-raising tasks."[1] Using concurrent think-aloud protocols, these studies coded learners' reported verbalizations of L2 *awareness*, "a particular state of mind in which an individual has undergone a specific subjective experience of some cognitive content or external stimulus" (Tomlin & Villa, 1994, p. 193). Awareness was generally coded at three levels—*noticing*, *reporting*, and *understanding*—the latter including various processes, such as *hypothesis formulation* (formulating hypotheses about the rule underlying a target item), *rule formulation* (providing a full or partial grammatical rule for the target item), and *activation of prior knowledge* (using some prior knowledge to encode or decode the target item).

Overall, studies have reported a positive correlation between greater awareness and, more generally, depth of processing (i.e., the "amount of cognitive effort . . . employed in decoding and encoding some grammatical or lexical item in the input"; Leow, 2015, p. 204) and amount of L2 development (Leow, 2012). This is consistent with Leow's (2015) model of instructed SLA, according to which processing L2 data in greater depth may result in higher levels of awareness that may facilitate the incorporation of intake[2] into the learner's developing system.

The present study coded learners' cognitive processes in terms of depth of processing and levels of awareness. However, these processes can be interconnected with other coding protocols. For example, *hypothesis* and *rule formulation* can be posited to correspond to *patterning*

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

and *rule formulation* in Toth et al.'s (2013) study. Our discussion section invokes these interconnections to illustrate the (dis)similarities between the coconstruction of explicit knowledge in classroom contexts (among learners, or between learners and the teacher) and online contexts (between learners and the video game).

The Complexity of Spanish *Gustar* Structures for English-Speaking L2 Learners

A multidimensional construct, L2 complexity has been viewed from at least four perspectives—developmental, psycholinguistic, linguistic, and pedagogical (see Housen, 2014; Housen & Simoens, 2016). In what follows, we discuss the complexity of Spanish *gustar* structures for our L1 English participants.

From a *developmental* perspective, Spanish *gustar* structures are typically learned relatively late by English speakers, if not in terms of emergence then at least in terms of accurate mastery (Montrul, 1997). From a *psycholinguistic* and *linguistic* perspective, Spanish *gustar* structures pose problems for L1 English speakers trying to process or produce them, due to, among other factors, the linguistic contrasts with equivalent "to like" structures.

First, *gustar* is an intransitive verb that requires a dative experiencer, whereas "to like" is a transitive agentive verb. In that sense, Spanish *gustar* structures behave similarly to English "to please" structures, as illustrated in (1), (2), and (3). Second, English and Spanish also differ in how word order determines functional assignment, a contrast that has been posited as a source of L2 complexity (Collins, Trofimovich, White, & Horst, 2009). English is a subject-verb-object (SVO) language that relies mainly on word order for functional assignment, whereas Spanish, given its morphologically richer nature, has more flexible word order, as shown in (1) versus (3a) or (3b):

(1) English:

I	like	the house.
I-NOM	1st Sg. like	ACC-the house

(2) English:

The house	is pleasing	to me.
NOM	3rd Sg. please	DAT-1st Sg.

(3) a. Spanish (OVS):

(A mí)	*me*	*gusta*	*la casa.*
DAT-To me	DAT CL-1st Sg.	pleases	NOM-the house

b. Spanish (SVO):

La casa	*me*	*gusta*	*(a mí).*
NOM-The house	DAT CL-1st Sg.	pleases	DAT-to me

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

When producing a sentence like (1) or interpreting sentences like (3a) or (3b), English speakers apply the so-called *first noun strategy* to map syntactic roles to noun phrases. They assign the subject role to the first noun (in preverbal position) and the object role to the second noun (in postverbal position), incorrectly rendering (1) as **(Yo) gusto las casas*, **(Yo) me gusto las casas*, or **Yo me gustan las casas*, or parsing (3b) as *"The house likes me."

Third, the verb *gustar* needs to agree with the subject (the liked thing) in number (*la casa* or *las casas*), which increases the complexity of the structure given its low communicative value (DeKeyser, 2005). Fourth, when the experiencer is a noun (rather than a dropped pronoun), it must be obligatorily preceded by the case-marking preposition *a* ("to") and accompanied by a redundant indirect object pronoun (e.g., *le*, "to him") as shown in (4a) and (4b). A similar process applies when the experiencers are two nouns, in which case the redundant indirect object pronoun must be plural (*les*, "to them") and the case-marking preposition *a* ("to") must also precede the second noun, as shown in (4c).

(4) a.	*A John*	*le*	*gusta*	*la casa.*
	DAT-to John	DAT CL-3rd Sg.	3rd Sg.	NOM-the house
	("To John, the house pleases to him.")			
b.	*A John*	*le*	*gustan*	*las casas.*
	DAT-to John	DAT CL-3rd Sg.	3rd Pl.	NOM-the houses
	("To John, the houses please to him.")			
c.	*A John y a Marta*	*les*	*gusta*	*la casa.*
	DAT-to John	DAT CL-3rd Pl.	3rd Sg.	NOM-the house
	("To John and to Marta, the house pleases to them.")			

From a L2 processing standpoint, morphological elements such as the case-marking preposition *a* and the redundant indirect object pronouns *le* and *les* in (4a–c) are difficult to perceive because both are minimally salient—they are unstressed, monosyllabic, and likely to appear after a pause (N. Ellis, 2016). As for production, L2 learners have to master the functionally complex Spanish clitic system, which marks for case, number, and gender simultaneously, an example of syntactic allomorphy that has been posited to increase the cognitive complexity of a structure (DeKeyser, 2005). Indeed, learners often produce (4b) incorrectly as **A John les gustan las casas* or **A John las gustan las casas* (Sanz, 1999).

Finally, from a pedagogical perspective, *gustar* structures are also complex. The explanations (or "rules") typically provided in Spanish L2 textbooks and classroom grammar sessions tend to be elaborate (with many concepts, steps, or subrules); they include a high amount of metalanguage and terms with a high degree of conceptual abstractness, and they are small in scope, as they apply to only a group of verbs (Dietz, 2002; Py, 1999).

