



RESEARCH ARTICLE

Optimizing the input for learning of L2-specific constructions: The roles of Zipfian and balanced input, explicit rules and working memory

Manuel F. Pulido 

Department of Spanish, Italian and Portuguese, The Pennsylvania State University, University Park, 442 Burrowes Building, 16802, USA
Email: mpulido@psu.edu

(Received 02 December 2022; Revised 26 December 2023; Accepted 08 January 2024)

Abstract

Usage-based theory has proposed that learning of linguistic constructions is facilitated by input that contains few high-frequency exemplars, in what is known as a skewed (or Zipfian) input distribution. Early empirical work provided support to this idea, but subsequent L2 research has provided mixed findings. However, previous approaches have not explored the impact that cognitive traits (e.g., working memory) have on the effectiveness of skewed or balanced input. The experiment reported here tested learners' ability to develop new L2 categories of adjectives that guide lexical selection in Spanish verbs of "becoming." The results showed that, when explicit rules are provided, low-working memory learners benefitted from reduced variability in skewed input, while high-working memory individuals benefitted from balanced input, which better allows for rule-based hypothesis testing. The findings help clarify the mixed findings in previous studies and suggest a way forward for optimizing the L2 input based on individual traits.

Introduction

Usage-based approaches propose that, to learn a language, speakers must accumulate experience of regular linguistic patterns through exposure to the input (Bybee, 2008; Ellis & Wulff, 2014; Ellis et al., 2013; Tomasello, 2005). But what work in recent years has highlighted is that exposure to just any form of input may not be sufficient for successful language learning. Importantly, previous research has reported that the ability to learn new constructions from the input may be facilitated by specific frequency distributions, such as low variability input distributions that present a few prototypical lexical exemplars with very high frequency. In fact, such low variability (or "skewed") input distributions are pervasive in natural language (Zipf's law; e.g., Lavi-Rotbain & Arnon, 2023) and are believed to facilitate novel pattern detection during language acquisition (Goldberg et al., 2004). It has been proposed that constructions are learnable in part thanks to the presence of highly prototypical exemplars, which facilitate the formation

of new categories and the integration of less frequent non-prototypical exemplars (e.g., Abbot-Smith & Tomasello, 2006; Bybee & Eddington, 2006; Ibbotson et al., 2012; Goldberg, 2013). In contrast, highly variable (or “balanced”) input, in which multiple lexical items are equally frequent, is believed to make pattern detection more difficult.

Support for the superiority of skewed input comes from early work with children and adults by Goldberg and colleagues. In a series of experiments that investigated the learning of a novel construction,¹ learning was found to be significantly improved when linguistic input presented a skewed distribution (Goldberg et al., 2004; 2007; Casenhiser & Goldberg, 2005). More specifically, in a study with adult native English speakers—which extended previous work with children (Casenhiser & Goldberg, 2005)—Goldberg and colleagues (2007) found that skewed input significantly improved learning when the same exemplars were presented repeatedly at the beginning of exposure (this was called “skewed first” input), relative to input in which frequencies were randomly skewed or balanced across lexical items. Therefore, later studies that have investigated skewed input have typically employed “skewed first”—rather than “skewed random”—input.²

However, subsequent L2 research on the effectiveness of skewed input has produced mixed results. Several subsequent experiments with L2 learners reported either no differences between skewed and balanced input or even an advantage for balanced input (Nakamura, 2012; McDonough & Nekrasova-Becker, 2014; McDonough & Trofimovich, 2013; Year & Gordon, 2009). In an experiment that investigated Thai speakers’ acquisition of the English double-object construction, McDonough and Nekrasova-Becker (2014) found a significant advantage for learners exposed to balanced, rather than skewed, input. The authors proposed that the failure to find an advantage for skewed input in L2 learners might be due to differences in the mechanisms engaged by adults during categorization and particularly their tendency towards explicit learning strategies and rule searching.

Specifically, rule-based learning allows individuals to engage in hypothesis testing upon each encounter with new stimuli—a strategy that may be most productive when input is more variable, as is the case with balanced, rather than skewed, input. This idea was supported by McDonough and Trofimovich (2013), who found that when explicit rules were provided (i.e., “deductive” exposure), only balanced input led to significant learning; in contrast, in the absence of rules (“inductive” exposure), balanced input led to below-chance accuracy and was inferior to skewed input. Therefore, some previous results suggest that learners’ ability to engage in hypothesis testing deeply influences the effectiveness of skewed or balanced input.

This idea is relatable to work that has found an important effect of individual learner cognitive resources under explicit, but not implicit, learning conditions. Previous data have shown that, even when different learners are exposed to the same input, individual differences in working memory (WM) are associated with variability in how exemplars become categorized (DeCaro et al., 2008), as well as variability in linguistic performance (Navarro et al., 2022). More specifically, within the literature on L2 acquisition, a considerable body of evidence has found that individual WM capacity is a significant

¹These studies used a novel “appearance” construction, in which a first noun phrase expresses the grammatical subject and a second noun phrase expresses the location. For example, the sentence “the rabbit the hat mooped” describes a rabbit appearing on a hat. Input distributions were manipulated based on the variability of the nonce-verbs used (e.g., “mooped”).

²Unless otherwise specified, in the rest of the text the bare term “skewed” is used to refer to “skewed first” input.

predictor of learning in explicit conditions (Reber et al., 1991; Robinson, 2005; Tagarelli et al., 2011; 2015). Given this, exploring the contribution of individual differences provides a promising avenue to clarify the mixed findings regarding the effect of exposure to input that is skewed or balanced.

The notion that specific input conditions may unequally benefit different learners, based on individual aptitude-based differences such as WM, is in line with research on so-called aptitude-by-treatment interactions (ATI) within second language acquisition (DeKeyser, 2012; Robinson, 2001). The ATI approach is based on the idea that the effectiveness of language learning depends on the interaction between an individual's skill and the type of treatment they receive (DeKeyser, 2021; Skehan, 1989; Vatz et al., 2013).³ Therefore, a primary goal focuses on identifying and matching individual learners' aptitudes with learning conditions that may enhance language learning outcomes. In this logic, it is conceivable—or, even, expected—that the effectiveness of a particular type of treatment (skewed or balanced input) may be modulated by individual learner traits.

Very limited previous work within the ATI framework has considered how individual differences might modulate the effect of skewed or balanced input, and none have—to the best of the author's knowledge—considered the role of individual WM. In a study that considered the role of individual learner traits during exposure to skewed and balanced input, Brooks and colleagues (2017) examined the potential effect of nonverbal intelligence on the acquisition of Russian morphology. Based on previous work in which the authors examined the effect of input variability (by manipulating the number of new vocabulary items presented, rather than input skewness; Brooks & Kempe, 2013), they predicted learners would benefit from balanced input only if they had higher nonverbal intelligence (as measured by section of the Culture Fair Intelligence Test). Contrary to their predictions, no significant interactions were found between skewed/balanced input and their measure of nonverbal intelligence on the acquisition of morphology.

So far, Brooks and colleagues' study appears to be the only previous work that explicitly examined the role of individual differences during exposure to skewed or balanced input, with no previous studies having considered the role of WM in this regard. That is, despite that fact that well-documented variability in learning outcomes has been associated with differences in WM capacity, the contribution of individual differences to learning from input that is skewed or balanced has not been adequately explored.

The limited current understanding of this issue limits L2 research and risks disfavoring at least a portion of learners for whom specific input properties may not be most adequate. Taking an ATI approach, the goal of the present study is to investigate how individual differences in cognitive skill—specifically in WM—modulate learners' ability to form categories from the input. The experiment reported below capitalizes on a well-studied set of verbs in Spanish, termed verbs of “becoming,” to investigate learners' ability to learn new L2 categories under different input conditions. The following section reviews prior evidence on how variability in WM, as a central cognitive resource for processing information, may modulate individuals' ability to learn from different types of input.

³In a different, but somewhat related, vein the “desirable difficulties” approach also considers how learning gains may be impacted by interactions between individual learner traits and training difficulty; see, for example, Bjork & Kroll (2015), Suzuki et al. (2019), and Pulido & Dussias (2020).

Alternative learning pathways: the role of WM and verbalizable rules in learning

The extant neuropsychological theory has underscored that category learning (as other aspects of human cognition) is characterized by the existence of more than one system with the ability to perform a similar function. In particular, previous work has pointed at competition between explicit verbal learning (i.e., when learning is guided by verbalizable rules) and implicit learning systems (Ashby et al., 1998; Green et al., 2006). There is good evidence that the explicit and implicit learning systems recruit different cognitive resources and that these systems are in fact neurally dissociable (Chandrasekaran et al., 2014).

