

Challenges and potential of quasi-experimental studies in Cognitive Linguistics applied to language teaching and learning

Beatriz Martín-Gascón

beatma22@ucm.es

Universidad Complutense de Madrid

Edificio D Planta 1ª izquierda. Ciudad Universitaria 28040 – Madrid (España)

Abstract: This review study focuses on the challenges faced when conducting effects-of-instruction research in Applied Cognitive Linguistics. It offers an overview of relevant research that has sought to demonstrate the effectiveness of a cognitive-based approach to second language (L2) pedagogy and provides illustrations from previous work in L2 Spanish by the author and colleagues of some of the factors that can influence findings in quasi-experimental research. More specifically, the study addresses some of the difficulties encountered in the design of materials and assessment tests, during the pedagogical intervention and the data collection and data analysis phases. These include choice of assessment, test effects, sample size, withdrawal, and time-on-task, among others. Along with these methodological issues, a discussion of possible solutions as well as pedagogical and methodological implications are discussed.

Key words: Effects-of-instruction, Applied Cognitive Linguistics, L2 Spanish, Quasi-experimental research, Challenges.

1. INTRODUCTION

Applied Cognitive Linguistics (ACL) has recently emerged as a prominent pedagogical approach, focusing on the study of acquisition processes and second language (L2) pedagogical methodologies. Since the onset of the 21st century, it has gained increasing attention in empirical research, finding notable application within L2 classrooms. Significant contributions during this period, as underscored by Llopis-García (2023), have played pivotal roles in delineating criteria and solidifying the theoretical and applied foundations of ACL's potential in L2 learning and instruction (Pütz, Niemeier & Dirven, 2001). The work of Achard and Niemeier (2004) elucidated how concepts formulated within the domain of Cognitive Linguistics (CL) offer insightful perspectives on various pivotal aspects of L2 pedagogy. Subsequently, De Knop and De Rycker (2008) advanced efforts in L2 instruction to establish a robust conceptual nexus between CL and L2 instruction. In a parallel vein, Robinson and Ellis (2008) emphasized the understanding of L2 acquisition as an immersive process of information processing and language production during social interactions, wherein intentions, concepts, and meaning are conveyed.

Building upon these theoretical underpinnings, empirical investigations, such as the study conducted by De Knop, Boers and De Rycker (2010), extended this reasoning with practical applications, contending that instruction inspired by this intersection can establish robust connections between form and meaning in learners' memory. Concurrently, Littlemore and Juchem-Grundmann (2010) featured real classroom studies, yielding positive outcomes for CL-based instruction in L2 Spanish (Llopis-García's study) and L2 English (Mueller & Ho's study).

Despite an increased degree of interest in the potential application of CL to L2 learning, there persists an urgent need to establish it in L2 classrooms (Piquer-Píriz, 2021; Llopis-García, in press) and empirical endeavors aiming to demonstrate the effectiveness of a CL-based approach to L2 pedagogy remain scarce. The past five years have witnessed a proliferation of substantial empirical and theoretical advancements in the discipline with a focus on L2 Spanish consolidated in two volumes edited by Ibarretxe-Antuñano, Cadierno and Castañeda Castro (2019), and Llopis-García and Hijazo-Gascón (2019). Yet, due to the relative youth of the discipline, much of the conducted research furnishes theoretical frameworks and pedagogical proposals yet awaiting empirical validation. While numerous studies have been conducted, there remains a scarcity of empirical evidence supporting these theoretical frameworks. Notably, only a handful of empirical studies, such as those by Llopis-García (2009), Llopis-García and Alonso Aparicio (2019), and Colassaco (2019), have delved into the efficacy of CL approaches integrating principles from cognitive grammar and processing instruction. The limited empirical literature on ACL appears, on one hand, to converge in asserting that the innovative nature of cognitive instruction sometimes yields outcomes that have yet to conclusively demonstrate the superior efficacy of ACL in the L2 classroom (Llopis-García, 2023, p. 34). On the other hand, studies yielding positive outcomes also present real challenges, stemming from their novelty in quasi-experimental design. Furthermore, this type of design in Applied Linguistics often entails controlled environments, rigorous statistical analyses, and large sample sizes (Gass & Mackey, 2016). These elements, however, can pose a certain level of challenge when applied to effects-of-instruction research.

This article therefore aims to delineate the potential and constraints inherent in the application of Quasi-Experimental Design (QED) to inform research within ACL, particularly in the context of pedagogical interventions within L2 (Spanish) classrooms. We aim to examine factors that can influence findings in such empirical investigations by delving into the challenges encountered in the design of materials and assessment tests, intervention, data collection and data analysis phases. These include choice of assessment, test effects (i.e., effect of the assessment *vs.* effect of the teaching intervention), sample size, withdrawal, and time-on-task, among others.

The structure for this paper is as follows. Initially, the underpinnings of QED within the context of effects-of-instruction research, with a specific focus on its application in the more general domain of Applied Linguistics will be elucidated (Sections 2 and 2.1). Subsequently, a review of empirical investigations within the field of CL particularly emphasizing their application in the L2 classroom, while critically assessing the factors contributing to their partial success will be undertaken (Section 3). This examination will set the stage for the subsequent presentation of QED studies conducted by the author in collaboration with colleagues (Martín-Gascón 2023a, 2023b, in press; Martín-Gascón, Llopis-García and Alonso-Aparicio 2023) in Section 4. These studies, delineated as three distinct studies (4.1-4.3) for clarity, serve as descriptive examples of the effectiveness of employing an ACL approach in teaching and assessing complex and neglected linguistic forms in an L2 Spanish instructional context. Additionally, they provide a nuanced analysis of both the strengths and limitations inherent in the chosen QED (4.4). Notably, these investigations are characterized by their quasi-experimental nature, specifically focusing on the L2 acquisition of Spanish, which represents a relatively underexplored target language in comparison to English within the realm of ACL and empirical research.