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

As can be seen, Spanish *gustar* structures pose multiple problems for L1 English speakers from a developmental, psycholinguistic, linguistic, and pedagogical perspective, all of which add to their overall learning difficulty (Housen & Simoens, 2016). Furthermore, we posit here that not all Spanish *gustar* structures are equally complex and that they can be divided into at least four types depending on the number of steps that L1 English speakers need to take to process or produce them:

- Type 1 (at least three steps): *Gustar* structures in which the experiencer is a first-person pronoun (singular *me* or plural *nos*) or second-person singular pronoun (*te*). Learners need to (a) process the thing liked as the subject (e.g., *las casas*, "the houses"), (b) make the experiencer the indirect object (e.g., *me* "to me"), and (c) conjugate the verb in agreement with the thing(s) liked (e.g., plural *gustan* "please"). Optionally, they may also have to (d) insert or decode a pleonastic dative pronoun preceded by a redundant preposition (e.g., *a mí*, "to me").
- Type 2 (at least four steps): *Gustar* structures in which the experiencer is a third-person pronoun (*él*, *ella*, *ellos*, or *ellas*). In addition to the first three processes (a–c) in Type 1 and optionally (d), learners must (e) process or produce the dative pronouns *le* and *les*, which are typically confused with the direct object pronouns *lo*, *la*, *los*, or *las* or the reflexive *se*, due to their similarity and low saliency.
- Type 3 (five steps): *Gustar* structures in which the experiencer is one noun. In addition to the first three (a–c) and fifth (e) processes in Type 2, learners have to obligatorily perform (d).
- Type 4 (six steps): *Gustar* structures in which the experiencers are two or more nouns. In addition to the five processes in Type 3, learners must repeat (d) for the second noun.

RESEARCH QUESTIONS

In light of the previous review, the present study asked the following research questions:

1. Does the type of instruction (guided induction vs. deductive instruction vs. control) have differential effects on the development of a structurally and cognitively complex L2 structure (Spanish *gustar* structures)? If so, do these effects hold after 2 weeks?
2. What cognitive processes are employed while processing a complex L2 form during guided induction in an online environment?

METHOD

Participants

Participants were 70 (initially 129) college-level students (38 female and 32 male) of beginning Spanish who met the following three

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

criteria: (a) they scored zero on the production pretests; (b) they reported no exposure to the targeted structure during the study; and (c) they completed all experimental sessions. Before taking part in the study, participants had received 10 weeks (approximately 25 hr) of formal exposure to Spanish under a communicative approach. The course textbook used was *Vistazos*, third edition (McGraw-Hill).

Treatment

Educational Video Game. Developed in JavaScript, *The Gustar Maze* used guided induction to gradually promote deeper processing of the target structures. Based on Leow's (2015) model of instructed SLA, *The Gustar Maze* sought to reap the reported benefits of *task-essentialness* (to focus learners' attention on the target form; Loschky & Bley-Vroman, 1993), *scaffolding* (to gradually introduce the complexity of the form; Ludwig-Hardman & Dunlap, 2003), *right/wrong feedback* (to trigger hypothesis and rule formulation; Cerezo, 2016; Rosa & Leow, 2004), and *guiding questions* (to help learners infer rules; Vogel et al., 2011). *The Gustar Maze* was initially pilot-tested among 14 participants, who reacted very positively overall and helped identify some issues (e.g., user-friendliness and glitches) that were subsequently fixed.

On starting the game, participants were greeted by a pop-up bubble asking, "*¿Cómo se dice en español* [How do you say in Spanish] I like the house*?*" The screen then displayed two options, *yo* (I) and *a mí* (to me). Participants selected one, received corrective feedback, and were presented with the next two options to complete the sentence. The treatment consisted of 20 items like this, sequentially presented according to the four types of *gustar* structures discussed earlier, across four video game levels (see Table 1).

If the learners' choice was right, they accumulated money and moved on to select new sentence constituents. If their choice was wrong, they

Table 1. Treatment items in *The Gustar Maze*

Video game level	Structure type	Type of experiencer	Number of items
1	1	1st sg. pronoun (*me*, "I")	(6)
2		2nd sg. pronoun (*te*, "you")	(2)
		1st pl. pronoun (*nos*, "we")	(2)
3	2	3rd person pronoun (*le*, "s/he")	(4)
4	3	One noun	(3)
	4	Two nouns	(3)

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

lost money and returned to a previous video game level. In addition to "right/wrong feedback" (Cerezo, 2016), guiding questions—such as, "Oops, seems like *yo* doesn't work with *gustar*"; "Wow, did this sentence follow the previous pattern?"; and "Did you see what took place between *A mí me gusta la casa* [I like the house] and *A mí me gustan las casas* [I like the houses]?"—were provided. Figure 1 provides a video game screenshot.

After successfully completing a video game level, the screen displayed a list of correct and incorrect rules for the *gustar* item types covered so far. Participants were asked to select those that they thought were correct and were transported to the next level regardless of their accuracy. Once they completed all 20 treatment items, participants exited the maze and were welcomed by a series of fireworks. Total dollars won and time spent appeared in a corner of the computer screen.

Instructional Conditions. Intact classes were randomly assigned to one of two experimental groups varying in the type of instruction—guided induction (GI, $n = 24$) and deductive instruction (DI, $n = 26$)—or to an uninstructed control group (control, $n = 20$). In the DI group the teacher covered all 20 exemplars of *gustar* described in the Educational Video Game section, in the same order. He first asked participants (in Spanish), "How do you say in Spanish 'I like the house'?" and invited participants to provide an answer. He then wrote on the blackboard *(A mí) me gusta la casa*; translated the sentence literally ("to me [emphatic] to me pleases the house"); explained that it was not the English subject "I" that was the subject of the verb *gustar* but, rather, the thing liked ("the house"); and asked participants to pay close attention to the verb agreement. As he introduced the other



Figure 1. A screen shot of one treatment item in the video game.

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

exemplars of *gustar* in the same fashion, he highlighted orally and visually on the blackboard the differences between the English and Spanish equivalents and the potential processing problems of this verb evidenced by English-speaking learners of Spanish. Questions were encouraged and responded to and corrective feedback was provided when necessary. In turn, the GI group played *The Gustar Maze* video game, whereas the control group simply performed the assessment tasks without any formal exposure to the targeted structure.[3]

Assessment Tasks

Assessment tasks included two productive tasks (controlled oral and written production), which measured participants' ability to produce orally and in writing the targeted *gustar* structure, and one receptive task (multiple-choice written recognition). These tasks likely measured explicit knowledge, given their controlled nature (Spada & Tomita, 2010) and our pedagogical treatments, which were designed to promote explicit learning (Leow, 2015). All pretests had six old items (identical to the ones presented in the treatment). Additionally, the immediate and delayed posttests had six new items, which included different experiencers and "liked things." The instruction for the controlled oral production task read as follows:

> In each of the items below, people are expressing their liking of something. Using the words below in each item, how would you say each item orally in full sentences? Speak loudly and clearly into the microphone. Example: (1) Tú/gustar/las casas.

The instruction for the controlled written production task read:

> Write the Spanish equivalents of the following sentences in English. Use the verb *gustar*. Example: "John and Mary like Chinese food."