Importantly, previous work in cognitive psychology and in L2 learning indicates that differences in cognitive profile (e.g., in WM) may modulate the effectiveness of specific pathways engaged during input processing and category learning. In particular, tasks that can be conducted by relying on explicit information depend on WM resources. WM is believed to play a particularly important role in rule-based (“deductive”) learning: conditions in which rules are available and thus encourage explicit hypothesis testing during processing (McDonough & Trofimovich, 2013) have been found to place greater demands on WM resources (Tagarelli et al., 2015, Waldron & Ashby, 2001). In contrast, implicit information-integration learning is thought to rely on the procedural learning system and functions independently of WM span (Ashby et al., 1998; Chandrasekaran et al., 2014).⁴

Some valuable insight on the engagement of WM comes from studies in cognitive science that have focused on learning of nonlinguistic stimuli, as well as learning of (semi)artificial languages, to examine the role of individual differences under controlled learning conditions. In an influential study on visual category learning, Waldron and Ashby (2001) found that in an explicit learning condition—in which distinctions between stimuli were easy to verbalize via an explicit rule—learning ability was predicted by individuals’ ability to rely on WM; in fact, when WM resources were overtaxed by a secondary task, learning was disrupted. In contrast, in implicit learning conditions—in which stimuli properties were difficult to verbalize via explicit rules—the level of WM load did not modulate learning ability (for further evidence of a dual-learning system see Chandrasekaran et al., 2014; Maddox et al., 2013; although for potential interactions between explicit and implicit learning, see, e.g., Forkstam & Petersson, 2005).

Similar evidence on the involvement of WM is found within the realm of language learning. In a study examining the learning of linguistic structure via explicit rules or incidentally, Tagarelli et al. (2011) employed a semi-artificial language consisting of English lexicon and German syntax; using a semi-artificial language allowed the authors to investigate learning of verb order in the absence of lexical learning demands. Tagarelli and colleagues hypothesized that learning performance would be correlated with WM, to the extent that learners relied on a rule to engage in explicit hypothesis testing; however, they also hypothesized that WM should not influence learning in a condition with no explicit rules, which favored implicit information-integration learning. To test this idea, they exposed participants to conditions in which learners could either rely on a verbalizable rule or not, while examining the contribution of WM in

⁴WM span is a relevant measure in that it reflects a higher ability to temporarily store and retrieve information (e.g., words), in the face of interference from another task (e.g., considering an explicit rule or solving arithmetic operations, as in the operation-span task; see the section “Individual Differences Measures” for further details).

each condition. Their results showed that only when a verbal rule was available—enabling hypothesis-testing during input exposure—did high-WM individuals achieve higher learning rates than those with lower WM.

Interestingly, work on category learning has also suggested that individuals with higher WM are not always at a learning advantage when acquiring the defining aspects of a new category. Some insightful evidence comes from the role of WM during learning of nonlinguistic categories, under exposure to verbalizable or non-verbalizable patterns (e.g., when visual categories are defined by cues such as shape, color). Indeed, in contexts in which learners cannot rely on verbalizable rules, and therefore require engaging in bottom-up categorization, individuals with *lower* WM abilities have been found to outperform higher-WM individuals (DeCaro et al., 2008). More generally, conditions that tax WM have been found to impair rule-based learning, while they tend to leave implicit category learning unaffected (e.g., Markman et al., 2006; Waldron & Ashby, 2001).

The acknowledgement of variability in cognitive pathways engaged during learning across individuals has important theoretical and practical ramifications for research on learning (Green et al., 2006). The studies reviewed above suggest that individual learners might benefit best from exposure to input conditions that encourage the optimal input-processing strategy based on their individual abilities (for a similar approach see, e.g., Perrachione et al., 2011). In particular, these findings indicate that considering WM capacity may allow to identify and match the optimal learning conditions for individuals with different levels of cognitive skill, in line with aptitude-by-treatment interaction (ATI) approaches in L2 acquisition (see also, e.g., Indrarathne & Kormos, 2018; Perrachione et al., 2011, or the studies in DeKeyser, 2021).

As noted above, however, almost no previous studies have examined the role of individual differences during exposure to skewed or balanced input (with the exception of Brooks et al., 2017), and no previous work has considered the role of WM on the ability to learn from different input distributions. Given the significant role of WM in previous studies that investigated explicit L2 learning conditions, the experiment reported below seeks to clarify the mixed evidence on the influence of skewed and balanced input by taking into account diversity in WM capacity among the learner population.

Present study

The goal of the present study is to investigate how individual differences in cognitive resources (specifically in WM span) modulate the effectiveness of skewed and balanced input, by examining learning in deductive (rule-based) and inductive (without explicit rules) learning conditions. This rationale is in line with an ATI approach (e.g., DeKeyser, 2012; 2021; Robinson, 2001). That is, given that learners differ in their internal resources available during input processing, it should not be assumed that the same input will engage the same processing strategies across different learners.

Taking direction from the evidence reviewed above, firstly, the present study manipulated the type of INPUT DISTRIBUTION. I hypothesized that, by facilitating the learning of high-frequency exemplars, skewed input distributions may alleviate learners' cognitive burden. In contrast, balanced input—which is, by definition, more highly variable—would pose a higher burden on WM. As such, balanced input may be beneficial for learners with relatively higher WM resources, while skewed input would be more beneficial for lower-WM learners.

Secondly, to create conditions that differed in their level of demand on WM, the present study manipulated the INSTRUCTION TYPE. Following McDonough and Trofimovich (2013), the present study employed deductive and inductive learning conditions, which differed on whether learners were provided with an explicit rule that guided their learning prior to being exposed to the input (“deductive” learning) or whether no rule was provided to guide input processing (“inductive learning”).⁵ In deductive learning, participants are better able to engage in hypothesis testing by testing an explicit rule upon each encounter with new stimuli. As reviewed above, previous work has indicated that deductive hypothesis testing places higher demands on WM and leads to learning outcomes that are predicted by individual WM span (DeCaro et al., 2008; Tagarelli et al., 2011).

The two variables (INPUT DISTRIBUTION and INSTRUCTION TYPE) will allow to explore the conditions that may facilitate (or hinder) learning for individuals of different cognitive profiles (see the detailed hypotheses and predictions described in “Hypotheses and Predictions”).

To investigate the influence of these variables on learners’ ability to learn new L2 categories, the experiment reported below capitalizes on a well-studied set of verbs in Spanish, termed verbs of “becoming.” Spanish verbs of becoming, which comprise a number of quasi-synonymous verbs including *ponerse*, *quedarse*, or *volverse* (all functionally equivalent to “become”), are a case in point for categorization that involves L1–L2 incongruent constructions. Because usage of the specialized verbs of becoming relies on knowledge of different semantic categories of adjectives, speakers of languages such as English must develop new categories into which to group known adjectives.

Analogical extension has been identified as a reliable rule for successful categorization (Bybee & Eddington, 2006). In other words, a prototypical adjective often becomes the basis for generalization when using a specific [verb + adjective] construction, such as [*ponerse* + adjective], or [*quedarse* + adjective]. In a study of L1 speaker corpus and similarity judgments, Bybee and Eddington (2006) showed that adjectives used in discourse with each verb of becoming—for example, *ponerse* and *volverse*—formed well-defined semantic categories around a high-frequency, prototypical adjective. For example, a prototypical verb–adjective combination such as *ponerse* + *nervioso* (“nervous”) accounted for verb choices with less frequent and even innovative adjectives. Given a prototypical adjective for a specific verb (*ponerse* + *nervioso*), less frequent and even innovative combinations (*histérico* “hysterical,” *agresivo* “aggressive,” or *furioso* “furious”) were readily accounted for through semantic similarity, as confirmed by semantic association networks based on native speaker semantic judgements.

Of the verbs of becoming, two verbs—namely, *volverse* and *ponerse*—are selected here, as they form categories that are best defined based on semantic clusters (Bybee &

⁵The inductive/deductive employed here, modelled after McDonough and colleagues’ work, has some differences and similarities to incidental/intentional, implicit/explicit conditions in other studies. For example, Tagarelli et al. (2015) employed “incidental and intentional conditions ... designed to promote the acquisition of implicit and explicit knowledge, respectively” (227). Their incidental learning condition instructed participants to repeat stimuli, but did not announce that they would be tested. In contrast, their intentional condition provided no rule, but warned participants that they would be tested (this would be most comparable to the inductive learning condition employed here). Rather than taking this approach, the present study capitalized on the inductive/deductive distinction used in previous studies on skewed input effects, which are more directly relevant to the goals of the study.

Eddington, 2006; Eddington, 1999). Previous work has highlighted the difficulty of formulating rules to guide categorization of Spanish verbs of becoming. Verb selection based on rules that describe, for example, slow/fast change or passive/active change, is somewhat overlapping and not clear-cut (Bybee & Eddington, 2006; Eddington, 1999; Ibarretxe-Antuñano & Cheikh-Khamis, 2019), making acquisition a notable challenge for L2 speakers. Instead, previous usage-based work that examined actual usage in corpus data (Brown & Cortés-Torres, 2012; Bybee & Eddington, 2006) suggests that a prototype-based rule may be most effective at guiding learners in the formation of semantic adjective clusters. Therefore, learners are expected to successfully categorize exemplars when directed by an explicit rule indicating that categories are built around a prototype (i.e., “*ponerse* is used with words related in meaning to the word ‘nervous,’ and *quedarse* with words related to the word ‘still’”). Mirroring the frequencies of natural corpus data, the skewed condition employed here presented with higher frequency two adjectives (i.e., *nervioso*, “nervous” and *quieto* “still”) that have been identified as central, given their high frequency of use with each verb in previous work (Bybee & Eddington, 2006). However, in order to develop *semantic categories* (i.e., groups of semantically related adjectives) structured around a prototypical member (e.g., *ponerse nervioso*), it is not clear whether, and to what degree, variability is beneficial.