Finally, this paper will explore the implications of the findings for prospective research endeavors (Section 5).

2. QUASI-EXPERIMENTAL DESIGNS (QED)

In this section, we will examine the foundational principles of Quasi-Experimental research Design (QED) within the realm of effects-of-instruction research, with particular emphasis on its use within the domain of Applied Linguistics. QED, as their name implies, make use of naturally occurring variations in the main independent variable of interest, rather than manipulating it artificially (Campbell & Stanley, 2015). In this type of design, some subjects are exposed to a treatment while others are not, but the assignment is not randomized like in true experimental designs. Hence, QED represent a category of non-experimental research aiming at replicating the rigor and structure of randomized true experiments, albeit without the element of random assignment (Cook & Wong, 2008; Kirk, 2009). By doing so, they seek to mitigate selection bias and minimize threats to internal validity. This research design is rated as having the second-highest level of causal evidence, with well-conducted randomized controlled trials being the highest (Zabor, Kaizer & Hobbs, 2020). Hence, the attraction towards employing QED stems from their capacity to yield robust, unbiased causal estimates for a diverse array of educational interventions, showcasing their potential. However, the hierarchy of evidence in terms of the strength and reliability of findings from QED studies is still evolving and remains an area of active research exploration.

2.1. QED effects-of-instruction research in Applied Linguistics

Within the field of Applied Linguistics, QED play a key role in establishing causal relationships between variables. In empirical studies conducted in instructional settings, like university language classrooms with pre-existing courses and students already enrolled in a class, achieving adequate randomization would be rather challenging. QED do not rely on random assignment; rather, students are assigned to groups based on non-random criteria, such as teacher-researcher decisions.

Furthermore, QED does not require to have a true control group, although a comparison group might be involved. This latter constitutes an additional experimental cohort that undergoes a distinct experimental intervention. For instance, when examining the effects of a particular instructional intervention, researchers may determine that intact classes provide the most ecologically valid environment for conducting the study, as discussed in prior works by Gass and Mackey (2016) or Plonsky (2017). In such instances, a comparison group engaged in standard classroom instruction can serve as the optimal reference point for the group in the experimental condition. In this vein, Hamrick and Sachs (2018) contend that employing a trained control group (i.e., a comparison group) as opposed to a real control group may assist in mitigating potential hidden biases among participants. Additionally, as opposed to non-experimental studies, which can also manifest as pre-experimental designs and do not employ either a control or comparison group (Nunan, 1992), QED empower researchers to make more definitive inferences regarding the causal connection between two variables (Marsden & Torgerson, 2012), such as teaching method and immediate (and deferred) learning gains.

While experimental studies offer advantages in minimizing biases and confounding variables, they are often associated with high costs, ethical constraints, and practical challenges. Conversely, QED studies are valuable in situations where conducting an

experiment is impractical, and they can be carried out with smaller sample sizes, as happens in empirical research within Applied Linguistics.

Within the domain of Applied Linguistics using QED, a key consideration lies in the specific contextual applications of research findings. It is important to validate experimental materials and procedures within these contexts. For instance, when endeavoring to extend the applicability of results and conclusions to other L2 classrooms, a preliminary investigation into the prevailing instructional methods and characteristic learning practices in such environments is essential. The insights gleaned from this examination can provide the foundational framework for the development of experimental materials and procedural design (Lightbown & Spada, 2019). By integrating these considerations, researchers might avoid the potential pitfalls associated with arbitrary experimental manipulations. This approach ensures that the findings not only uphold scholarly rigor but also resonate with and cater to the practical contexts they are intended to serve. It establishes a robust connection between theoretical research and its pragmatic implementation, increasing the impact of research within effects-of-instruction research in Applied Linguistics.

3. EMPIRICAL RESEARCH IN ACL TO THE L2 CLASSROOM

In this section, a review of empirical studies within the domain of CL with a specific emphasis on their application within the L2 educational setting will be presented. Concurrently, an assessment of the potential contributing factors underlying their (sometimes partial) efficacy will be undertaken.

Few QED studies aimed at investigating the effectiveness of applying CL to teaching linguistic forms in the L2 classroom have been carried out in the last decades. Researchers

with a CL background have proposed ways of exploiting non-arbitrary, symbolic, and usage-based aspects of language largely contributing to highlighting the potential of applying CL to L2 learning (e.g., Achard & Niemeier, 2004; De Knop & De Rycker, 2008; Pütz, Niemeier & Dirven, 2001). Yet many of their proposals are either small-scale, show small effect sizes or are hard to interpret due to confounding variables (Boers, 2013).

Studies like the ones conducted by Tyler, Mueller and Ho (2010, 2011), Hung, Vien and Vu (2018), Colassaco (2019), Llopis-García (2019), or Alonso-Aparicio and Llopis-García (2019) report results of QED effects-of-instruction exploring the (relative) efficacy of applying a CL approach to either L2 English or L2 Spanish. For the purpose of brevity, yet acknowledging the existence of additional pertinent research, attention will be directed towards these selected studies along with strengths and limitations encountered.

