

Computerized dynamic assessment of implicature comprehension in L2 Chinese

Tianyu Qin, *University of North Georgia*

Rémi A. van Compernelle, *Carnegie Mellon University*

Abstract

The focus of this article is on the design and administration of an online computerized dynamic assessment of an important area of pragmatics: implicature comprehension (i.e., indirect meaning). The assessment centered on indirect acceptances, refusals, and opinions. Drawing on Vygotskian psychology, this assessment had a dual aim: (a) to evaluate test-takers' current capacities as evidenced by independent performance and (b) to evaluate test-takers' emerging capacities as evidenced by their responsiveness to assistance, or mediation, during the test. Sixty-seven US university learners of Chinese as a foreign language, evenly distributed across beginning, intermediate, and advanced course enrollment levels, took the assessment. The results show that while all test-takers were able to improve with assistance, there was a large degree of individual variation in responsiveness to mediation. In line with the notion of the zone of proximal development, we interpret our findings as evidence that the performances observed with mediation are more informative as a diagnosis of learner capability than independent performance alone. Our discussion therefore focuses on the implications of our testing approach for developing individual teaching and learning plans for test-takers, as well as directions for future research.

Keywords: *Dynamic Assessment, Implicature Comprehension, Chinese as a Foreign Language, Pragmatics*

Language(s) Learned in This Study: *Chinese*

APA Citation: Qin, T., & van Compernelle, R. A. (2021). Computerized dynamic assessment of implicature comprehension in L2 Chinese. *Language Learning & Technology*, 25(2), 55–74.
<http://hdl.handle.net/10125/73433>

Introduction

Assessment in second language (L2) research and practice “refers to the act of collecting information and making judgments about a language learner’s knowledge of a language and ability to use it” (Chapelle & Brindley, 2002, p. 268). In most approaches to assessment, relevant information about a learner’s knowledge of and/or ability to use a language is limited to what he or she is capable of doing in unassisted, or solo, performance. An alternative approach, known as dynamic assessment (DA) (see Poehner, 2008; Lantolf & Poehner, 2014), integrates instructional intervention as part of the procedure as a means of probing how far a learner is able to go beyond their individual capacity. Consequently, DA aims to collect information and make judgments not only about what a learner has already achieved, as evidenced by solo performance, but more importantly about what capacities are in the process of formation, as evidenced by a learner’s responsiveness to assistance.

In this paper, we report the findings of a study that used a computerized dynamic assessment (C-DA) instrument to evaluate pragmatic comprehension in Chinese as a second/foreign language (L2). Following the work of Poehner and Lantolf (2013) and Poehner et al. (2015) on listening and reading comprehension, the C-DA instrument offered assistance in the form of graduated prompts (i.e., from implicit to explicit help) when participants encountered difficulty during a multiple-choice test. The procedure resulted in the

calculation of three different test scores: (a) an unmediated, or “actual,” score based on first responses only (i.e., no assistance); (b) a mediated score that included additional points earned with assistance; and (c) a learning potential score (LPS) (Kozulin & Garb, 2002) that represents responsiveness to assistance by accounting for the difference between mediated and actual scores. We chose to focus on the comprehension of implicature (i.e., indirect meaning) since several studies have shown that this is a challenging skill to develop in L2 (Taguchi, 2005, 2011; Taguchi et al., 2013). Our study therefore extends previous C-DA work from general comprehension skills to the specific domain of pragmatics. We also hope to contribute to the L2 pragmatics assessment literature (e.g., Ross & Kasper, 2013) by illustrating the potential of C-DA to provide in-depth diagnoses of learner capacities (e.g., implicature comprehension), including those that are in the process of developing.

Implicature Comprehension in a Second Language

Implicature refers to the indirect, or implied, meaning of an utterance (Grice, 1975). Factors such as expected level of politeness can impact the ways in which speakers shape their utterances, including by using indirect language. For example, the utterance “I have a lot of work to catch up on” can be an indirect refusal of an invitation to go out for a drink. Implicature comprehension can be challenging for L2 learners because they not only have to understand the words spoken but also, and more importantly, what they imply in context.

Research in interlanguage pragmatics (Bouton, 1992, 1994; Roever, 2005; Taguchi, 2005, 2008; Taguchi et al., 2013; Taguchi et al., 2016) has identified several linguistic and learner factors that influence implicature comprehension. For example, Taguchi (2005, 2011) has reported that not all types of implicature are equally accessible to L2 learners. Indirect refusals usually follow predictable conventionalized patterns (e.g., providing an excuse to refuse an invitation) and are therefore easier to comprehend than indirect opinions, which tend to be more idiosyncratic. For instance, “You should run a restaurant!” could be interpreted as a positive or a negative opinion of one’s cooking abilities, depending on tone of voice, the assumed shared knowledge of how a dish turned out, and so on (example adapted from Taguchi et al., 2013). Taguchi (2005) reported that less conventional implicatures, namely, indirect opinions, led to lower accuracy rates and slower comprehension speed than the conventional indirect refusals. Taguchi et al. (2013) found that L2 Chinese learners received higher scores on indirect refusals, which was in line with the results in L2 English, but the response times did not differ significantly across implicature types. In addition, Taguchi (2008) also reported a significant effect of item type on accuracy but not on comprehension speed among L2 Japanese learners.