The instruction for the multiple-choice recognition task read:

> This is a speed test! Select the letter that best describes each sentence as quickly as you can. Make sure you read through all the options and do your best at selecting the right answer quickly! PLEASE DO NOT RETURN TO A PREVIOUS ITEM, THAT IS, JUST DO THE ITEMS QUICKLY ONE AT A TIME WITHOUT CHECKING PREVIOUS ONES AS THIS WILL AFFECT YOUR SPEED. GO! Example: "She likes the houses" (a. *A ella gustan las casas*; b. *Ella gusta las casas*; c. *A ella le gustan las casas*; d. *Ella le gusta las casas*; e. None of the above).

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

Procedure

Participants reported to the language laboratory and read and signed the institutional review board form, which informed them that they would be assigned to either a classroom or the language laboratory to receive online or offline exposure to certain Spanish grammatical items. They then completed the pretest. As explained previously, intact classes were assigned to the three instructional conditions. To address the second research question and control for reactivity (i.e., whether thinking aloud impacted subsequent performance), about half of the participants in the GI group were also requested to think aloud while playing *The Gustar Maze*, after a training session with a mathematical problem (see online Appendix A). The sessions for the DI and GI groups lasted about 20 and 15 min, respectively. Directly after, participants performed an unannounced immediate posttest. The control group simply performed the immediate posttest near the end of the session, which involved activities without *gustar*. An unannounced delayed posttest was administered 2 weeks later, together with a debriefing questionnaire to address potential data contamination during the entire experimental period. For all assessment stages, the order of tests was as follows: (a) controlled oral production, (b) controlled written production, and (c) multiple-choice written recognition. To minimize formal exposure to the target form, all references to *gustar* were removed from the syllabus and homework assignments, and instructors were requested not to entertain questions on it during the experimental period.

Scoring

All items on the assessment tasks were conservatively scored one point or zero points based on whether or not they were fully accurate and were coded as old (included in the treatment) or new items. The maximum scores were 6 on the pretests and 12 on the immediate and delayed posttests. The analysis of the think-aloud protocols is presented in the next section.

Analysis

Reliability coefficients on the immediate posttests, computed using Cronbach's alpha, were high (0.97, 0.97, and 0.92 for controlled oral production, controlled written production, and multiple-choice written recognition, respectively), proving the validity of the assessment tests.

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

Separate *t* tests on the scores from the immediate posttests revealed no significant difference between the think-aloud and non-think-aloud GI groups on either controlled oral production, $t(22) = .74$, $p = .468$, controlled written production, $t(22) = .39$, $p = .701$, or multiple-choice written recognition, $t(22) = .51$, $p = .618$, indicating no reactivity effects.

To answer the first research question, which asked whether the type of instruction had differential effects on the development of Spanish *gustar*, separate repeated-measures ANOVAs were performed per type of item (old, new, and all items), entering type of instruction and time as the between- and within-subject factors, respectively. Because the pretests included old items only, achievement data were analyzed in a two-step process: a pretest-posttest-delayed posttest analysis of old items and a posttest–delayed posttest analysis of new and all items. Effect sizes (partial eta-squared, η_p^2) and observed power (*OP*) were calculated. Effect sizes around .01, .06, and .14 were considered small, medium, and large, respectively. An observed power of .8 was considered acceptable. The alpha level for all analyses of significance was set at .05.

To answer the second research question, which addressed the cognitive processes employed by the GI group, two coders transcribed the participants' think-aloud protocols, highlighting instances of hypothesis formulation, rule formulation, and activation of prior knowledge, as defined previously, together with instances of metacognition, defined as any comment describing the participants' feelings about their progress. Interrater reliability was 94.5%, and for the remaining 5.5% an agreement was reached. For each verbal protocol, the total number of words, the number of words used to verbalize cognitive processes, and the number of instances of each type of cognitive process at each of the four levels of the video game were tallied. Online Appendix A contains a participant's full protocol. The following list provides examples of cognitive processes from several participants:

- Hypothesis formulation: "*A mí. A mí me. A mí* and *me* mean the same thing. Yes they both mean I" (Diana).
- Rule formulation: "Because *gustar* agrees to the thing being liked" (Tom).
- Activation of recent prior knowledge: "*A ella* I feel like we already did this one" (Sally).
- Metacognition: "I'm getting the hang of it" (Alex).

RESULTS

Table 2 provides descriptive statistics of accuracy scores by type of assessment task, testing stage, item, and instruction. Although separate statistical analyses were performed for old, new, and all items, for brevity's sake the following sections report only the differences

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

Table 2. Descriptive statistics: All groups, tests, and items

	Pretest mean (*SD*)			Posttest mean (*SD*)			Delayed mean (*SD*)		
Source	Old	New	All	Old	New	All	Old	New	All
Controlled Oral Production									
GI (*n* = 24)	0.00	–	–	4.88	5.04	10.00	4.38	4.29	8.71
	(0.00)			(1.57)	(1.65)	(3.04)	(2.24)	(2.16)	(4.29)
DI (*n* = 26)	0.00	–	–	3.73	3.85	7.58	2.15	2.42	4.58
	(0.00)			(2.52)	(2.51)	(5.01)	(2.48)	(2.61)	(5.05)
Control	0.00	–	–	0.10	0.15	0.25	0.85	0.70	1.55
(*n* = 20)	(0.00)			(0.30)	(0.37)	(0.55)	(1.09)	(0.87)	(1.47)
Controlled Written Production									
GI (*n* = 24)	0.00	–	–	5.42	5.54	10.96	4.96	4.83	9.79
	(0.00)			(0.93)	(0.78)	(1.54)	(1.80)	(1.68)	(3.43)
DI (*n* = 26)	0.00	–	–	3.62	3.62	7.23	2.42	2.35	4.77
	(0.00)			(2.17)	(2.47)	(4.52)	(1.94)	(2.04)	(3.94)
Control	0.00	–	–	0.05	0.05	0.10	0.50	0.55	1.10
(*n* = 20)	(0.00)			(0.22)	(0.22)	(0.31)	(0.51)	(0.76)	(1.16)
Written Recognition									
GI (*n* = 24)	1.75	–	–	5.38	5.46	10.83	5.04	4.96	10.00
	(1.07)			(0.88)	(0.83)	(1.55)	(1.46)	(1.52)	(2.93)
DI (*n* = 26)	1.50	–	–	5.19	4.92	10.12	4.73	5.00	9.73
	(1.36)			(1.33)	(1.79)	(2.80)	(2.13)	(2.19)	(4.21)
Control	1.65	–	–	1.55	1.35	2.90	2.70	2.80	5.55
(*n* = 20)	(1.35)			(0.94)	(0.58)	(1.25)	(1.49)	(1.61)	(3.00)

between them. A full report of statistical analyses is included in online Appendices B–D.