The studies conducted by Tyler, Mueller and Ho (2010, 2011) focus on the application of CL in learning the semantics of English modals and prepositions. In the first study, the authors identify a persistent challenge faced by L2 learners in understanding modal verbs. These verbs, though frequently encountered in the input, prove complex due to their dual semantic nature—root and epistemic senses. Traditional teaching materials (i.e., textbooks), tend to treat these meanings as interchangeable, neglecting systematic semantic patterns within the modal system. In contrast, the authors defend that CL analyses provide students with a motivated representation of the relationship between root and epistemic meanings, offering a precise semantic delineation of each modal. To assess the pedagogical efficacy of the CL approach to modals, an effects-of-instruction study was conducted with high-intermediate adult L2 English learners. Three groups were

formed: a cognitive treatment group, a Speech Acts treatment group, and a control group. Results revealed that the cognitive treatment group exhibited greater improvement compared to the Speech Acts treatment group. This empirical evidence substantiates the assertion that CL not only furnishes a compelling analytical framework for language, but also forms the basis for more effective grammar instruction than prevalent teaching materials.

A limitation of the study lies in the observed moderate treatment effect, amounting to a 2.7-point gain in performance. This outcome could be attributed to the restricted duration of the treatment, during which both instruction groups received approximately two hours of teaching on modals. It is noteworthy that, within this constrained timeframe, the cognitive group also faced a potential disadvantage that pervades in empirical studies within ACL, and which is grappling with the introduction of novel and substantially different grammatical concepts, such as *force dynamics* and *metaphor*. These cognitive demands, coupled with the need to assimilate new meanings for modals and their contextual use, might have contributed to the observed moderate effect. A more comprehensive analysis should consider whether an extended duration of instruction could have yielded more significant treatment effects.

The research underscores a limitation pertaining to the scope of the modal system under examination. The study acknowledges that the selected target forms do not offer a comprehensive representation of the modal system. Numerous modals and periphrastic constructions remained unexplored throughout the research. This limitation bears implications for the generalizability of the findings, as the study's outcomes may not entirely capture the intricacies of modal use. As a potential avenue for future exploration, the study suggests embarking on a longitudinal study that seeks to encompass a more

comprehensive array of modal constructs. Such an approach would be pivotal in establishing a more robust understanding of the cognitive linguistics-based instructional method.

Another drawback pertains to the alignment between the instructional approach and assessment tests. Specifically, the assessment used in the study did not match the CL-based instruction. Instead, tests employed a multiple-choice, fill-in-the-blank format. This disjunction between instruction and evaluation methodologies constitutes an important constraint, as it potentially compromises the study's capacity to accurately measure the learning results linked with the cognitive approach.

In their subsequent study in 2011, Tyler, Mueller and Ho extend their inquiry to the semantics of English prepositions. They propose a CL-based model that conceptualizes the manifold meanings associated with prepositions as a motivated polysemy network. According to the authors, this framework holds promise in providing a structured framework for L2 learners grappling with the intricate array of meanings attributed to prepositions. They further emphasize the imperative nature of experimental investigations in advancing the field of ACL and underscore that while a theoretically compelling description of language is key, it alone does not suffice as a panacea for L2 pedagogy and acquisition. The challenge, which is one faced by ACL research in general, lies in translating these theoretical underpinnings into effective teaching materials that are accessible to both learners and instructors. To this end, the authors advocate for a multifaceted approach that incorporates the significance of noticing, the incorporation of pushed input, and explicit instruction followed by communicative tasks, among others.

In evaluating the study, a few limitations have been identified that allow for discussion of a more comprehensive understanding of their findings. These restraints encompass the

study's sample size, treatment duration, assessment methods and absence of control group. The first limitation pertains therefore to the narrow population it represents, consisting of only 14 advanced learners. While the study provides valuable insights into the CL-based instruction's impact on this specific group, its generalizability to other advanced groups or to a broader range of proficiency levels remains disputed. In future studies, it would be worth including a larger sample or participants with varying levels of proficiency to ascertain the treatment's effectiveness across a wider spectrum. Furthermore, all participants in the study were native speakers of Italian, a language closely related to English which features numerous prepositions with shared meanings. This shared linguistic background may have played a role in the participants' performance.

Another limitation observed refers to the brief duration of the treatment, consisting of only two teacher-fronted instruction sessions lasting 50 minutes each, accompanied by two pair-work sessions of 30 minutes each. The limited exposure to the treatment raises questions about the depth of the participants' understanding and retention of the extended meanings and meaning extension processes. To address this limitation, future research could consider incorporating a delayed posttest, which would provide insights into the sustainability of the gains observed. Besides, the absence of control groups in the study poses an important limitation. Without a control group, it is challenging to assert confidently that CL-based instruction offers a superior basis for L2 learning. To establish the comparative effectiveness of this approach, follow-up experiments should include control groups. Lastly, the choice of assessment methods used in the study may not have been well-aligned with the CL instruction. To mitigate this limitation, future studies could consider designing assessments that better align with the CL approach.

More recently, but also in the domain of teaching prepositions in L2 English, Hung, Vien and Vu (2018) present another study that constitutes a quasi-experimental endeavor aiming at assessing the efficacy of integrating CL to the L2 classroom. The research design employed in this investigation is of a pretest-posttest between-group nature. The selection of participants is based on several criteria, including their prior learning experiences, consistent exposure to English outside of formal classroom settings, willingness to participate in the study, and performance on the pretest. Through a structured curriculum spanning four sessions, each lasting 90 minutes, the study focuses on elucidating the spatial and metaphorical connotations of a list of prepositions. Within this pedagogical framework, the cognitive group received instruction rooted in CL principles, while the traditional group was instructed using vivid visual. The outcome of the study revealed a notable disparity in favor of the cognitive group during the posttest phase. This confirms the hypothesis that CL-based instruction yields better results in comparison to traditional pedagogical approaches.

Furthermore, their study included a cohort of 50 first-year students from distinct intact classes. While this sample size allows for broader generalizations, it is important to acknowledge a noteworthy constraint in the experimental framework: the absence of a control group. This omission precludes the establishment of a comparative baseline against which the effects of the interventions can be rigorously and reliably assessed. Apart from this, the study used a pretest-posttest design, which helps assess changes over time. However, the absence of a delayed posttest diminishes the power to discern the potential long-term impacts of the interventions.