Proficiency level is positively correlated with comprehension accuracy (Taguchi, 2008; Taguchi, 2011; Taguchi et al., 2013). However, the findings are mixed with respect to study abroad (SA) experience. Taguchi (2008b) found L2 English learners with SA experience made significant improvement on comprehension speed but not on accuracy of comprehension. Taguchi (2011) reported no effect of SA experience on response times; however, she identified SA students’ advantage in the comprehension of non-conventional implicatures and routine expressions but not in indirect refusals. Finally, Taguchi et al. (2013) compared L2 Chinese learners and Chinese heritage speakers. They found that Chinese heritage speakers outperformed their counterparts on comprehension accuracy but not on comprehension speed.

While previous studies have evaluated L2 learners’ already developed abilities as evidenced by solo performance on tasks, the focus of dynamic assessment is to provide a dual assessment of current (i.e., independent) abilities and those that become possible with assistance. Provoking development during an assessment through assistance allows test administrators and researchers to gain insight into capacities that are currently in the process of formation. Our aim in the present study is to extend L2 implicature comprehension research by assessing learners’ abilities that are still developing and only observable with assistance.

Dynamic Assessment

Zone of Proximal Development

Dynamic assessment (DA) is predicated on Vygotsky's (1978) concept of the zone of proximal development (ZPD), often described as the difference between what a person can do alone and what becomes possible when help is provided. Vygotsky argued that assessments needed to include both kinds of information because they revealed different, but related, aspects of development. The first kind of information, what can be referred to as the zone of *actual* development (ZAD), links a person's past developmental trajectory to the present—essentially, what the person is currently able to do without further assistance. The second kind of information (i.e., the ZPD) is suggestive of the person's immediate, or *proximal*, future—essentially, what the person is becoming able to do but only with assistance at the moment.

Most current L2 DA work derives from the analysis of graduated and contingent feedback (i.e., feedback that is as implicit as possible and only available when needed) presented by Aljaafreh and Lantolf (1994). The study investigated the negotiation of corrective feedback in the dialogic activity co-constructed by L2 English learners and their tutor. Learners were given help during the one-on-one tutorial sessions on the errors they made in the in-class essays. The tutor did not determine the corrective feedback beforehand since corrective procedures in the ZPD must be negotiated between the expert and the novice (Aljaafreh & Lantolf, 1994). The authors' analysis showed how the tutor began with rather implicit, strategic forms of assistance (e.g., "Is there anything wrong with the sentence?") but would provide more targeted, explicit help as needed.

Aljaafreh and Lantolf (1994) applied two criteria to determine learners' microgenetic growth: (a) product-oriented criterion showing signs of improvement, and (b) learner change from other-regulation to self-regulation during the negotiation. The change from other-regulation to self-regulation is what Vygotsky (1978, p. 57) considered to be a form of internalization, a developmental process during which interpersonal processes are transformed into intrapersonal ones. In particular, the second criterion concerning learner change was reflected in the *frequency* and *quality* of help elicited from the tutor to deal with similar errors in subsequent episodes. According to the authors, this data showed whether learners have moved from primarily relying on the tutor to relying on themselves.

Approaches to Mediating Performance in DA

Sternberg and Grigorenko (2002) distinguished two general approaches to integrating mediation in DA: a 'cake format' and a 'sandwich format.' In the cake format, mediation is integrated throughout the test, like layers of a cake, and test-takers have continuous access to support when it is needed. By contrast, in the sandwich format, mediation is provided, or sandwiched, between a pretest and a posttest of independent performance (e.g., as a teaching session). In this format, the assessment focuses on learner responsiveness to mediation as a whole in terms of its 'effect' on posttest performance.

Expanding on Sternberg and Grigorenko's (2002) discussion of DA procedures, Lantolf and Poehner (2004) have noted that mediation—however integrated into the test—may be provided in two ways. One method is dialogical in nature, meaning that assistance is open-ended and negotiable during the assessment. It also may be done in a pre-scripted standardized form (e.g., graduated prompts). Open-ended dialogic mediation has the advantage of being acutely sensitive to learners' ZPDs since the test-proctor, or mediator, is free to pursue any and all means of supporting the learner. However, the approach is difficult to scale up because it is labor intensive. By contrast, mediation that is standardized for all learners (e.g., pre-scripted prompts), while less sensitive to individual ZPDs, is scalable since the mediation can be scripted in advance and used for all test takers. In addition, it is easily automated as in the case of a computerized DA instrument (Poehner & Lantolf, 2013).