Controlled Oral Production

The ANOVA performed on the controlled oral production scores (old items) yielded a significant main effect for time, $F(2, 134) = 89.76$, $p = .000$, $\eta_p^2 = .57$, $OP = 1.00$, a significant main effect for type of instruction, $F(2, 67) = 32.81$, $p = .000$, $\eta_p^2 = .49$, $OP = 1.00$, and a significant interaction between time and type of instruction, $F(4, 134) = 20.75$, $p = .000$, $\eta_p^2 = .38$, $OP = 1.00$. These results are best interpreted in conjunction with their visual representation in Figure 2.

The significant main effect for time indicated that, overall, participants experienced learning of the target form. Separate within-subject ANOVAs revealed that both experimental groups, GI and DI, experienced statistically significant learning from pretest to immediate posttest.

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

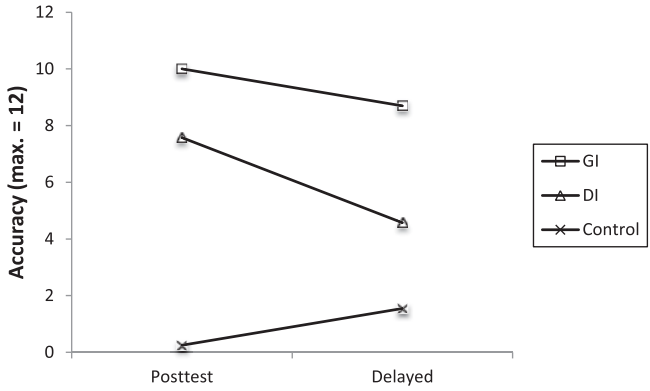


Figure 2. Controlled oral production accuracy by group (old items only).

The GI group increased 4.87 items, $F(1, 23) = 231.67$, $p = .000$, $\eta_p^2 = .91$, $OP = 1.00$, and the DI group increased 3.73 items, $F(1, 25) = 56.86$, $p = .000$, $\eta_p^2 = .69$, $OP = 1.00$. However, whereas the GI group retained learning from the immediate to the delayed posttest, with a nonsignificant loss of .50 items, $F(1, 23) = 1.38$, $p = .252$, $\eta_p^2 = .06$, $OP = .20$, the DI group experienced a statistically significant decrease of 1.58 items, $F(1, 25) = 12.40$, $p = .002$, $\eta_p^2 = .33$, $OP = .92$. In turn, the control group did not improve significantly from the pretest to the immediate posttest, $F(1, 19) = 2.11$, $p = .163$, $\eta_p^2 = .10$, $OP = .28$, but it did experience statistically significant learning of .75 items from the immediate to the delayed posttest, $F(1, 19) = 12.04$, $p = .003$, $\eta_p^2 = .34$, $OP = .91$.

The significant interaction between time and type of instruction indicated that the groups' learning trajectories were statistically different. A post hoc Scheffé test indicated that at the immediate posttest both experimental groups outperformed the control group (GI: $M = 4.78$, $SE = .55$, $p = .000$; and DI: $M = 3.63$, $SE = .54$, $p = .000$). The GI group was the top scorer, but it did not significantly outperform the DI group ($M = 1.14$, $SE = .51$, $p = .088$). At the delayed posttest, however, only the GI group outperformed the control group (GI: $M = 3.53$, $SE = .63$, $p = .000$; and DI: $M = 1.30$, $SE = .62$, $p = .118$). Additionally, the GI group outperformed the DI group ($M = 2.22$, $SE = .59$, $p = .002$).

The preceding statistical procedure was repeated for new and all items (Figure 3) at the immediate to delayed posttest interval, yielding the same statistically significant differences, with a couple of exceptions (see online Appendix B). First, the DI group did outperform the control group on delayed oral production of both new items ($M = 1.72$, $SE = .62$, $p = .026$) and all items ($M = 3.03$, $SE = 1.21$, $p = .049$). And second, the GI group's edge over the DI group approached significance on the immediate oral production of all items ($M = 2.42$, $SE = 1.01$, $p = .062$).

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

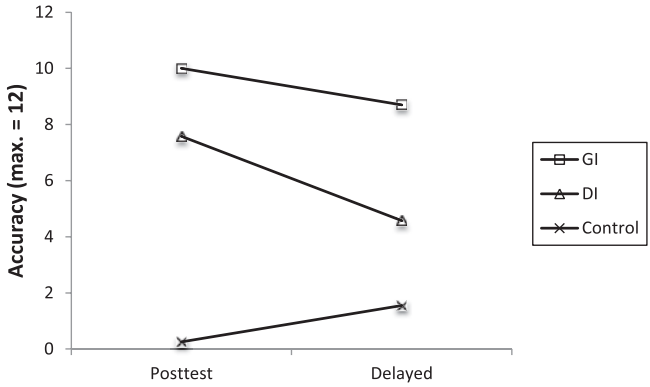


Figure 3. Controlled oral production accuracy by group (all items).

Controlled Written Production

Similarly to oral production, the ANOVA on the controlled written production scores (old items) yielded a significant main effect for time, $F(2, 134) = 142.36$, $p = .000$, $\eta_p^2 = .68$, $OP = 1.00$, a significant main effect for type of instruction, $F(2, 67) = 85.74$, $p = .000$, $\eta_p^2 = .72$, $OP = 1.00$, and a significant interaction between time and type of instruction, $F(4, 134) = 36.30$, $p = .000$, $\eta_p^2 = .52$, $OP = 1.00$. Figure 4 provides a visual representation.

Separate within-subject ANOVAs revealed that both experimental groups improved significantly from the pretest to the immediate posttest. The GI group increased 5.42 items, $F(1, 23) = 816.59$, $p = .000$, $\eta_p^2 = .97$, $OP = 1.00$, and the DI group increased 3.61 items, $F(1, 25) = 71.91$,

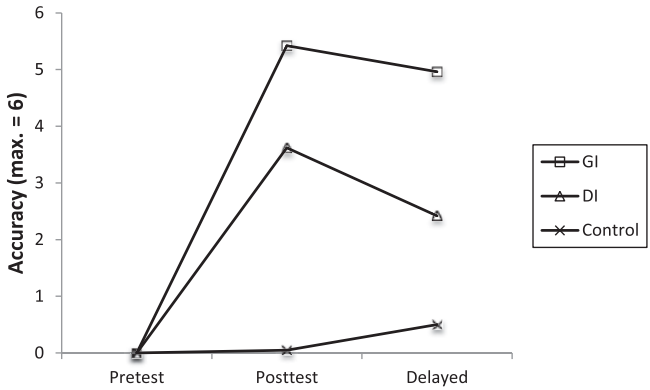


Figure 4. Controlled written production accuracy by group (old items only).