In addition to the foregoing, another limitation, already observed in previous studies, deals with the alignment of assessment material with instructional method. Tests included

exercises rooted in multiple-choice questions, sentence and text completion. This raises a concern about the congruence between the evaluation tools and the instructional approach. A more robust evaluation might have entailed assessments tailored to mirror the specific linguistic and cognitive demands of the CL instruction, thereby ensuring a more precise measure of the intervention's impact on the targeted skills and competencies. If research focusing on L2 English has been relatively scarce, the state of investigations in L2 Spanish has been even more overlooked. Recent contributions to address this gap can be found in studies published in Ibarretxe-Antuñano, Cadierno and Castañeda Castro (2019), and Llopis-García and Hijazo-Gascón (2019). Noteworthy among these contributions is Colassaco's (2019) study, which explores the efficacy of two instructional methodologies, cognitive and traditional, in teaching deictic motion verbs to Italian and German learners of L2 Spanish (B1 level). The cognitive approach, which incorporated principles from cognitive grammar and processing instruction, demonstrated higher outcomes in deictic motion verbs use compared to the traditional approach.

Regarding data collection tools designed by Colassaco (2019), tests exhibited both commendable aspects and certain limitations. On the positive side, a strong emphasis on meaning and the deictic center of the speech act was found, underscoring the importance of context and speakers' reference point in linguistic interpretation. However, tests still relied on certain conventional formats, such as fill-in-the-gaps and the need to choose between two options, which departs from a more CL view of interpreting and producing deictic motion verbs. On the other hand, the study measured the target form knowledge through a pretest and three post-tests: post-test 1, administered immediately after the intervention; post-test 2 one week later; and post-test 3 one month after. This temporal spread of assessments provides a comprehensive view of the potential long-term effects

of the experimental conditions. It allows for an exploration of the retention of the learning outcomes, shedding light on the persistence of any observed improvements over time. Another important aspect of Colassaco's study is the big sample size, consisting of a total of 274 students. This robust sample not only enhances the statistical power of the analysis but also contributes to the generalizability of her findings. Furthermore, the presence of two experimental groups, each receiving distinct instructional approaches (CL and traditional), in addition to a control group, allows for a more nuanced understanding of the impact of a CL instructional method.

In a similar vein, Llopis-García (2019) contextualizes and presents results from a previous empirical study (Llopis-García, 2009) that ascertain whether a CL instructional approach, coupled with a focus on interpretation and production of verb mood via processing instruction, facilitates the learning of the indicative-subjunctive contrast among L2 Spanish learners. The research design included a total of 81 German students with A2/B1 proficiency level who were divided and methodically assigned through random allocation into three groups: two experimental groups undergoing instruction based on CL (group 1 focusing on interpretation, $n=35$, and group 2 on production, $n=31$), in addition to a control group devoid of pedagogical treatment ($n=15$). Her findings showed that the experimental groups receiving CL instruction for modal selection exhibited a more substantial impact in both interpretation and production tasks compared to the control group.

It is worth noting that an experimental group receiving instruction from a more traditional linguistic perspective could have helped corroborate the productivity of the CL pedagogical approach. Additionally, the tools used for data collection assessed both interpretation and production abilities, despite the fact that the instruction primarily

emphasized either production (experimental group 1) or comprehension (experimental group 2). This indicates to some extent a potential misalignment between the evaluation methods and the instructional emphasis. The study exhibited several notable strengths. Among many others, it spanned a duration of eight weeks, during which participants attended one 90-minute session per week. This longitudinal approach allowed for comprehensive examination through an initial assessment (pretest) and two subsequent posttests. Yet, as the author (2009, p. 155) highlights, the longitudinal nature of the study also introduced a potential drawback in the form of fluctuating levels of student participation across various stages.

Alonso Aparicio and Llopis-García (2019) carried out an empirical study focusing on the acquisition of the imperfect/past simple contrast and examining whether a CL approach surpasses the effectiveness of the more pervasive notional-functional method. Their study involved 131 students in a university in North America who had an A2 level of Spanish. Students were randomly assigned to each experimental condition (CL $n=45$, traditional $n=41$, and control $n=45$). This sample size was reduced to 56 students (CL $n=17$, traditional $n=18$, and control $n=21$), as those who did not attend all sessions, spoke another romance language or had a mark equal or superior to 6 out of 10 in the pretest were excluded. Their findings revealed that in one of the tasks (interpretation), both experimental groups outperformed the control condition, though there were no significant differences between the two experimental groups. Yet, in the production task, students from the cognitive group did demonstrate greater improvement compared to those who received traditional instruction.

As the authors argue, the sample size might have hindered the generalizability of the results and the lack of significance in the case of the interpretation task. Additionally, the

nature of the instruments used for data collection and the absence of delayed post-tests also constrain the conclusions drawn. In the first instance, given that the tests used for data collection favored monitoring, it is not possible to conclude to what degree the acquired knowledge would have transferred to the execution of more spontaneous and natural tasks, such as an oral interview, for example. Furthermore, a grammaticality judgment may not be the most suitable instrument for measuring knowledge, as, despite researchers' insistence during the pre and post-tests that participants reflect and not choose randomly, participants already had a 50% success rate prior to task completion, as asserted by the authors (p. 328). In the second case, due to a conflict with the curriculum of the institution where the research was conducted, no delayed post-tests were included. Consequently, it is hard to ascertain the extent of long-term retention of the acquired imperfect/past simple contrast.