Following previous work discussed above (e.g., Goldberg et al., 2007; McDonough & Trofimovich, 2013), the present study employed a paradigm consisting of exposure followed by generalization. That is, to learn the new linguistic categories, participants are first exposed to a small set of items under different experimental learning conditions. Subsequently, they are asked to make generalizations on a new set of untrained items based on the recently learned categories.

Thus, to summarize, the present experiment was based on a 2×2 design. Learners were exposed to one of two types of INPUT DISTRIBUTION: skewed input, in which a prototypical adjective was presented with higher frequency for each verb; or balanced input, in which all adjectives associated with each verb appeared with equal frequency. Additionally, training was manipulated based on INSTRUCTION TYPE: inductive learning, for which no explicit rules were provided; or deductive learning, for which an explicit verbal rule encouraged hypothesis testing during exposure. Finally, the effect of individual WM on learning under each context was examined.

Research questions

Guided by the findings reviewed above, this study examined the following research questions.

RQ 1: How does skewed input, which provides high-frequency exposure to a prototypical adjective, facilitate the learning of the categories of the verbs of “becoming” constructions?

RQ 2: How does the deductive instruction type, in which a verbal rule is explicitly provided, modulate the effect of skewed or balanced input?

RQ 3: How does individual WM capacity modulate the effects of input distribution and instruction type?

Hypotheses and predictions

The experiment reported below will allow to test three different hypotheses and their associated predictions grounded in the previous literature:

Skewed input hypothesis

This hypothesis would predict a main effect of input distribution, such that skewed input facilitates acquisition of constructions by maximizing exposure to a highly frequent “pathbreaker” item.

Previous mixed findings suggest that, even within this view, it must be assumed that there is variability in skewed input benefitting some learners, while not having an effect in others. In this scenario, skewed input may still bear an advantage even if only for some learners. For example, *at the group level*, skewed input may provide stability in the transmission of linguistic structures by allowing all speakers to acquire regular patterns in the input with no trade-off (e.g., Goldberg et al., 2007; Lavi-Rotbain & Arnon, 2022), while no advantage for balanced input would be predicted. This hypothesis remains agnostic toward the potential effect of instruction type and WM.

Rule-Testing hypothesis

This hypothesis would predict that, rather than seeing a benefit of skewed input, balanced input should be beneficial for L2 learners, particularly in deductive learning conditions.

Previous L2 studies with adult learners have proposed that balanced input is more beneficial because it provides more variability for hypothesis testing. This view is supported by work that found an advantage of balanced input when combined with an explicitly stated rule (McDonough & Trofimovich, 2013; McDonough & Nekrasova-Becker, 2014). No specific prediction is made for WM on the basis of this hypothesis, although a general benefit of higher WM would be expected.

Available cognitive resources hypothesis

This hypothesis attempts to reconcile previous findings by predicting that WM will modulate the effect of input distributions: Higher-WM learners may benefit from balanced input in deductive conditions (rule-testing hypothesis), whereas lower-WM learners will benefit from skewed input.

The view proposed here is that both hypotheses above are true under specific circumstances. Taking direction from evidence in the adult categorization literature and in L2 learning (e.g., DeCaro et al., 2008; Tagarelli et al., 2011), it is hypothesized that skewed input may be more beneficial when cognitive resources are burdened or limited. This may be due to learner-internal causes (individual learners inherently have lower cognitive resources available, e.g., lower WM) or due to learner-external causes (cognitive load is high, e.g., during explicit hypothesis testing in deductive learning). In conditions that promote hypothesis testing, balanced input may be beneficial but *only* as long as learners’ cognitive resources are on par with task demands. In other words, this view makes the specific prediction that skewed input should be more beneficial than balanced input at baseline (inductive learning) but that in more cognitively demanding learning conditions (deductive learning), skewed input will mainly benefit low-WM learners, while high-WM learners will benefit from balanced input. In concrete terms, an interaction between WM, input distribution and instruction type would be expected.

Methodology

Participants

Monolingual native speakers of English (N=115) were recruited through Prolific (www.prolific.co) to participate in an online experiment implemented on Gorilla (www.gorilla.sc; Anwyl-Irvine et al., 2019). Five participants were excluded because they reported more than minimal knowledge of Spanish in their responses to a language history questionnaire (i.e., self-ratings greater than the minimum value of 1 in a scale from 1 to 7). Nine additional participants were excluded due to inattentiveness during the training phase (accuracy < 80%), and 17 were excluded because they failed to adequately learn the verb-adjective pairings after training exposure (test accuracy < 70%). The final pool for analysis consisted of 82 participants (29 female, 52 male, 1 not specified), distributed across inductive skewed (N=21), inductive balanced (N=18), deductive skewed (N=25), and deductive balanced (N=18); the total exclusion rates were roughly evenly distributed across inductive balanced (27.36%), inductive skewed (29.3%), deductive balanced (26.33%), and deductive skewed (27.36%). Twenty-two participants reported low-proficiency knowledge of other languages (French, German, Japanese, Chinese, Gaelic, Welsh), which do not mirror the verb-based categories targeted here (see Table 1 for a summary of proficiency). All provided informed consent and received compensation at a rate of 10 USD/hour. As part of a larger project, participants in the inductive learning condition were invited to complete a second session, which will be reported elsewhere.

Individual differences measures

Participants completed a series of individual differences measures to characterize their linguistic profile and WM span. Basic demographic information was available through Prolific. The tasks are described in this section and the results are summarized in Table 1 (by-condition information is provided in Table S1). The materials of the individual differences tasks are available at https://osf.io/5dcx8/?view_only=1a1fc062f4fd4286884f29b9d05c3e94.

Language history questionnaire

A language history questionnaire was used to gather data on relevant aspects of individuals' linguistic background, including experience and proficiency with any additional languages. Basic demographic data (age and gender) was obtained from Prolific.

Picture naming

The picture naming task was employed to measure participants' individual lexical ability (based on Gollan et al., 2008) and was employed to control for individual

Table 1. Summary of individual differences measures

Measure	M	SD	Range	CI
Age (in years)	31.85	6.81	20–45	[30.35, 33.36]
English (L1) picture naming (/1)	0.97	0.04	0.81–1	[0.95, 0.98]
WM (/60)	50.89	7.50	17–60	[49.24, 52.54]
Other languages (self-reported /7)	2.26	1.06	1–3.75	[1.78, 2.74]

differences in vocabulary skill, by including it as a covariate in the analysis. The task took approximately five minutes. Participants saw 80 black and white line drawings, which were presented one at a time in randomized order. In each trial, participants were asked to type the name for the image displayed in a text box. The first nine trials were used as practice. The accuracy of typed responses was coded off-line; up to one character permutation or misspelling was allowed.

Operation span

Participants completed the Operation Span (O-Span) task as a measure of WM. In this task participants were presented with simple math problems alongside a potential solution and were asked to indicate whether the solution was correct or not. The O-Span task was employed given its widespread use in the language learning literature (e.g., Linck & Weiss, 2015; Tagarelli et al., 2011; 2015), and it relies to a lesser extent on verbal skills than other tasks such as the reading span.⁶ Following each operation, an English word appeared on the screen, which participants should memorize for later recall at the end of the trial. There were three trials at each set size ranging from a two to a six word span. The individual score of memory span was calculated as the total number of words correctly recalled. Completing the task took approximately 10 minutes.

Materials

A list of 50 Spanish adjectives were selected based on previous corpus work reported by Bybee and Eddington (2006), with each half of the list consisting of adjectives, respectively, associated with *ponerse* and *quedarse*. Data from the Corpus del Español (web/dialects section, containing over two billion words) confirmed that the selected adjectives were used preeminently with their corresponding target verb (mean proportion of use with target verb: 89.5%, SD: 12.37). Given that the goal was to examine the reorganization of existing concepts into new categories, the use of English translations allowed learners to fully understand the meaning of adjectives, without the considerable burden of learning the meaning of 50 items in a new language. Adjectives were thus translated into English (e.g., *caliente* “hot,” *nervioso* “nervioso,” *muerto* “dead,” *delgado* “slim,” etc.) with contextualized searches of verb+adjective using Linguee (linguee.com), a machine learning translation tool based on a corpus of paired bilingual texts, and were checked with two online bilingual dictionaries (the *Collins Spanish Dictionary* and *WordReference*, which are combined on one same site, wordreference.com). From the list of 50 adjectives, 10 adjectives were selected for training (half associated with each verb) and the remaining 40 were not trained but were reserved for the generalization task.

Two lists of training materials—one for each of the two conditions—were created based on the subset of training items. In the list for the balanced input condition, each

⁶As noted by an anonymous reviewer, there are a number of possible operationalizations of WM; for instance, in addition to the O-Span, Tagarelli and colleagues (2011; 2015) also employed a letter-number ordering task (for a review of WM in SLA, see, e.g., Li, 2022; Linck et al., 2014).