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

p = .000, η_p^2 = .74, *OP* = 1.00. However, whereas the GI group retained learning from the immediate to the delayed posttest, with a nonsignificant loss of .46 items, $F(1, 23)$ = 1.04, p = .319, η_p^2 = .04, *OP* = .16, the DI group experienced a statistically significant decrease of 1.19 items, $F(1, 25)$ = 9.62, p = .005, η_p^2 = .28. *OP* = .85. The control group did not improve significantly from the pretest to the immediate posttest, $F(1, 19)$ = 1.00, p = .330, η_p^2 = .05, *OP* = .16, but it did experience a statistically significant increase of .45 items from the immediate to the delayed posttest, $F(1, 19)$ = 11.07, p = .004, η_p^2 = .37, *OP* = .88.

Post hoc Scheffé tests showed that the experimental groups outperformed the control group, both immediately after the treatment (GI: M = 5.37, SE = .44, p = .000; and DI: M = 3.57, SE =.43, p = .000) and two weeks later (GI: M = 4.46, SE = .49, p = .000; and DI: M = 1.92, SE = .48, p = .001). Additionally, the GI group outperformed the DI group at both the immediate posttest (GI: M = 1.80, SE = .41, p = .000) and the delayed posttest (GI: M = 2.54, SE = .46, p = .000).

Replication of the preceding statistical procedure for new and all items (Figure 5) at the immediate to delayed posttest interval yielded exactly the same statistically significant differences (see online Appendix C).

Multiple-Choice Written Recognition

The ANOVA on the multiple-choice written recognition scores (old items) yielded a significant main effect for time, $F(2, 134)$ = 125.39, p = .000, η_p^2 = .65, *OP* = 1.00, a significant main effect for type of instruction, $F(2, 67)$ = 23.25, p = .000, η_p^2 = .41, *OP* = 1.00, and a significant interaction between time and type of instruction, $F(4, 134)$ = 22.82, p = .000, η_p^2 = .40, *OP* = 1.00. Figure 6 provides a visual representation.

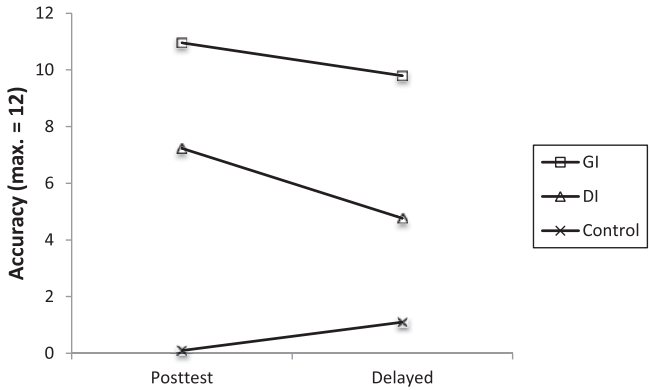


Figure 5. Controlled written production accuracy by group (all items).

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

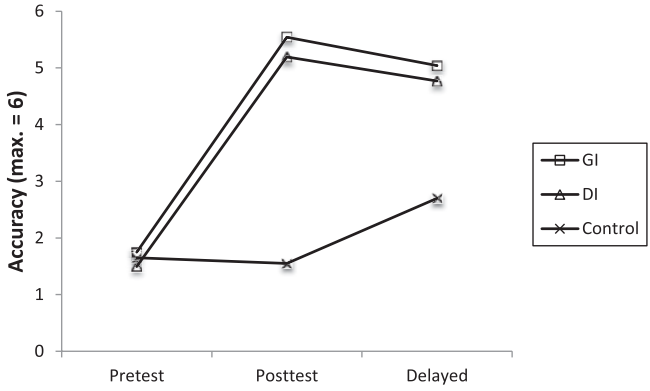


Figure 6. Multiple-choice written recognition accuracy by group (old items only).

Separate within-subject ANOVAs revealed that both experimental groups improved significantly from pretest to immediate posttest. The GI group increased 3.62 items, $F(1, 23) = 166.27$, $p = .000$, $\eta_p^2 = .88$, $OP = 1.00$, and the DI group gained 3.69 items, $F(1, 25) = 139.47$, $p = .000$, $\eta_p^2 = .85$, $OP = 1.00$. Both the GI and DI groups retained learning from the immediate to the delayed posttest, $F(1, 23) = .94$, $p = .343$, $\eta_p^2 = .04$, $OP = .15$, and $F(1, 25) = 1.43$, $p = .242$, $\eta_p^2 = .05$, $OP = .21$, respectively. The control group did not improve significantly from pretest to posttest, $F(1, 19) = .24$, $p = .629$, $\eta_p^2 = .01$, $OP = .07$, but did gain 1.15 items from the immediate to the delayed posttest, which was statistically significant, $F(1, 19) = 22.29$, $p = .000$, $\eta_p^2 = .54$, $OP = .99$.

Post hoc Scheffé tests showed that the experimental groups outperformed the control group, both immediately after the treatment (GI: $M = 3.83$, $SE = .33$, $p = .000$; and DI: $M = 3.64$, $SE = .32$, $p = .000$) and two weeks later (GI: $M = 2.34$, $SE = .53$, $p = .000$; and DI: $M = 2.03$, $SE = .52$, $p = .001$). However, both experimental groups performed comparably on both posttests, immediately after the treatment ($M = .18$, $SE = .31$, $p = .83$) and two weeks later ($M = .31$, $SE = .49$, $p = .82$).

Replication of the preceding statistical procedures for new and all items (see Figure 7) at the immediate-delayed posttest interval yielded the same statistically significant differences (see online Appendix D).

Cognitive Processes during the GI Treatment

Participants verbalized, on average, 1,245.8 words (min. = 240; max. = 2733; $SD = 734.42$). Of those, they used 222 words to verbalize the cognitive processes in our coding protocol, which conservatively

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

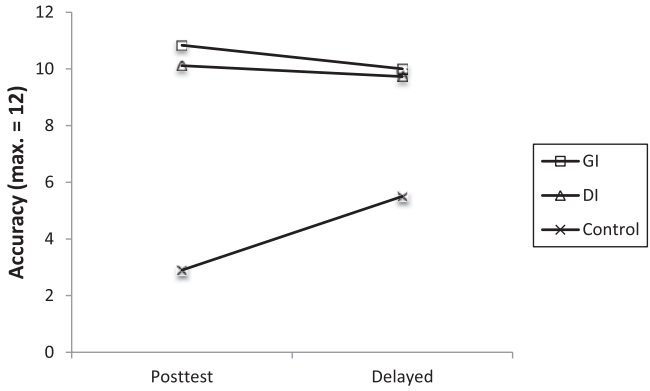


Figure 7. Multiple-choice written recognition accuracy by group (all items).