Given the preceding context, the following section delineates studies that, taking into consideration the reviewed literature, approached certain methodological inquiries — namely, assessment type, student cohort size, pseudo-randomization, and pre, post, and delayed post-test designs— in a distinctive manner. We will now proceed to detail this in Section 4.

4. THE CASE OF THREE RECENT QED EFFECTS-OF-INSTRUCTION INVESTIGATIONS

This section introduces three QED investigations carried out by the author and colleagues with the objective of furnishing demonstrative instances of the efficacy of employing an ACL methodology in instructing complex linguistic forms within the L2 Spanish

instructional setting. Furthermore, these studies furnish a discerning examination of the strengths and constraints inherent in the selected QED methodology (4.4).

This series of ACL research was motivated by an observable absence of experimental research within the realm of CL and L2 Spanish pedagogy and assessment regarding the advantages of integrating these disciplines in an articulated manner. Each investigation applied concepts from CL in an instructional setting focused on teaching and assessing L2 Spanish at the university level.

This methodological approach sought to ensure a comprehensive evaluation of L2 learners' performance, in a manner that harmonized with the theoretical underpinnings and instructional strategies grounded in CL. By doing so, we aimed to provide a more nuanced understanding of the cognitive approach's effectiveness in L2 learning, addressing to some extent the shortcomings identified in prior research endeavors. The outcomes from the empirical studies reported here showed statistically significant benefits for the CL cohort across comprehension and production tasks at each proficiency tier. These encouraging findings elucidated the efficacy of this approach, as it facilitated the acquisition of the target linguistic forms, thereby increasing learners' abilities in communication, metaphorical expression, and ironic expression.

The linguistic phenomena comprised psych verbs such as *me gusta* 'I like', *me molesta*, 'It bothers me' (investigation 1: Martín-Gascón, Llopis-García, and Alonso-Aparicio 2023), metaphorical expressions involving *ponerse* as in *me pongo nerviosa* 'I get nervous' and *tocar* as in *toqué fondo* 'I hit rock bottom' (investigation 2: Martín-Gascón 2023a, 2024) as well as verbal irony utterances (e.g., *Claro que sí* 'Yeah, sure') (investigation 3: Martín-Gascón 2023b). These constructions exhibited several common aspects: they served as a means for expressing emotion, their acquisition in an educational

context posed a challenge for English native speakers (the target population in all studies), and their integration into the curriculum had hitherto received limited attention.

As previously discussed, over the past two decades there has been a significant surge in empirical research endeavors aimed at examining the efficacy of CL in facilitating L2 acquisition. However, these studies have met with only partial success in obtaining data favoring the CL condition. Our contention in these investigations is that this limited success can be attributed to the prevailing assessment design, which predominantly relies on conventional evaluation tasks for measuring L2 proficiency. Hence, a distinguishing feature of the QED studies presented here was the meticulous construction of assessment tests, which were intricately tailored to align with both the chosen theoretical framework and the application of CL principles within pedagogical contexts.

Also advocated in these investigations is the incorporation of a broader spectrum of psych-verbs at introductory levels, the introduction of constructions related to change-of-state and tactile experiences through heightened metaphor awareness, and the inclusion of verbal cues for irony in the L2 Spanish curriculum, all interpreted through a CL perspective. The authors also call for the practical implementation of these principles in the day-to-day L2 classroom experience and in empirical investigations evaluating the ramifications of an ACL pedagogy. The positive findings underscore the significance of adopting an ACL methodology in the realms of teaching and assessing L2 Spanish. In what follows, each QED study is examined looking at the derived strengths and limitations.

4.1. Investigation 1: psych verbs at basic levels

In the study by Martín-Gascón, Llopis-García, and Alonso-Aparicio (2023), the authors undertook an analysis to ascertain whether L2 assessment makes a difference in ACL

when learning the L2 Spanish psych-verb construction. Hence, as a novelty in the field of ACL to the L2 classroom, the pilot and main study included addressed overlooked effects of assessment typology in L2 learning. The empirical studies encompassed a pilot study ($n= 59$) and a larger-scale replication ($n= 160$), both conducted following a pretest/posttest/delayed posttest design for three research conditions (control, experimental-cognitive and comparison-traditional). Participants were enrolled in eleven intact sections and attending their first semester of an online Spanish/L2 basic course as part of their core curriculum at a university in North America. Data collection consisted of an assessment grounded in CL principles for tasks involving interpretation and production. In the interpretation task, participants engaged in activities such as matching images to linguistic forms to comprehend meaning, using motivated concepts such as experiencer/performer and stimulus/main idea. This task emphasized understanding meaning through visual representation, with a focus on form for motivated meaning rather than metalanguage, and without strict categorization of correct versus incorrect answers. Notably, it incorporated a blend of traditional linguistic concepts and CL approaches, such as distinguishing between subject and idea/performer.

Conversely, the production task required students to identify components such as experiencer, verb of emotion, and stimulus/idea within sentences and discourse, with a similar emphasis on meaning comprehension. Participants were prompted to showcase their understanding through the organization of sentences in an Object-Verb-Subject (OVS) structure, emphasizing clarity on "who's who." This task also integrated both traditional linguistic and CL concepts, ensuring accessibility for participants unfamiliar with the latter. Results evinced that, after instruction, the cognitive group demonstrated

markedly higher performance in comparison to the traditional group in both tasks – Mann-Whitney U tests for interpretation task: posttest ($Z = -8.345, p = .000$) and delayed posttest ($Z = -5.675, p = .000$) and for production: posttest ($Z = -4.895, p = .000$) and delayed ($Z = -5.168, p = .000$). These findings, in line with the authors' hypothesis, lend support to the efficacy of pairing cognitive instruction and cognitive assessment for complex grammatical forms in the L2.