Because the WM task involves a verbal component via lexical recall, it is expected that this measure might be weakly correlated with L1 lexical skill. This was consistent with a correlational analysis performed between WM (word recall) and English picture naming, showing a reliable but weak association ($r(80) = 0.28$, $p = 0.01$). As indicated, the picture naming scores were included in the analysis reported below as a covariate, allowing to control for individual variability attributable to L1 lexical skill.

adjective was repeated nine times, for a total of 90 training trials. In the list for the skewed input condition, the prototypical adjective within each verb category (i.e., “nervous” for the verb *ponerse* and “still” for *quedarse*) was repeated 25 times each, while the rest of the adjectives were repeated 5 times each; that is, the prototypical verbs were five times more frequent than any of the other non-prototypical verbs. For the training phase, a recording of each of the two target Spanish verbs was made by the author, who is a native speaker of Spanish.

Procedure

Participants completed one session lasting approximately one hour. They first completed the English picture naming task, followed by the training task, an immediate test, and the generalization task. After this, they completed the O-Span task and language history questionnaire. Participants were assigned to one of the two types of instruction (inductive or deductive) and of input distribution (balanced or skewed).

Training task procedure

During the training task, participants saw an adjective displayed in the center of the screen, as well as each of the two verbs, one on the bottom left and the other on the bottom right corner of the screen. The position of the verbs was counterbalanced across participants. Participants were instructed to listen to the verb presented auditorily and press a button on their keyboard assigned to each verb; all verb forms were presented in the infinitive. At the beginning of each trial, the adjective and verb were displayed for 2,000 ms; this long interval was selected to encourage participants to pay attention to the trained adjective and actively anticipate the upcoming verb.⁷ The target verb was then presented auditorily, and participants pressed the button on their keyboard corresponding to the location of the target verb (“z” for left, “m” for right). The timing of the stimuli allowed time for participants to anticipate the upcoming verb; as a result, hearing the verb confirmed or disconfirmed their prediction, allowing them to learn from erroneous predictions, even without getting explicit correction (for a recent review on prediction error in L2 learning, see, e.g., Bovolenta & Marsden, 2022).⁸

Participants had to make a response to advance to the next trial; responses were disabled before the audio for the verb was played. Immediately after a response was recorded, the adjective disappeared and only the verbs remained on display for 300 ms before the next trial. The sequencing is illustrated in Figure 1. Five practice trials were presented to allow participants to familiarize themselves with the tasks; to avoid any

⁷The approach employed here focuses on explicit attention to the construction, and explored the effect of different types of exposure under carefully timed experimental conditions. This was considered an appropriate first step for a lexical-semantic construction like the Spanish verbs of becoming. The lexical nature of the construction did not require that additional linguistic context be provided, in contrast with previously tested constructions that targeted morphosyntax and required full sentences (e.g., Goldberg et al., 2007; McDonough & Nekrasova-Becker, 2014; Year & Gordon, 2009). Future work in progress will extend the present paradigm by embedding target verb-adjective pairs in sentences.

⁸Importantly, learning to apply an abstract rule requires that learners be able to relate it to specific instances. For learners exposed to explicit instructions, the rule guided the reconfiguration of adjectives into two new categories that do not exist in the L1, and therefore each exemplar provides evidence to test whether a given adjective should be paired with one verb or the other.



Figure 1. Sample trial for the adjective “nervous” and the target verb “ponerse.” The audio was played 2,000 ms after the adjective was displayed. The position of each of the two verbs on the screen (bottom left and right corners) was counterbalanced across participants but remained constant for a given participant.

descriptive adjectives that might be semantically related to the experimental materials, the proper adjectives “Chilean” and “Mexican” were employed for practice.

Importantly, different instructions were presented to participants in each condition. Participants in the inductive learning condition were simply warned that they would see and hear two verbs, both of which were equivalent to “become” but that Spanish required using one or the other based on the accompanying adjective. Participants were simply asked to do their best to learn the verb-adjective pairing in the task. This was consistent with an inductive learning approach, in which any generalizations are derived directly from a body of observations, without a priori rules that guide input processing (Motha, 2013; Seliger, 1975). In contrast, participants in the deductive learning group were explicitly presented with a “critical rule,” which indicated that *ponerse* was used with words related in meaning to the word “nervous” and *quedarse* about words related to the word “still” (this rule was directly based on the research on verbs of becoming reviewed above by Bybee and Eddington, 2006). They were also warned that they should not try to come up with other rules, as that would confuse them. The instructions text for the training, test and generalization tasks is provided in Appendix B. Completing the training took less than 10 minutes.⁹

Test procedure

Immediately after the training task, participants were asked to complete a test using the same procedure as for training but without the auditory cues guiding selection. This allowed to assess whether participants had learned the verb-adjective combinations during the previous exposure phase. Each item was presented twice in order to mitigate the possibility that a participant might correctly guess the correct answer.

Generalization task procedure

Participants were informed that they would be seeing new untrained adjectives and that they should choose the verb that they deemed most appropriate in each case. Critically, as in the training task, participants in the inductive learning group received no additional guidance, while participants in the deductive learning group were reminded of the rule described above.

The procedure of the generalization task was very similar to that of the test, with the important difference that the adjectives presented were the rest of the items that had not

⁹Despite its short, intensive exposure, the training presented learners with 90 trials of verb-adjective pairings, providing a substantial amount of exposure to verb-adjective mappings.

been previously trained, and each item was presented only once. Participants were given up to 10 seconds to make a decision; if no response was made, they were prompted to respond faster. No feedback was provided during this task and, given that it closely mirrored the training task, no practice trials were included.¹⁰ Completing the test and generalization tasks took less than five minutes.

Analysis

The results from the immediate test revealed that a majority of participants remembered the target verb-adjective pairings with high accuracy following exposure during the training phase (mean: 95.15%, SD: 7.61); there were no significant differences in the immediate test results across the skewed and balanced subgroups ($t(69.30) = -0.98, p = 0.33$). Participants who failed to reach 70% test accuracy were excluded; these were approximately evenly distributed across the exposure conditions (seven skewed, nine balanced). Because it would not be possible for learners to generalize on recently learned knowledge unless sufficient knowledge is gained during training, this threshold was set as a minimum level of learned items following exposure upon which to make generalizations. The 70% minimum was selected as it was the cut-off threshold employed in previous work using a similar paradigm to examine the effect of skewed input (McDonough & Trofimovich, 2013) and in previous work in our lab (e.g., Pulido, 2021); this threshold provides a practical compromise for accuracy higher than chance (50%) while it avoids setting an overly high threshold.

Data from the Generalization task were submitted to a generalized mixed effects regression analysis using R (R Core Team, 2023). Data and script are available at https://osf.io/5dcx8/?view_only=1a1fc062f4fd4286884f29b9d05c3e94. Response accuracy during generalization was employed as the dependent variable. The model included fixed effects for input distribution (skewed, balanced) and instruction type (inductive, deductive), as well as individual scores for test accuracy, English picture naming, and WM (O-Span) and their two-way interactions with input distribution and instruction type and a three-way interaction for WM, input distribution, and input type, to test the available cognitive resources hypothesis described above. Following attempts that resulted in convergence failure, the final random-effects structure included random intercepts for subjects and items, as well as by-item random slopes for learning type. Categorical variables were sum-contrast coded, and continuous variables were centered. Collinearity variance inflation scores were found to be acceptable (< 2.0 for all variables). Cronbach's alpha was calculated as a reliability index for the O-Span task and found to be acceptable at $\alpha = 0.86$ (i.e., within a range from 0.70 to 0.90; Streiner, 2003; Tavakol & Dennick, 2011). Because the raw O-span data had a skewed distribution (skewness = -1.52 , kurtosis = -6.74); the analysis was performed on Box-Cox transformed O-Span data (skewness = -0.81 , kurtosis = -3.47); this was employed instead of log-transformation, which had the opposite effect (skewness = -2.95 , kurtosis = -16.76). More importantly, the residuals of the regression model were normally distributed, as confirmed with the DHARMA package (Hartig, 2022).¹¹ The results of the final model are presented in Table 2.

¹⁰Due to experimental error, one item in the *ponerse* verb list was not displayed during the generalization task (i.e., the stimuli displayed contained 19 adjectives for *ponerse* and 20 for *quedarse*).

¹¹Two participants could be considered outliers based on their O-Span scores (< 2.5 SDs from the mean). However, removing these two outliers did not alter the pattern of results. The results from the full dataset are reported here for completeness.