excluded potential instances of noticing and reporting while participants read aloud the practice items and feedback (min. = 0; max. = 457; *SD* = 174). This represented 15.49% of total talk time (min. = 0%; max. = 29.3%; *SD* = 10.91%; see Figure 8 for individual participant percentages). On average, participants verbalized the cognitive processes in our protocol 20.8 times during the treatment (min. = 0; max. = 44; *SD* = 15.39; see Figure 9 for individual participant counts). Only 2 participants out of 10 did not report any of the cognitive processes in our protocol.[4] The remaining 8 were able to reach the highest level of awareness (rule formulation) 17.75 times on average (min. = 6; max. = 34; *SD* = 10.4). As Figure 10 shows, for all 10 participants the most frequently verbalized cognitive process was rule formulation, with 14.2 average counts (68.27%), followed by metacognition

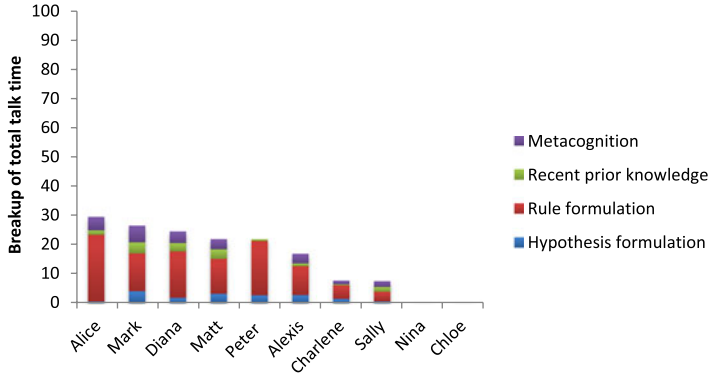


Figure 8. Percentage of words used to verbalize cognitive processes by participant.

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

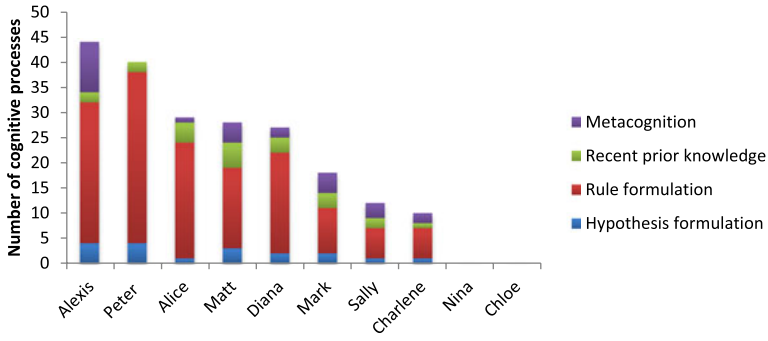


Figure 9. Number of verbalized cognitive processes by participant.

(2.6 counts, 12.5%), activation of recent prior knowledge (2.2 counts, 10.58%), and hypothesis formulation (1.8 counts, 8.65%). This breakup of cognitive processes was homogenous across all four levels of the video game, although instances of metacognition seemed to diminish gradually.

DISCUSSION

The first research question asked whether type of instruction (guided induction vs. deductive instruction vs. control group) would have differential effects on the development of complex Spanish *gustar* structures. The answer to this question is affirmative for our productive assessment tasks (controlled oral and written production) and negative for our receptive task (multiple-choice written recognition). As noted

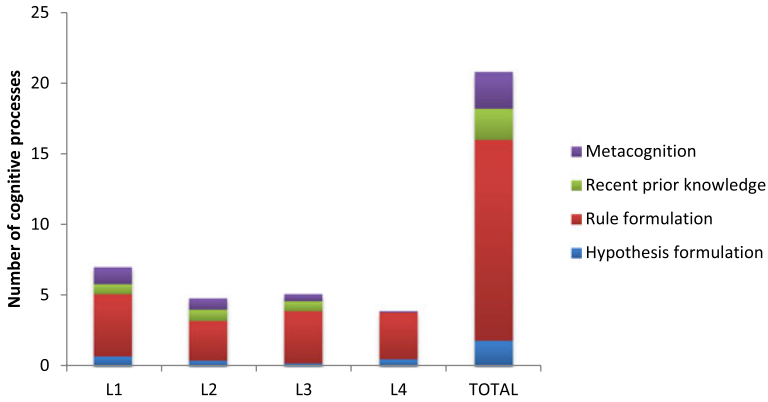


Figure 10. Number of verbalized cognitive processes by level of the video game.

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

earlier, our tasks likely measured our participants' explicit knowledge of the targeted structure, given their controlled nature and the explicit orientation of our treatments.

Experimental groups GI and DI both outperformed the control group on all posttests (save the DI group on delayed oral production of old items). Additionally, GI outperformed DI on all controlled production posttests (save immediate oral production, in which it approached statistical significance for all items). This edge increased over time. For written production, GI's edge over DI went from 3.73 items (31% of 12) on the immediate posttest to 5.02 items (42%) 2 weeks later. For oral production, it went from 2.42 items (20%) to 4.13 items (34%). This was because whereas GI retained learning, DI experienced significant decreases in retention (3 items, 25%, on oral production, and 2.46 items, 20%, on written production). A closer look at the written and oral posttests revealed that those decreases were localized in two sources: largely Type 4 items (the double noun structure, particularly the insertion of the *a* dative marker before the second noun) and to a lesser extent the two *nos* items at Type 1. These findings are remarkable, given that both experimental groups practiced with the same 20 treatment items and DI actually spent 5 min more on the items than the average GI participant. As discussed subsequently, GI may have promoted deeper cognitive processing than DI, helping learners internalize a greater portion of intake in their developing systems (Leow, 2015), particularly for the more difficult types of structures.

Our results are much in line with previous studies, in which guided induction (more than deductive instruction) helped learners develop various linguistic structures in classroom contexts, both immediately and up to 2 weeks later (e.g., Haight et al., 2007; Herron & Tomasello, 1992; Smart, 2014; Vogel et al., 2011). However, our study further illustrates that guided induction can prove effective in fully online learning, in which teacher mediation is not possible.

Our experimental groups, however, performed comparably on all multiple-choice recognition posttests. This is consistent with Flynn's (1986) point that receptive tasks are less challenging than productive tasks. In other words, the depth of processing elicited by DI may have been enough to parallel GI in accurately recognizing *gustar* structures. This is further illustrated by the larger delayed gains of our control group for multiple-choice recognition (compared to the production tests), as explained subsequently.