4.2. Investigation 2: metaphorical constructions at intermediate level

In a subsequent work, Martín-Gascón (2023a, 2024) followed the QED approach employed in the preceding study, this time comparing the CL and traditional methodologies for teaching and learning metaphorical constructions in L2 Spanish. This investigation also aimed at fostering the development of learners' metaphoric proficiency in the L2 domain. The cognitive instruction combined cognitive parameters and insights stemming from the conceptual metaphor theory with multimodal content, while the traditional approach adhered to a communicative and formalistic paradigm grounded in prevailing L2 textbooks. A cohort of 33 A2+ students from a North American institution partook in the two experiments. Similarly, the assessments were devised in alignment with CL tenets, measuring learners' overarching metaphorical comprehension (Task 1), original metaphor production (Task 2), and performance in the comprehension (Task 3) and production (Task 4) of *ponerse* change-of-state constructions (Martín-Gascón, 2023a) and metaphorical tactile constructions with *tocar* (Martín-Gascón, 2024).

All tasks prompted students to contemplate their own embodiment, both perceptual and bodily, and to compare and contrast with their native language. Rather than emphasizing a rigid right or wrong dichotomy and focusing on metalanguage, they encouraged learners

to derive meaning from context and their own physical and perceptual experiences. Visual representation was incorporated to aid in comprehension of the target linguistic forms. Hence, exercises included emojis, GIFs, and excerpts depicting emotional scenes from popular media, such as TV shows and Almodóvar's films. These assessments also included questions designed to heighten learners' awareness of metaphorical language and to prompt reflection on their own experiential knowledge. This facilitated deeper engagement with the language and encouraged students to explore its nuances within the context of their lived experiences.

These studies followed the same pretest/posttest/delayed posttest design for the three research conditions (control, cognitive, and traditional). Results showed that an ACL methodology proves to be significantly more beneficial for all four tasks targeting at the two constructions (*ponerse* and *tocar*) under study. Although students exposed to traditional instruction exhibited progress over time, those in the CL group exhibited statistically better performance in metaphoric competence and in the comprehension and production of the target constructions. This would underscore that a CL and metaphor-based instructional and assessment approach predicts a fruitful pathway for developing metaphoric awareness and acquiring figurative constructions at intermediate proficiency levels.

4.3. Investigation 3: verbal irony at intermediate and advanced levels

The last empirical study implemented a CL-based pedagogical material to teach irony in relation to the expression of positive and negative emotions. The cohort was formed by 87 intermediate and 82 advanced students. Data collection was gathered from a 75-minute classroom session, adhering to a pretest/posttest design to retrieve ironic production and

identification. A questionnaire about linguistic background and language use was administered prior to instruction. Findings evinced a significant improvement in both ironic production and identification subsequent to the intervention, across both proficiency levels. While advanced students demonstrated notably superior performance in the production task ($Z = -2,907, p = .004$), no significant discrepancy emerged between the two groups in irony recognition tasks. These findings highlight the importance of explicitly imparting instruction in irony even at lower proficiency levels, rather than reserving it solely for C1-C2 levels as recommended by the Curricular Plan of the Instituto Cervantes and the Common European Framework of Reference for Languages, to potentiate learners' intercultural sensitivity and communicative expertise.

4.4. Strengths and methodological challenges

Overall, this collection of work contributes to the pioneering application of CL-based assessment tasks, a departure from conventional and pervading testing formats like fill-in-the-gap exercises and multiple-choice tests. This shift, while promising, engendered a significant challenge in the development of test materials, particularly in contexts of lower proficiency levels, encompassing novice learners with less than two months of exposure to the course (in the case of the psych-verb study), as well as A2+ level students (targeting at metaphorical constructions). At this early stage, venturing into more intricate and desirable tasks, such as composing full sentences or furnishing extensive texts could have been to some extent premature. In both the pilot and the main study targeting psych-verbs, tests exhibited inclusivity by incorporating references to both subject (in the traditional sense) and stimulus/main idea (cognitive), or object (traditional) and experiencer (cognitive). The terminology featured in the instructions was consistent

across these task formats. Another strength lies within the main findings: the outcomes underscore the significant efficacy of the CL instruction and assessment. Notably, for both the psych-verb construction study, involving a total of 219 participants, and ironic utterances, encompassing 169 participants, the sample size was markedly robust. This larger sample size lends weight to the generalizability of our findings, reinforcing the reliability of the observed effects.

Another methodological strength relies on the use of “pseudo-randomized” groups. Specifically, in the investigations pertaining to psych-verb and metaphorical construction, three distinct groups were employed, each pre-existing at the onset of the academic year at their universities. The sampling criteria were carefully devised to maximize homogeneity, ensuring that all students were positioned at the same proficiency level, with students scoring equal to or above 55% in the pretest being excluded from further data analysis. Last, in terms of time allocation, the design of the tests, encompassing pre, post, and delayed assessments, was calibrated to be completed within a 20-minute timeframe each. This temporal framework was rooted in pedagogical practicality, aligning with the standard duration of a classroom session in L2 Spanish at the participating university, which spans 75 minutes. It is noteworthy that the instructional phase was relatively short, primarily because the target linguistic forms—comprising each pedagogical intervention i) positive and negative psych-verbs, ii) metaphorical constructions featuring *ponerse* and *tocar*, and iii) ironic utterances—did not require an extended duration for effective delivery.