Results

The results of the mixed-effects regression analysis (Table 2) revealed a significant effect of test scores, suggesting that higher knowledge of the exemplars presented during training increased performance in generalization. A crucial three-way interaction emerged between instruction type, WM, and input distribution ($\beta = -0.21$, $SE = 0.09$, $p = 0.02$). The interaction, which is visualized Figure 2, suggested no effect of input in the inductive learning group, and an interaction of WM and input type within the deductive learning group. Further, the results suggested that skewed input was more beneficial only for learners in the deductive group but that this effect was reversed at high levels of WM span. To follow up on this interaction, as a first step, the model was rerun using treatment contrasts. With the instruction type reference level set to inductive learning, results for both skewed and balanced input indicated no significant effects of WM or input distribution and no significant interaction (all p -values > 0.85). In comparison, a main effect of WM emerged only for the deductive balanced exposure ($\beta = 0.64$, $SE = 0.22$, $p < 0.004$). Additionally, the output for deductive learning indicated a significant interaction of WM and input distribution, suggesting that within deductive learning, WM was associated with higher accuracy for balanced input ($\beta = 0.78$, $SE = 0.28$, $p < 0.004$) but with an opposite polarity effect for skewed input ($\beta = -0.78$, $SE = 0.28$, $p < 0.004$). The effect of WM in the interaction is further illustrated in Figures 2–3 (the median split in Figure 3 is employed only for visualization; the analysis was performed on the continuous WM variable).

These results guided the follow-up comparisons with the emmeans package (reported comparisons are False Discovery Rate-corrected). For WM, comparisons were performed for ± 1 SD from the mean WM scores; this made it possible to examine the effect of WM as a continuous variable, rather than as a dichotomized factor. Post-hoc comparisons for the interaction between WM and input distribution revealed no differences within the inductive instruction type (all p -values > 0.99); in contrast, for deductive learners, comparisons indicated that lower-WM learners exposed to balanced input performed significantly worse than both higher-WM learners also exposed

Table 2. Output summary of the mixed-effects regression analysis

Fixed effects	Estimate	SE	z	p
(Intercept)	0.70	0.19	4.14	< 0.001
Input distribution	-0.05	0.09	-0.55	0.583
Instruction type	-0.15	0.11	-1.39	0.165
Test accuracy	0.20	0.09	2.18	0.030
Picture naming	-0.02	0.11	-0.22	0.828
WM	0.12	0.09	1.26	0.208
Input distribution * Instruction type	0.09	0.09	1.01	0.310
Input distribution * Picture naming	0.12	0.10	1.24	0.215
Input distribution * WM	0.19	0.10	1.94	0.052
Input distribution * Test accuracy	-0.10	0.09	-1.06	0.287
Instruction type: Picture naming	0.08	0.11	0.77	0.443
Instruction type * WM	-0.13	0.09	-1.41	0.159
Instruction type * Test accuracy	0.01	0.10	0.13	0.895
Input distribution * Instruction type * WM	-0.21	0.09	-2.31	0.021
Random effects	Variance	SD	Correlations	
Intercept Participant	0.42	0.65		
Intercept Item	0.73	0.86		
Instruction type Item	0.12	0.35	-0.99	
Conditional R^2 : 0.31; marginal R^2 : 0.04				

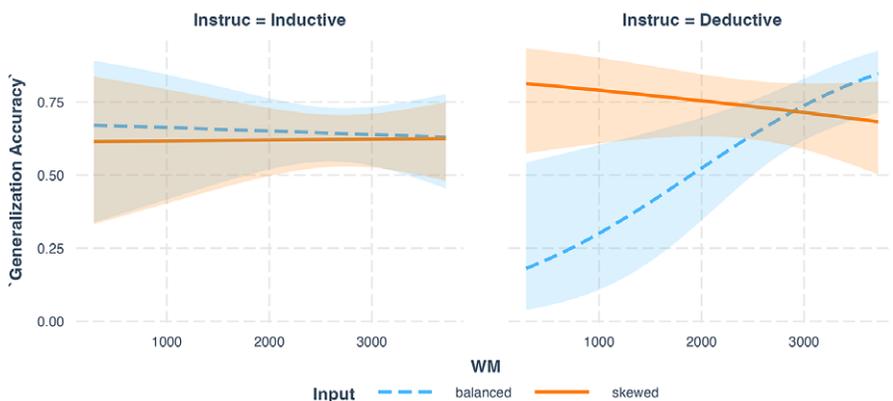


Figure 2. Interaction between input distribution, instruction type and WM based on mixed-effects model estimates. Ribbons represent 95% CIs.

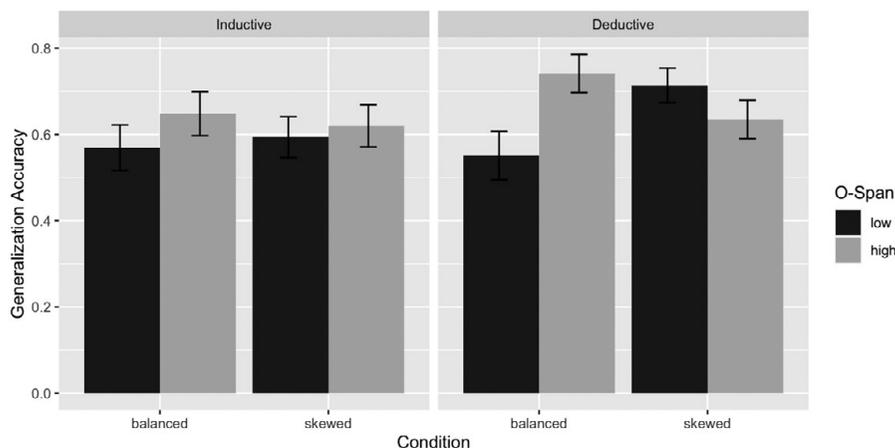


Figure 3. Accuracy in generalization by input distribution, input type, and WM. A median split of WM scores is applied here only for illustration purposes. Error bars represent 95% CIs.

to balanced input ($p = 0.021$) and *lower*-WM learners exposed to skewed input ($p = 0.037$) but was not different from *higher*-WM learners exposed to skewed input ($p = 0.234$). Finally, comparisons across instruction type at the same level of WM confirmed that, at higher-WM, balanced input led to significantly higher performance for deductive than for inductive learning ($p = 0.039$), and skewed input at lower-WM was marginally higher for deductive than inductive learning ($p = 0.074$; for the complete set of FDR-corrected comparisons, see Tables S2–S3 in the appendix).

Discussion

Previous work has produced various findings on the effect of skewed input. While some studies found it to be advantageous (e.g., Casenhiser & Goldberg, 2005; Goldberg et al., 2007; Madlener, 2016; Zhang and Dong, 2019, study 2; Zhang & Mai, 2023, study 2),

subsequent L2 studies reported either no advantage for skewed input (Nakamura, 2012; Year & Gordon, 2009; Zhang & Mai, 2023, study 1) or even an advantage for balanced input (Brooks et al., 2017; McDonough & Trofimovich, 2013; McDonough & Nekrasova-Becker, 2014; Zhang & Dong, study 1). In line with previous work that explored ATI in SLA (for a review, see, e.g., Vatz et al., 2013), this study explored the effect of individual WM capacity in modulating the effect of skewed and balanced input under inductive and deductive conditions. The experiment reported here capitalized on Spanish verbs of “becoming,” a set of verbs (approximately equivalent to “become” in English), which selectively collocate with subsets of adjectives, requiring L2 learners to develop a new categorization of adjective groups. Specifically, the materials targeted the selection of adjective subsets that collocated with two verbs (*ponerse* and *quedarse*), requiring the formation of two new semantic categories for adjectives. The experiment employed a 2 × 2 design by crossing the instruction type (deductive or inductive—that is, with or without explicit rules, respectively) and input distribution (skewed or balanced) to test the role of each factor and their potential interaction with WM.

A main goal of the study was to test the hypothesis that individual diversity in cognitive resources (specifically, in WM) mediates the effect of skewed input. Couched within evidence in the adult categorization literature and in L2 learning (e.g., DeCaro et al., 2008; Kempe et al., 2010; Tagarelli et al., 2011), I hypothesized that skewed input would facilitate learning when cognitive resources are limited—whether due to learner-internal traits (e.g., lower WM) or to learner-external causes (e.g., when engaging in real-time rule testing during input processing). On the other hand, under explicit rule-based learning conditions, balanced frequency distributions may provide optimal input to facilitate hypothesis testing but *only* as long as learners’ cognitive resources are on par with task demands. In short, the available cognitive resources hypothesis would predict that, when an explicit rule is being concurrently held in WM, skewed input will tend to benefit low-WM learners, while high-WM learners will benefit from balanced input. This hypothesis is in line with previous studies within the ATI approach, which found that lower-aptitude learners benefitted from reduced variability in the input, in contrast with higher-aptitude learners, who tended to benefit from higher variability in the input (Brooks et al., 2006; Perrachione et al., 2011). The results indeed showed ample variability across conditions based on individual WM in support of previous ATI research suggesting that learners with different aptitude levels may benefit from learning under different input conditions (e.g., Brook et al., 2006; Perrachione et al., 2011). Critically, when WM capacity was factored in, variability in subgroups was systematic and replicated previous seemingly contradictory findings in the literature. The results are discussed next in connection with the three research questions.