The control group improved statistically—though minimally—from the immediate to the delayed posttests (1.3 items on oral production, 1 on written production, and 2.6 items on written recognition). A post hoc analysis revealed that these gains were experienced in the Type 1 structures of *gustar* (specifically with pronouns *me* and *te*), but not in the types posited to be more advanced. Possible causes

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

include potential (but not statistically significant) test effects after the immediate posttests and some potential exposure to the target structure in the participants' textbook and workbook despite their reports to the contrary.

With the mentioned exceptions, our results were consistent for old, new, and all items. One possible interpretation of these results is that our item types were not fundamentally different. In other words, introducing new experiencers and liked objects in *gustar* structures did not pose an added difficulty to our participants, who were able to generalize their recently gained explicit knowledge to the new items.[5]

In sum, our results for the first research question show that guided induction can successfully promote explicit learning of a structurally and cognitively complex grammatical structure in a fully online environment. Furthermore, they indicate that guided induction may have an edge over deductive instruction in more cognitively taxing assessment tasks (production over recognition) and for more complex structures or substructures, as evidenced by the GI group's retention versus the DI group's decrease of Type 4 *gustar* substructures in production tasks.

The second research question asked which cognitive processes were employed during the learning of complex Spanish *gustar* structures by the GI group. An analysis of this group's think-aloud protocols revealed that participants verbalized awareness at the level of understanding (hypothesis and rule formulation), activation of recent prior knowledge, and metacognition during 15.49% of their talk time. Although this may seem a small percentage, it should be noted that we did not code as cognitive processes the participants' read-alouds of the video game script and feedback, which may well overlap with instances of noticing and reporting of the target structure. In fact, 8 out of 10 participants reported verbalizations of the correct rules, on average 17.75 times and ranging from 6 to 34 times. As is discussed subsequently, rule formulation was indeed the most frequently verbalized cognitive process, on average 14.2 times (68.27%), followed by metacognition (12.5%), activation of recent prior knowledge (10.58%), and hypothesis formulation (8.65%).

The high percentage of rule formulation (68.27%) contrasts sharply with that in Toth et al.'s (2013) study, in which the participants only formulated the rules during 20% and 14% of their talk time in small-group and whole-class interaction, respectively. Those participants spent most of their time patterning (44.4% and 49.9% in small-group and whole-class interaction, respectively), as opposed to the participants in our study, who showed hypothesis formulation only 8.65% of the time. One likely cause of this divergence is the different type of task in both studies. Much less controlled, the task in Toth et al. (2013) involved reading texts to find and compare instances of the targeted structure (Spanish *se* structures) and counterexamples (personal sentences).

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

In contrast, the item-by-item translation nature of our task and the recurrent feedback with guiding questions helped our participants quickly notice the gap in their interlanguage, allowing them to start formulating rules earlier. Additionally, our video game promoted a more individualized learning experience, with feedback and guiding questions that were in line with each learner's needs.

Online Appendix A, which includes a coded full verbal protocol (reported in Leow, 2015), provides an insightful account of the cognitive processes employed by one participant in our GI group, Mark. As can be observed in the protocol, Mark's initial production attempts demonstrated the difficulty of our target structure. His initial inaccurate responses were clearly driven by L1 processing. However, with the help of guiding questions, he soon began to process the L2 data deeper by noticing mismatches, making hypotheses, showing evidence of rule formulation, activating recently learned knowledge of exemplars, and displaying metacognition. Mark's deeper processing led him to reach awareness at the level of understanding of several target rules, overriding his initial hypotheses and readjusting his subsequent productions. This protocol thus appears to provide empirical support for Leow's (2015) model at the intake processing stage. At the same time, it illustrates that guided induction that is theoretically driven can be successfully used to promote deeper processing and potential development of structurally and cognitively complex linguistic structures in an online learning platform.

CONCLUSIONS, LIMITATIONS, AND FURTHER RESEARCH

Our study showed that whereas both traditional deductive instruction in a classroom setting and guided induction via a video game effectively promoted the development of explicit knowledge of the complex *gustar* structures, the latter had an edge, particularly on productive assessment measures that were more cognitively taxing and for the more difficult aspects of the target structure. This was likely due to the specific nature of our guided induction task, which promoted deeper cognitive processing, and to the individualized learning that our video game was able to promote.

This study, however, is not without limitations. First, as pointed earlier, our experimental groups were guided by ecological considerations. As a result, we did not explicitly control for learner attention during the DI treatment and conflated type of instruction with instructional setting. Baralt's (2013) research suggests that the effects of a pedagogical intervention (e.g., task complexity) may be moderated by the instructional setting, with more complex and cognitively demanding tasks working better face-to-face than online. Thus, future studies may want to control for learner attention in DI treatments and compare our two

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

types of instruction under the same setting, or address the interaction between type of instruction (DI vs. GI) and type of setting (face-to-face vs. online). Second, we only controlled for learners' prior knowledge of the target structure. Future studies must therefore control for cognitive and psychosocial individual differences. Third, although the same 20 treatment items were practiced in both experimental conditions and the GI group took five fewer minutes to complete the treatment, the potential backtracking nature of the video game did allow a (minimal) number of participants to practice again with previously encountered items. Future studies should investigate the moderating role of amount of practice, in line with Cerezo (2016). Fourth, given the relatively similar level of processing reported by our participants, we did not address whether there was a correlation between depth of processing and learning outcomes. Future studies on larger populations of learners should be best suited to do so. Lastly, we only targeted one grammatical structure with four posited levels of complexity. Future studies may want to target various grammatical structures or various types of L2 features (e.g., lexical, morphological, or syntactic) in the same design, to investigate more clearly how the type of instruction may be moderated by the type or complexity of L2 features.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0272263116000139

Received 1 April 2015
Accepted 4 February 2016
Final Version Received 10 February 2016

NOTES

1. According to Toth et al. (2013, p. 4) "such tasks constitute guided induction, as they engage learner agency in analysing L2 grammar while providing support through task materials and teacher feedback."

2. "Intake . . . is that part of the input that has been attended to by second language learners while processing the input. Intake represents stored linguistic data that may be used for immediate recognition and does not necessarily imply language acquisition" (Leow, 1993, p. 334).

3. In comparing these experimental groups, we were driven by ecological considerations (the prevalent type of classroom-based instruction and the affordances of video games). However, as pointed out by one reviewer, by conflating type of instruction (deductive instruction or guided induction) with instructional setting (face-to-face or online learning), we introduced learners' attention as a potential confound. Nonetheless, it is noteworthy that both the DI and GI groups performed not only statistically similarly on the recognition assessment task but also at more than 90% accuracy, which suggests that the DI group was indeed paying high rates of attention. The differential performances on the production tests could be attributed to *how* participants processed the target data.