In the pursuit of advancing pedagogical research within ACL, it is essential to reflect upon the challenges that emerge during empirical investigations. In what follows, we reflect upon the negative aspects encountered during the three investigations and

highlight critical aspects that demand consideration in the context of future research. Conducting QED research embraces a range of intricate logistical considerations, among others, the need for oversight by an ethics committee, obtaining informed consent from students, and ensuring sustained participation throughout the study. Furthermore, the design of the study must navigate potential interference with the regular curriculum of both the researchers' own classes and those of other teachers involved in the research with control groups. One notable challenge encountered pertains to the sample size for the study involving metaphorical constructions with the verbs *ponerse* and *tocar*, which was notably restricted, totaling only 33 participants. The protracted nature of teaching interventions and data collection resulted in non-attendance and participant withdrawals, which reduced the initial samples for all studies, and especially for this one. Additionally, in the case of the QED study focusing on irony, there were inherent issues with the sample composition. Participants exhibited a diverse linguistic background, with some being heritage speakers while others self-identified as English-Spanish bilinguals within advanced proficiency groups. Furthermore, in this study, the absence of a delayed posttest attributed to time constraints imposed by curricular demands, and the absence of a comparison group hindered the contrast of cognitive versus more traditional instructional approaches.

Another pivotal aspect for consideration is the design of the evaluation tasks. All participant groups could have undertaken a unified assessment task featuring a balanced assortment of diverse item types, encompassing both traditional and ACL-based items. This, indeed, represents a promising avenue for future research. However, the researchers considered it imperative to acknowledge the cognitive orientation of the evaluation design. The premise underlying this approach is the need to first establish, through

rigorous examination, that the assessment design itself is not inadvertently skewing results in favor of one instructional approach over another. The deliberate emphasis on ACL serves to substantiate this hypothesis. Previous studies have predominantly favored traditional instruction in assessment, employing a binary correct-versus-incorrect framework for both traditional and cognitive groups. By inverting this paradigm and employing cognitive assessment for both groups, these empirical studies contribute to a novel methodological approach. This seeks to affirm that assessment design significantly influences the bridge between instruction and assessment. Consequently, future research endeavors should aim at adopting more balanced arrays of assessment tasks, comprising both cognitive and notional-functional components, to afford both instructional approaches equal opportunities for evaluation.

In considering the broader limitations of these studies, it is crucial to acknowledge the pervasive impact of the COVID-19 pandemic. The necessity to conduct all studies within an online learning environment introduced a series of challenges. Prolonged periods of distance learning engendered class burnout, leading to participant withdrawal and, consequently, extended data collection periods. These factors collectively underscore the multifaceted challenges inherent in conducting QED research within the constraints of a rapidly evolving educational landscape impacted by external exigencies. Table 1 provides a summary of the above-mentioned guidelines for conducting effective QED research in effects-of-instruction studies.

1. Use CL-based assessment that matches the CL-based instruction
2. Ensure terminology inclusivity (trad. and CL) and consistency in test design

3. Ensure robust sample sizes for generalizability and reliability of the observed effects
4. Use pseudo-randomized groups
5. Maximize homogeneity ensuring proficiency (class level + pretest score)
6. Follow a pre-post-delayed test design
7. Manage time allocation aligning with real-classroom experience

Table 1. Guidelines for QED in ACL.

5. CONCLUSIONS AND FURTHER RESEARCH

This study has reviewed previous relevant investigations that have sought to demonstrate the effectiveness of an ACL approach in L2 pedagogy. It has also provided illustrations from previous work by the author and colleagues that align with this body of research. We have also addressed critical factors that may impact findings in quasi-experimental research, including aspects such as adequate randomization, the absence of control and/or comparison groups, small sample sizes, varied treatment durations, and assessment methodologies encompassing both quantitative (e.g., the absence of delayed posttests) and qualitative (e.g., tests lacking a cognitive-based nature).

It is noteworthy that the empirical studies within ACL remain somewhat limited, particularly those exhibiting positive outcomes for the cognitive condition. We have aimed at exploring the reasons underlying this scarcity, taking into consideration these influential factors. In line with our findings and in alignment with Llopis-García's (2022, in press) claims, we assert the importance of rigorously examining the assessment design

itself to ensure it does not inadvertently favor one instructional approach over another. This study places deliberate emphasis on ACL to substantiate this hypothesis. Traditionally, assessments have favored conventional instructional methods, employing a binary correct-versus-incorrect framework for both traditional and cognitive groups. By inverting this paradigm and implementing cognitive assessment for experimental groups, our empirical studies introduce a novel methodological approach. This innovative approach seeks to underscore the significant influence of assessment design on the connection between instruction and evaluation. As such, this study opens up new avenues for further research and exploration in this critical area of ACL and L2 (Spanish) pedagogy and assessment.

REFERENCES

- Achard, M., & Niemeier, S. (Eds.) (2004). *Cognitive linguistics, second language acquisition, and foreign language teaching*. Berlin: Mouton de Gruyter.
- Alonso-Aparicio, I., & Llopis-García, R. (2019). La didáctica de la oposición imperfecto/perfecto simple desde una perspectiva cognitiva. In *Lingüística cognitiva y español LE/L2* (pp. 274-299). Routledge.
- Boers, F. (2013). Cognitive linguistic approaches to teaching vocabulary: Assessment and integration. *Language Teaching*, 46(2), 208-224.
- Campbell, D. T., & Stanley, J. C. (2015). *Experimental and quasi-experimental designs for research*. Ravenio books.
- Colasacco, M. A. (2019). A cognitive approach to teaching deictic motion verbs to German and Italian students of Spanish. *International Review of Applied Linguistics in Language Teaching*, 57(1), 71-95.