To recapitulate, the first research question asked whether and how skewed input (i.e., higher frequency in exposure to a prototypical adjective for each category) facilitates the formation of new L2 categories. The results revealed no main effect of input distribution, i.e., no overall effect of skewed or balanced input exposure. At face value, this result is congruent with later studies that, in contrast to earlier work (e.g., Goldberg et al., 2004; 2007), reported either no advantage for skewed input or even an advantage for balanced input (McDonough & Trofimovich, 2013; McDonough & Nekrasova-Becker, 2014; Nakamura, 2012). However, a beneficial effect of skewed input emerged precisely where predicted by the available cognitive resources hypothesis during the generalization task: After being exposed to deductive training, in which an explicit verbalizable rule was provided, lower-WM learners experienced a significant benefit from reduced variability in the skewed input, while balanced input resulted in lower accuracy for lower-WM learners. This finding supports the proposal originally

made by L1 researchers that lower variability in skewed input conditions may facilitate the development of a new category through a “pathbreaker” prototypical exemplar. However, an essential qualification must be made in what concerns adult L2 learners, where the effect was far from being present in a majority of learners.

Instead, the more common finding in later L2 studies has suggested that a combination of deductive instruction type, in which a verbal rule is explicitly provided, and balanced input might be more conducive to learning by allowing adult learners to engage in hypothesis testing (McDonough & Trofimovich, 2013). This possibility was directly addressed through the second research question, which asked whether the effect of skewed or balanced input was contingent on the provision of an explicit verbal rule during instruction. This account leads to the straightforward prediction of an interaction between input distribution and instruction type, by which balanced input conditions should become beneficial when combined with an explicit verbal rule that allows for hypothesis testing. Once again, this hypothesis was partly born out but with a critical qualification. As predicted by the cognitive resources hypothesis, the availability of an explicit rule in combination with balanced input facilitated learning only for those learners with higher WM (recall that at higher-WM, balanced input led to significantly higher performance for deductive than for inductive learning) but not for lower-WM participants. As noted above, for learners with lower WM, when instruction provided explicit rules, skewed input was significantly more beneficial than balanced input.

Interestingly, when exposure occurred in the absence of explicit rules, no differences between skewed and balanced input were found. This finding is congruent with data previously reported by Year and Gordon (2009) and McDonough and Trofimovich (2013) who found no contribution of skewed input relative to balanced input in the absence of rules (although in the case of Year and Gordon’s study, the manipulation involved skewed random, rather than skewed first input; see also McDonough & Nekrasova-Becker, 2014, for an advantage of balanced input).

Similarly, no significant influence of WM was observed for inductive conditions in the present results. This is despite the fact that learning was quite explicit in all cases, with no other concurrent task nor additional linguistic context. However, the finding that WM did not modulate learning during inductive exposure is in line with the previous work that indicated an effect of WM only when learners were exposed to explicit rules (e.g., Tagarelli, 2011) and that suggested that the absence of rules favored a bottom-up or “information-integration” approach.

Comparison of present and previous findings on the effect of input distributions

Altogether, the pattern of results based on learners ranging across the WM spectrum fits well within the previous literature. Crucially, by considering the joint effect of input distribution, instruction type, and WM, the present study helps reconcile previous mixed findings. The present data fit well within previous studies that reported aptitude-based interactions and help identify the contexts in which skewed input presents an advantage in adults. In particular, the results provide insight into the cognitive mechanisms that underlie the skewed-input effect and suggest that populations with lower WM (subsets of adult learners but also, e.g., younger children) may particularly benefit from exposure to skewed input.

Nonetheless, some unanswered questions and discrepancies remain. For example, in some cases, previous data still found an advantage for balanced input in the absence of

rules (McDonough & Nekrasova-Becker, 2014). In other cases, balanced input was advantageous with explicit rules but, in the absence of rules, numerically inferior to skewed learning (McDonough & Trofimovich, 2013). Results were also mixed in work from other labs (Namakura, 2012; Year & Gordon, 2009). A relevant aspect that appears to have received little attention is the potential influence of the specific linguistic constructions investigated. For example, Nakamura (2012) examined the influence of skewed and balanced input on the acquisition of the “appearance” construction used in the original studies by Goldberg and colleagues, as well as on the Samoan ergative construction. It seems revealing that, although no clear effect skewed or balanced input was observed for the Samoan ergative, the results did replicate the advantage of skewed input for the “appearance” construction. In this regard, it is worth considering how the input distribution affected linguistic elements across studies. The verbs of becoming construction targeted in the present study may have been more similar to the lexically based “appearance” construction than some of the morpheme-based constructions examined in other studies (e.g., the Samoan ergative, which is marked by the particle “e,” in Nakamura, 2012; or the Esperanto direct object-marker “-n,” in McDonough & Trofimovich, 2013). Notably (and unlike the lexically based constructions), in those morpheme-based constructions, the skewed/balanced manipulation affected the lexical items that co-occurred with the target particles but not the distribution of the target particles themselves. While examining and discussing the potential differences across the constructions examined in previous work is beyond the scope of the present paper, this is an aspect that merits further investigation in future studies.

Another important issue concerns the question of the conditions that may give rise to an advantage of skewed input. As discussed above, the present study found that, modulated by WM, skewed input influenced learning in the deductive learning condition. In the original studies by Goldberg and colleagues, an advantage of skewed was found in learning conditions that might seem most comparable to the inductive learning condition in the present study, as no explicit rule was provided. Why these differences across instruction types? It is also worth noting that, although L2 studies have typically failed to find an effect of skewed input, a previous study by McDonough and Trofimovich (2013) reported a numerical advantage for skewed input in inductive conditions, somewhat resembling the pattern in the original studies by Goldberg and colleagues (e.g., 2004; 2007). It is plausible that differences across studies may also lie in the level of explicitness of learning conditions (as well as, no doubt, on the level of complexity of the linguistic construction under consideration). In the original studies by Goldberg and colleagues, learning was incidental and meaning oriented during exposure to sentences and visual input. Similarly, McDonough and Trofimovich (2013) employed a meaning-focused task in which learners were presented with pictures and sentences and had to identify the object (rather than the agent). In contrast, given the complexity of the construction targeted, in the present study exposure was more explicit and involved a categorization task. In this study exposure was based on collocations (two-word verb-adjective combinations), rather than on sentences. This lexical approach (without sentences) was employed as a first step because learning the verbs of “becoming” construction does not involve morphosyntax, in contrast with the previous studies that explored morphosyntactic constructions (and thus required stimuli consisting of sentences, e.g., Goldberg et al., 2007; McDonough & Trofimovich, 2013; McDonough & Nekrasova-Becker, 2014; Year & Gordon, 2009). Additionally, no visual aid was employed, given the difficulty of accurately depicting a large number of often related adjectives. The focus on explicit learning of the target categories in the

present task (rather than on incidental learning during sentence processing) would seem to have substantially altered the level of attentional resources allocated to the target structure, rather than to meaning processing. That is, because previous meaning-oriented tasks required learners to focus their attention on processing sentence meanings, while attending to visual input, these conditions may have imposed greater demands on WM than in the inductive learning condition employed here. Therefore, in what concerns WM demands, it is not clear that the incidental conditions of previous studies (with no explicit rules) employing multimodal input are directly comparable to the inductive condition of the present study.¹² The question of how different paradigms may impinge on WM resources remains a question for future research to address.

Clearly, future work will be needed to replicate and extend the present findings to increase our understanding of how skewed input impacts learning, including in classroom-based settings (e.g., Madlener, 2016). In this sense, the present dataset provides an opportunity to inform future experimental efforts. Based on a power simulation performed on the present dataset with the *mixedpower* package in R (Kumle et al., 2018), future work seeking to replicate the three-way interaction reported here (with 80% power) may need a sample of ≥ 140 participants.

Implications for input optimization and personalization

At the core of the interest in investigating the potential benefits of skewed L2 input is the idea that the input can be optimized for learning. Ellis (2009) proposed that language learning can be regarded as a sampling problem, in that speakers must develop linguistic representations based on limited data. Given the limited exposure to input affecting L2 speakers in particular, designing input based on criteria such as function, frequency, and representativeness, among others, seems of particular importance. This proposal is in line with research conducted within an ATI approach during the last two decades or so (e.g., Vatz et al., 2013; DeKeyser, 2012), which suggested that input optimality may be relative to individual traits. For instance, in a study that examined L2 learners' ability to learn tone contrasts, Perrachione et al. (2011) found that higher variability in exposure to multiple speakers helped learners with high acoustic perception acquire L2 tones; however, for learners with lower acoustic perceptual abilities, high-variability training was in fact detrimental, and they benefited from reduced variability instead. Research by Brooks et al. (2006) found that a cognitive measure of executive function predicted the extent to which individual participants learning Russian gender benefitted from greater lexical variability in the input. The approach taken here adds evidence indicating that, even if each individual is exposed to the same stimuli, input efficacy should be expected to vary systematically based on individual cognitive traits and specifically WM.