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

4. Nevertheless, both participants spent the average amount of time on task and, like the non-think-aloud group, also performed very well on all the immediate and delayed posttests.

5. One reviewer suggested that new items should have been verbs similar to *gustar*, such as *interesar*, *encantar*, and so on. This point is well taken, though at the same time we did not want to introduce into the study additional information that would link *gustar* to other verbs.

REFERENCES

Adair-Hauck, B., Donato, R., & Cumo-Johanssen, P. (2010). Using a story-based approach to teach grammar. In J. L. Shrum & E. W. Glisan (Eds.), *Teacher's handbook: Contextualized foreign language instruction* (4th ed., pp. 216–244). Boston, MA: Heinle, Cengage Learning.

Ayoun, D. (2001). The role of negative and positive feedback in the second language acquisition of the passé composé and imparfait. *Modern Language Journal*, *85*, 226–243.

Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition*, *35*, 689–725.

Brooks, F. B., & Donato, R. (1994). Vygotskyan approaches to understanding foreign language learner discourse during communicative tasks. *Hispania*, *77*, 262–274.

Carroll, S. (2001). *Input and evidence: The raw material of second language acquisition*. Amsterdam, the Netherlands: John Benjamins.

Cerezo, L. (2016). Type and amount of input-based practice in CALI: The revelations of a triangulated research design. *Language Learning & Technology*, *20*, 100–123.

Collins, L., Trofimovich, P., White, J., & Horst, M. (2009). Some input on the easy/difficult grammar question. *The Modern Language Journal*, *93*, 336–353.

Craik, F. I. M., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, *17*(3), 379–410.

DeKeyser, R. (2005). What makes second-language grammar difficult? A review of issues. *Language Learning*, *55*, 1–25.

Dietz, G. (2002). On rule complexity: A structural approach. *EUROSLA Yearbook*, *2*, 263–286.

Elder, C., & Manwaring, D. (2004). The relationship between metalinguistic knowledge and learning outcomes among undergraduate students of Chinese. *Language Awareness*, *13*, 145–162.

Ellis, N. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology*, *5*(3), 289–318.

Ellis, N. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*(2), 143–188.

Ellis, N. (2016). Salience, Cognition, Language Complexity, and Complex Adaptive Systems. *Studies in Second Language Acquisition*, *38*(2), 341–351.

Erlam, R. (2003). Evaluating the relative effectiveness of structured-input and output-based instruction in foreign language learning. *Studies in Second Language Acquisition*, *25*(4), 559–582.

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*(2), 147–171.

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, *14*(3), 393–417.

Flynn, S. (1986). Production vs. comprehension: Differences in underlying competences. *Studies in Second Language Acquisition*, *8*, 135–164.

Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443–482). Amsterdam, the Netherlands: John Benjamins.

Haight, C. E., Herron, C., & Cole, S. P. (2007). The effects of deductive and guided inductive instructional approaches on the learning of grammar in the elementary foreign language college classroom. *Foreign Language Annals*, *40*(2), 288–310.

Hall, C. (1998). Overcoming the grammar deficit: The role of information technology in teaching German grammar to undergraduates. *Canadian Modern Language Review*, *55*(1), 41–60.

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press

Herron, C., & Tomasello, M. (1992). Acquiring grammatical structures by guided induction. *The French Review*, *65*(5), 708–718.
Housen, A. (2014). Difficulty and complexity of language features and second language instruction. In A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Malden, MA: John Wiley & Sons. doi:10.1002/9781405198431.wbeal1443.
Housen, A., & Simoens, H. (2016). Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition*, *38*(2), 163–175.
Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86.
Leow, R. P. (1993). To simplify or not to simplify: A look at intake *Studies in Second Language Acquisition*, *15*, 333–355.
Leow, R. P. (2012). Explicit and implicit learning in the L2 classroom: What does the research suggest? *The European Journal of Applied Linguistics and TEFL*, *2*, 117–129.
Leow, R. P. (2015). *Explicit learning in the L2 Classroom: A student-centered approach*. New York, NY: Routledge.
Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes & S. M. Gass (Eds.), *Task and language learning: Integrating theory and practice* (pp. 123–167). Clevedon, UK: Multilingual Matters.
Ludwig-Hardman, S., & Dunlap, J. C. (2003). Learner support services for online students: Scaffolding for success. *International Review of Research in Open and Distance Learning*, *4*(1), 1–15.
Montrul, S. (1997). Spanish gustar psych verbs and the unaccusative se construction: The case of dative experiencers in SLA. In A. G. Pérez-Leroux & W. R. Gass (Ed.), *Contemporary perspectives on the acquisition of Spanish: Vol. I. Developing grammars* (pp. 189–207). Somerville, MA: Cascadilla Press.
Oded, B., & Walters, J. (2001). Deeper processing for better EFL reading comprehension. *System*, *29*(3), 357–370.
Py, B. (1999). Enseignement, apprentissage et simplification de la langue. In J. Billiez (Ed.), *De la didactique des langues à la didactique du plurilinguisme* (pp. 145–151). Grenoble, France: Presses de l'Université Stendhal.
Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition*, *18*(1), 27–67.
Rosa, E. M., & Leow, R. P. (2004). Awareness, different learning conditions, and second language development. *Applied Psycholinguistics*, *25*(2), 269–292.
Rosa, E., & O'Neill, M. D. (1999). Explicitness, intake, and the issue of awareness. *Studies in Second Language Acquisition*, *21*(4), 511–556.
Sanz, C. (1999). What form to focus on? Linguistics, language awareness, and the education of L2 teachers. In J. F. Lee & A. Valdman (Eds.), *Form and meaning: Multiple perspectives* (pp. 3–23). Boston, MA: Heinle & Heinle.
Shaffer, C. (1989). A comparison of inductive and deductive approaches to teaching foreign languages. *The Modern Language Journal*, *73*(4), 395–403.
Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL 26*(2), 184–201.
Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*, 263–308.
Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*(2), 183–203.
Toth, P. D., Wagner, E., & Moranski, K. (2013). "Co-constructing" explicit L2 knowledge with high school Spanish learners through guided induction. *Applied Linguistics*, *34*, 279–303.
Vogel, S., Herron, C., Cole, S. P., & York, H. (2011). Effectiveness of a guided inductive versus a deductive approach on the learning of grammar in the intermediate level college French classroom. *Foreign Language Annals*, *44*(2), 353–380.
Wagner, E., & Toth, P. (2013). Building explicit L2 Spanish knowledge through guided induction in small-group and whole-class interaction. In K. McDonough & A. Mackey (Eds.), *Second language interaction in diverse educational contexts* (pp. 89–108). Amsterdam, the Netherlands: John Benjamins.

https://doi.org/10.1017/S0272263116000139 Published online by Cambridge University Press