- Cook, T. D., & Wong, V. C. (2008). Empirical tests of the validity of the regression discontinuity design. *Annales d'Economie et de Statistique*, 127-150.
- De Knop, S., & De Rycker, T. (Eds.). (2008). *Cognitive approaches to pedagogical grammar*. Berlin: De Gruyter Mouton.
- De Knop, S., Boers, F., & De Rycker, A. (Eds.). (2010). *Fostering language teaching efficiency through Cognitive Linguistics*. Berlin: De Gruyter Mouton.
- Gass, S. M., & Mackey, A. (2016). *Stimulated recall methodology in Applied Linguistics and L2 research*. Taylor & Francis.
- Hamrick, P., & Sachs, R. (2018). Establishing evidence of learning in experiments employing artificial linguistic systems. *Studies in Second Language Acquisition*, 40(1), 153-169.
- Hung, B. P., Vien, T., & Vu, N. N. (2018). Applying Cognitive Linguistics to Teaching English Prepositions: A Quasi-Experimental Study. *International Journal of Instruction*, 11(3), 327-346.
- Ibarretxe-Antuñano, I., Cadierno, T., & Castañeda Castro, A. (Eds.). (2019). *Lingüística cognitiva y español LE-L2*. New York: Routledge.
- Kirk, R. E. (2009). Experimental design. *Sage handbook of quantitative methods in psychology*, 23-45.
- Littlemore, J., & Juchem-Grundmann, C. (Eds.). (2010). *Applied cognitive linguistics in second language learning and teaching*. Amsterdam: John Benjamins.
- Llopis-García, R. (2019). Gramática cognitiva y selección modal en la enseñanza del español LE/L2, In I. Ibarretxe-Antuñano, T. Cadierno, & A. Castañeda Castro (Eds.), *Lingüística cognitiva y español LE/L2*, pp. 255–273. New York: Routledge/Taylor & Francis Group.

- Llopis-García, R. (2019). Gramática cognitiva y selección modal en la enseñanza del español LE/L2, In I. Ibarretxe-Antuñano, T. Cadierno, & A. Castañeda Castro (Eds.), *Lingüística cognitiva y español LE/L2*, pp. 274–299. New York: Routledge/Taylor & Francis Group.
- Llopis-García, R. (2022). Plenary: Lingüística Cognitiva Aplicada: Fundamentos teóricos e investigación empírica en ELE. Presented at the Conference of the Association for the teaching and learning of Spanish in Higher Education in the United Kingdom (ELE-UK), University of Warwick.
- Llopis-García, R. (2023). Compatibilidad de la lingüística cognitiva aplicada con la enseñanza de ELE. Apuntes desde el aula. *Revista Española de Lingüística*, 53(1), 17-40.
- Llopis-García, R. (in press). *Applied Cognitive Linguistics and L2 instruction. Elements in Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Llopis-García, R., & Hijazo-Gascón, A. (Eds.). (2019). Special issue: Applied Cognitive Linguistics to L2 acquisition and learning: Research and convergence. *International Review of Applied Linguistics in Language Teaching*, 57(1).
- Marsden, E. & Torgerson, C. J. (2012). Single group, pre- and post-test research designs: Some methodological concerns, *Oxford Review of Education*, 38:5, 583-616, DOI: 10.1080/03054985.2012.731208
- Martín-Gascón, Beatriz (2023a). Developing L2 learners' metaphoric competence: a case study of figurative motion constructions. *International Review of Applied Linguistics in Language Teaching*, vol. 61, no. 1, 2023, pp. 79-109. <https://doi.org/10.1515/iral-2022-0043>

- Martín-Gascón, B. (2023b). Teaching irony in the Spanish/L2 classroom. *Porta Linguarum*. 39, 213-230. <https://doi.org/10.30827/portalin.vi39.24084>.
- Martín-Gascón, B. (2024). The effects of cognitive-based instruction on metaphoric competence and figurative constructions production and comprehension in intermediate Spanish learners. *Porta Linguarum*.
- Martín-Gascón, B., Llopis-García, R., & Alonso-Aparicio, I. (2023). Does L2 assessment make a difference? Testing the empirical validity of applied cognitive linguistics in the acquisition of the Spanish/L2 psych-verb construction. *Language Teaching Research*, 0(0). <https://doi.org/10.1177/13621688231190981>
- Nunan, D. (1992). *Research methods in language learning*. Cambridge University Press.
- Piquer-Píriz, A. M. (2021). *Un recorrido por algunas de las principales aportaciones de la lingüística cognitiva a la enseñanza de lenguas extranjeras*. AALiCo Online Conferene.
- Plonsky, L. (2017). Quantitative research methods. In *The Routledge handbook of instructed second language acquisition* (pp. 505-521). Routledge.
- Pütz, M., Niemeier, S. & Dirven, R. (Eds.). (2001). *Applied Cognitive Linguistics, II, Language pedagogy*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110866254>
- Robinson, P. J., & Ellis, N. C. (Eds.). (2008). *Handbook of cognitive linguistics and second language acquisition*. New York: Routledge.
- Rogers, J., & Revesz, A. (2019). Experimental and quasi-experimental designs. In *The Routledge handbook of research methods in applied linguistics* (pp. 133-143). Routledge.
- Spada, N., & Lightbown, P. M. (2019). Second language acquisition. In *An introduction to applied linguistics* (pp. 111-127). Routledge.

Tyler, A., Mueller, C. M., & Ho, V. (2010). Applying cognitive linguistics to instructed L2 learning: The English modals. *AILA Review*, 23(1), 30-49.

Tyler, A., Mueller, C., & Ho, V. (2011). Applying Cognitive Linguistics to Learning the Semantics of English to, for and at: An Experimental Investigation. *Vigo International Journal of Applied Linguistics*, (8), 180-205.

Zabor, E. C., Kaizer, A. M., & Hobbs, B. P. (2020). Randomized controlled trials. *Chest*, 158(1), S79-S87.