The implications of the present and previous studies that revealed ATIs are potentially far-reaching, suggesting that no one type of input is beneficial across the board but that optimized input must necessarily be, to some extent, individualized input. This

¹²It is also possible that participants under inductive learning conditions may have engaged in rule searching and hypothesis testing, even if no explicit rule was provided. In this case, it can be expected that learners would have come up with rules that may have misled their learning attempts (given difficulties even among experts to formulate rules for the Spanish verbs of becoming; see, e.g., Ibarretxe-Antuñano & Cheikh-Khamis (2019)). Earlier work has also shown that engaging in the search for a complex rule without being given the rule, may not be necessarily conducive to better language learning (Reber, 1976).

approach opens opportunities for L2 researchers to identify the contexts that may trigger specific learning pathways and engage alternative systems to achieve a similar outcome (Ashby et al., 1998; Green et al., 2006). Relying on well-established constructs in psycholinguistics—such as WM (Tagarelli et al., 2011), executive control (Chen et al., 2023), or acoustic perception (Perrachione et al., 2011)—should allow to develop a systematic program of research on the optimization of input. Future work should also consider other well-studied individual-differences measures in the SLA literature—for example, language analytic ability or metalinguistic awareness (Jessner, 2008; Ranta, 2002). In sum, taking into consideration diversity among learners is a prerequisite to develop a more inclusive science of L2 learning.

Conclusion

By investigating diversity in cognitive skill—specifically, WM—among L2 learners, the present study found significant variability in the type of input that could be considered as “optimal” for L2 category learning. Following training in learning conditions that exposed participants to skewed or balanced input, and with or without explicit rules, the results indicated an important role of individual WM capacity in modulating the type of input that improved learning gains for subgroups of individuals. While the availability of explicit rules tended to enhance learning gains, in line with previous work, its combination with balanced input was detrimental for lower-WM learners, while being beneficial only for high-WM individuals. Altogether, the results help clarify the variety of results in previous research and suggest that the issue of what L2 input is optimal for learning must be examined in conjunction with individual variability in cognitive resources.

Acknowledgments. The author is grateful to Dr. Andrea Révész and to five anonymous reviewers for their helpful input. This work was supported by the College of Liberal Arts at Penn State. Thanks Carly Danielson, Rachyl Hietpas, Erica Hsieh, and Tiff Rodríguez-Cruz for their help with the implementation of the experiment and with data processing.

Competing interest. The author declares none.

References

- Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review*, 23(3), 275–290. <https://doi.org/10.1515/TLR.2006.011>
- Anwyl-Irvine, A. L., Massoné, J., Flitton, A., Kirkham, N. Z., Evershed, J. K. (2019). Gorilla in our midst: An online behavioural experiment builder. *Behavior Research Methods*, 52, 388–407 <https://doi.org/10.3758/s13428-019-01237-x>
- Ashby, F. G., Alfonso-Reese, L. A., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, 105(3), 442.
- Bjork, R. A., & Kroll, J. F. (2015). Desirable difficulties in vocabulary learning. *The American Journal of Psychology*, 128(2), 241–252.
- Bovolenta, G., & Marsden, E. (2022). Prediction and error-based learning in L2 processing and acquisition: a conceptual review. *Studies in Second Language Acquisition*, 44(5), 1384–1409. <https://doi.org/10.1017/S0272263121000723>
- Brooks, P. J., & Kempe, V. (2013). Individual differences in adult foreign language learning: The mediating effect of meta-linguistic awareness. *Memory & Cognition*, 41, 281–296. <https://doi.org/10.3758/s13421-012-0262-9>
- Brooks, P. J., Kwoka, N., & Kempe, V. (2017). Distributional effects and individual differences in L2 morphology learning. *Language Learning*, 67(1), 171–207.

- Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics*, 27, 185–209. <https://doi.org/10.1017/S0142716406060243>
- Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 226–246). Routledge.
- Bybee, J., & Eddington, D. (2006). A usage-based approach to Spanish verbs of “becoming.” *Language*, 323–355.
- Casenhiser, D. M., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science*, 8, 500–508.
- Chandrasekaran, B., Yi, H. G., & Maddox, W. T. (2014). Dual-learning systems during speech category learning. *Psychonomic Bulletin & Review*, 21(2), 488–495. <https://doi.org/10.3758/s13423-013-0501-5>
- Chen, J., Zhao, Y., Xu, X., Liu, J., & Deng, C. (2023). Predictive relations between executive function and Mandarin vocabulary ability among Uyghur-Mandarin bilingual preschoolers. *Cognitive Development*, 65, 101270. <https://doi.org/10.1016/j.cogdev.2022.101270>
- DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition*, 107(1), 284–294.
- DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, 62, 189–200. <https://doi-org.ezaccess.libraries.psu.edu/10.1111/j.1467-9922.2012.00712.x>
- DeKeyser, R. M. (2021). *Aptitude-treatment interaction in second language learning*. John Benjamins Publishing Company.
- Eddington, D. (1999). On “becoming” in Spanish: A corpus analysis of verbs expressing change of state. *Southwest Journal of Linguistics*, 18(2), 23–35.
- Ellis, N. C. (2009). Optimizing the input: Frequency and sampling in usage-based and form-focused learning. *The Handbook of Language Teaching*, 139.
- Ellis, N. C., O'Donnell, M. B., & Römer, U. (2013). Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning*, 63, 25–51.
- Ellis, N. C., & Wulff, S. (2014). Usage-based approaches to SLA. In *Theories in second language acquisition* (pp. 87–105). Routledge.
- Forkstam, C., & Petersson, K. M. (2005). Towards an explicit account of implicit learning. *Current Opinion in Neurology*, 18(4), 435–441. <https://doi.org/10.1097/01.wco.0000171951.82995.c4>
- Goldberg, A. E. (2013). Constructionist approaches. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 15–31). Oxford University Press.
- Goldberg, A. E., Casenhiser, D., & White, T. R. (2007). Constructions as categories of language. *New Ideas in Psychology*, 25(2), 70–86.
- Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, 15(3), 289–316. <https://doi.org/10.1515/cogl.2004.011>
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58(3), 787–814
- Green, D. W., Crinion, J., & Price, C. J. (2006). Convergence, degeneracy, and control. *Language Learning*, 56, 99–125.
- Hartig, F. (2022). DHARMA: Residual diagnostics for hierarchical (multi-level/mixed) regression models. R package version 0.4.6. <https://CRAN.R-project.org/package=DHARMA>
- Ibarretxe-Antuñano, I., & Cheikh-Khamsi, F. (2019). “How to become a woman without turning into a Barbie”: Change-of-state verb constructions and their role in Spanish as a foreign language. *International Review of Applied Linguistics in Language Teaching*, 57(1), 97–120. <https://doi.org/10.1515/iral-2018-2008>
- Ibbotson, P., Theakston, A. L., Lieven, E. V., & Tomasello, M. (2012). Semantics of the transitive construction: Prototype effects and developmental comparisons. *Cognitive Science*, 36(7), 1268–1288.
- Indrarathne, B., & Kormos, J. (2018). The role of working memory in processing L2 input: Insights from eye-tracking. *Bilingualism: Language and Cognition*, 21(2), 355–374.
- Jessner, U. (2008). A DST model of multilingualism and the role of metalinguistic awareness. *The Modern Language Journal*, 92(2), 270–283.
- Kempe, V., Brooks, P. J., & Kharkhurin, A. (2010). Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning*, 60(1), 127–153. <https://doi-org.ezaccess.libraries.psu.edu/10.1111/j.1467-9922.2009.00553.x>

- Kumle, L., Vö, M. L. H., & Draschkow, D. (2018). Mixedpower: A library for estimating simulation-based power for mixed models in R. Computer software. <https://github.com/DejanDraschkow/mixedpower>
- Lavi-Rotbain, O., & Arnon, I. (2022). The learnability consequences of Zipfian distributions in language. *Cognition*, 223, 105038. <https://doi.org/10.1016/j.cognition.2022.105038>
- Lavi-Rotbain, O., & Arnon, I. (2023). Zipfian distributions in child-directed speech. *Open Mind*, 7, 1–30. https://doi.org/10.1162/opmi_a_00070
- Li, S. (2022). Working memory and second language learning: A critical and synthetic review. In *The Routledge handbook of second language acquisition and psycholinguistics*, Routledge (pp. 348–360).
- Linck, J. A., & Weiss, D. J. (2015). Can working memory and inhibitory control predict second language learning in the classroom? *Sage Open*, 5(4), 2158244015607352.
- Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, 21, 861–883.
- Nakamura, D. (2012). Input skewedness, consistency, and order of frequent verbs in frequency-driven second language construction learning: A replication and extension of Casenhiser and Goldberg (2005) to adult second language acquisition. *International Review of Applied Linguistics (IRAL)*, 50, 31–67.
- Maddox, W. T., Chandrasekaran, B., Smayda, K., & Yi, H.-G. (2013). Dual systems of speech category learning across the lifespan. *Psychology and Aging*, 28(4), 1042–1056. <https://doi.org/10.1037/a0034969>
- Madlener, K. (2016). Input optimization. In H. Behrens and S. Pfänder (Eds.), *Effects of type and token frequency manipulations in instructed second language learning* (pp. 133–173). Walter De Gruyter.
- Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science*, 17, 944–948. doi:10.1111/j.1467-9280.2006.01809.x
- McDonough, K., & Trofimovich, P. (2013). Learning a novel pattern through balanced and skewed input. *Bilingualism: Language and Cognition*, 16(3), 654–662.
- McDonough, K., & Nekrasova-Becker, T. (2014). Comparing the effect of skewed and balanced input on English as a foreign language learners' comprehension of the double-object dative construction. *Applied Psycholinguistics*, 35(2), 419–442.
- Motha, H. (2013). *The effect of deductive and inductive learning strategies on language acquisition*. (Unpublished master's thesis). Tilburg University, Tilburg.
- Navarro-Torres, Christian A., Dussias, Paola E., & Kroll, Judith F. (2022). When exceptions matter: Bilinguals regulate their dominant language to exploit structural constraints in sentence production. *Language, Cognition and Neuroscience*, 1–26. <https://doi.org/10.1080/23273798.2022.2105915>
- Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, 130(1), 461–472.
- Pulido, Manuel F. (2021). Individual chunking ability predicts efficient or shallow L2 processing: Eye-tracking evidence from multiword units in relative clauses. *Frontiers in Psychology*, 11, 4004. <https://doi.org/10.3389/fpsyg.2020.607621>
- Pulido, M. F., & Dussias, P. E. (2020). Desirable difficulties while learning collocations in a second language: Conditions that induce L1 interference improve learning. *Bilingualism: Language & Cognition*, 23(3), 652–667. <https://doi.org/10.1017/S1366728919000622>
- R Core Team, R. (2023). R: A language and environment for statistical computing.
- Ranta, L. (2002). The role of learners' language analytic ability in the communicative classroom. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 159–179). John Benjamins.
- Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, 2(1), 88–94. <https://doi.org/10.1037/0278-7393.2.1.88>
- Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(5), 888.
- Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, 17, 368–392. <https://doi.org/10.1177/026765830101700405>
- Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, 25, 46–73.

- Seliger, H. W. (1975). Inductive method and deductive method in language teaching: A re-examination. *International Review of Applied Linguistics in Language Teaching*, 13(1–4), 1–18. <https://doi.org/10.1515/iral.1975.13.1-4.1>
- Skehan, P. (1989). *Individual differences in second language learning*. Edward Arnold.
- Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80, 99–103. https://doi.org/10.1207/S15327752JPA8001_18
- Suzuki, Y., Nakata, T., & DeKeyser, R. (2019). Optimizing second language practice in the classroom: Perspectives from cognitive psychology. *The Modern Language Journal*, 103(3), 551–561.
- Tagarelli, K. M., Mota, M. B., & Rebuschat, P. (2011). The role of working memory in implicit and explicit language learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33) Cognitive Science Society, Austin, TX pp. 2061–2066.
- Tagarelli, K. M., Mota, M. B., & Rebuschat, P. (2015). Working Memory, Learning Conditions and the Acquisition of L2 Syntax. In Z. Wen, M. Borges Mota, and A McNeill (Eds.), *Working memory in second language acquisition and processing*. Multilingual Matters.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medicine Education*, 2, 53–55. doi: 10.5116/ijme.4dfb.8dfd
- Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Vatz, K., Tare, M., Jackson, S. R., & Doughty, C.J. (2013). Aptitude-treatment interaction studies in second language acquisition. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 273–292). John Benjamins. <https://doi.org/10.1075/lllt.35.11vat>
- Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, 8(1), 168–176.
- Year, J., & Gordon, P. (2009). Korean speakers' acquisition of the English ditransitive construction: The role of verb prototype, input distribution, and frequency. *The Modern Language Journal*, 93(3), 399–417.
- Zhang, X., & Dong, X. (2019). Input frequency and construction interference interactions in L2 development. *Second Language Research*, 35(4), 505–527. <https://doi.org/10.1177/0267658318791651>
- Zhang, X., & Mai, C. (2023). Impact of constructional complexity and intralingual influence on the effectiveness of skewed input. *Language Teaching Research*, 27(5), 1217–1245. <https://doi.org/10.1177/1362168820981395>

Appendix A

Table S1. Individual measures by group (measures were included as covariates in analysis)

Measure	Age (SD)	Range	O-Span (SD)	Range	Picture naming (SD)	Range
Balanced inductive	29.4 (6.59)	20–41	49.8 (7.88)	36–60	0.950 (0.54)	0.806–1
Balanced deductive	33.7 (8.06)	21–45	52.8 (5.91)	37–59	0.980 (0.35)	0.875–1
Skewed inductive	29.9 (5.49)	22–40	50.7 (9.35)	17–61	0.950 (0.49)	0.806–1
Skewed deductive	33.9 (6.33)	23–44	50.6 (6.94)	32–60	0.986 (0.026)	0.875–1

Table S2. Post-hoc contrasts for input distribution at ± 1 SD WM scores (FDR-corrected)

Contrast	Instruction type	Estimate	SE	df	z ratio	p
WM 1962.92 balanced – WM 3331.68 balanced		0.072	0.390	Inf	0.185	0.998
WM 1962.92 balanced – WM 1962.92 skewed		0.137	0.358	Inf	0.382	0.981
WM 1962.92 balanced – WM 3331.68 skewed		0.121	0.351	Inf	0.346	0.986
WM 3331.68 balanced – WM 1962.92 skewed		0.064	0.366	Inf	0.176	0.998
WM 3331.68 balanced – WM 3331.68 skewed		0.049	0.347	Inf	0.141	0.999
WM 1962.92 skewed – WM 3331.68 skewed	Inductive	-0.015	0.316	Inf	-0.049	1.000
WM 1962.92 balanced – WM 3331.68 balanced		-1.286	0.447	Inf	-2.879	0.021
WM 1962.92 balanced – WM 1962.92 skewed		-1.067	0.398	Inf	-2.678	0.037
WM 1962.92 balanced – WM 3331.68 skewed		-0.786	0.416	Inf	-1.887	0.234
WM 3331.68 balanced – WM 1962.92 skewed		0.219	0.334	Inf	0.656	0.914
WM 3331.68 balanced – WM 3331.68 skewed		0.500	0.353	Inf	1.419	0.488
WM 1962.92 skewed – WM 3331.68 skewed	Deductive	0.281	0.326	Inf	0.863	0.824

Table S3. Post-hoc contrasts for instruction type at ± 1 SD WM scores (FDR-corrected)

Contrast	O-Span (transformed)	Input distribution	Estimate	SE	df	z ratio	p
Deduc. – Induc.	1,962.92	balanced	0.564	0.439	Inf	1.284	0.199
Deduc. – Induc.	3,331.68	balanced	-0.795	0.385	Inf	-2.063	0.039
Deduc. – Induc.	1,962.92	skewed	-0.640	0.358	Inf	-1.789	0.074
Deduc. – Induc.	3,331.68	skewed	-0.343	0.361	Inf	-0.951	0.342

Appendix B

Instructions for training task:

Inductive and deductive conditions:

“This is an experiment investigating language learning. In this task, you will see English words presented on the screen as text. These words will be adjectives (i.e., descriptor words such as ‘beautiful,’ ‘small,’ ‘big,’ ‘difficult’). In addition, you will also *hear* one of two Spanish verbs in each screen (‘ponerse’ and ‘quedarse’). Both Spanish verbs translate as ‘become’ in English, but Spanish requires that you sometimes use one verb and sometimes the other, depending on the particular adjective. Your task is to listen to the verb and select the correct verb, while trying to learn what Spanish verb to use with each English word.

The experiment contains two short practice blocks before the experimental block. The whole task takes less than 10 minutes. There will be one break during the main task.”

Additional rule provided for deductive condition only: “In addition to the examples, there is one critical RULE: One category is associated with the verb ‘ponerse.’ These are words that are related to the word ‘nervous’ and are to some extent similar in meaning. The other category is associated with the verb ‘quedarse.’ This group of words is related to the word ‘still.’ You shouldn’t try to come up with other rules, as that will confuse you! The rule to remember is that ‘ponerse’ is about words related to ‘nervous,’ and ‘quedarse’ about words related to ‘still.’”

Instructions for test following training: “Now you will complete a test. This will look very similar to the training you just completed. However, you will *not* hear what verb each word should go with. Do your best to remember and choose the correct verb.”

Instructions for generalization task: “This task is related to a previous task in which you practiced selecting and learning what Spanish verb to use with different adjectives. This task will work the same way. But this time you will see different adjectives and you will *not* be told what verb is correct.

Your task will be to select the verb that seems most appropriate based on the critical rule that you learned.

REMEMBER: ‘Ponerse’ will be associated with words that are related to the word ‘nervous’ and are to some extent similar in meaning. # ‘Quedarse’ is associated with words related to the word ‘still.’ You shouldn’t try to come up with other rules, as that will confuse you. Do your best to categorize words based on this rule.”

Cite this article: Pulido, M. F. (2024). Optimizing the input for learning of L2-specific constructions: The roles of Zipfian and balanced input, explicit rules and working memory. *Studies in Second Language Acquisition*, 46: 379–403. <https://doi.org/10.1017/S0272263124000081>