In L2 settings, the cake format of DA has been conducted in the evaluation of speaking ability (Antón, 2009; Davin, 2013; Poehner & Lantolf, 2005), listening comprehension (Poehner & Lantolf, 2013; Poehner et al., 2015), reading comprehension (Davin et al., 2014; Poehner & Lantolf, 2013; Poehner et al., 2015), grammatical competence (van Compernelle & Zhang, 2014), sociolinguistic development (van Compernelle, 2014), and pragmatic competence (van Compernelle, 2013; van Compernelle & Kinginger, 2013). Most of these studies have presented in-depth analyses of one-on-one interactions between expert and novice, while the idea of providing standardized graduated feedback has only been operationalized in a few studies (Davin et al., 2014; Poehner & Lantolf, 2013; Poehner et al., 2015; van Compernelle & Zhang, 2014). In addition, the four standardized DA studies all reported within-task design, and their results showed quantified evidence of L2 learners' different actual and mediated performance, suggesting their decreasing reliance on support within individual DA tasks.

Poehner and Lantolf (2013) was an initial attempt to use standardized DA procedures via an online tool (see also Poehner et al., 2015), focusing on reading and listening comprehension in L2 Chinese, French, and Russian. The C-DA test used a multiple-choice response format and standardized mediation was provided when test-takers encountered difficulty. The mediation design followed Aljaafreh and Lantolf's (1994) graduated and contingent feedback, but in a standardized format to make the results scalable. Test-takers had four attempts to answer each question. The initial attempt was unassisted, which evaluated learners' independent performance. If the initial attempt failed, the first prompt offered the most implicit assistance (i.e. test-takers were told the answer was incorrect and to try again, the reading test highlighted a smaller part of the text while the listening test replayed a chunk of the text). If test-takers failed to respond correctly again, the second prompt, which was more explicit than the preceding one (i.e. the search space was further narrowed), was provided. The last prompt in each item was the most explicit, and test-takers either selected the correct answer or were unable to respond correctly. An optional explanation was offered for the incorrect response after all four attempts. Poehner and Lantolf also included transfer items in their C-DA test. They claimed that the transfer tasks followed the same principles as preceding ones but were more difficult and complex in order to track whether learners had internalized the mediation offered earlier.

Extending current standardized DA research to L2 pragmatics, the present study applied a cake format of DA that comprised of two parts operationalized in terms of target speech act. Part one focused on indirect acceptances and refusals, while part two (transfer) focused on indirect opinions. This research design aimed to evaluate whether, and to what extent, participants were able to handle implicature items among different contexts with the implementation of DA procedures.

Transfer in DA

As Vygotsky (1978) argued, development involves going beyond improvement on assessment tasks to include the recontextualization of abilities to novel and/or more complex situations (see also Feuerstein et al., 1979; Poehner, 2007). Transfer in DA therefore refers to a test-taker's ability to deploy newly formed or improved abilities to more difficult problems. For example, the C-DA test reported in Poehner and Lantolf (2013) included five reading comprehension transfer items that were more difficult in terms of linguistic demands than the preceding test items. Such items are important components in DA as they allow researchers to diagnose learner change when similar principles in preceding tasks are applied in different and/or more complex contexts. Thus, researchers are able to investigate whether novel abilities that emerged in earlier tasks have been internalized by L2 learners and whether those newly formed functions can be recontextualized in new and potentially more challenging situations. Transfer in the current DA design was operationalized as comprehending a new speech act, indirect opinions, in the second half of the test that was more challenging (i.e., because of increased idiosyncrasy, fewer conventionalized strategies) than the speech act of indirect refusals encountered in the first half of the test.

Design Principles for C-DA

Extending the work of Poehner and Lantolf (2013) and Poehner et al. (2015), three design principles informed the development of the C-DA instrument used in this study. In this section we provide only a sketch of the principles. Specific details of the test design are provided in later sections.

The first principle is common to all dynamic assessment approaches: support provided in the test needs to be *contingent on learner need* and *graduated in explicitness* (see Aljaafreh & Lantolf, 1994). Test-takers should receive support only when it is necessary to help them accomplish the task at hand. In turn, the form of support they are provided during the test should be implicit at first, and should only become more explicit if test-takers are unable to benefit from a more implicit form of assistance. Contingent assistance links to Vygotsky's (1978) ZPD concept: Test administrators are able to distinguish independent abilities from those that become possible with assistance. Graduated assistance goes a step further in terms of diagnosing learner abilities, the assumption being that an inverse relationship exists between a learner's need for explicit assistance and his/her proximity to independent performance.

The second principle focuses on the idea of *transfer* or *recontextualization*. As noted earlier, dynamic assessment is concerned with responsiveness to mediation, including one's ability to recontextualize newly formed or further developed abilities to new contexts and/or more difficult tasks (Feuerstein & Rand, 1974; Feuerstein et al., 1979; for L2, Poehner, 2007; Poehner & Lantolf, 2013). This helps to distinguish test effects that have limited scope (e.g., mastery of a specific task) from broader developmental processes (e.g., internalization of new concepts or metacognitive strategies). Consequently, C-DA instruments should incorporate more difficult items or tasks at the end of the test in order to determine the extent to which gains made earlier in the test transfer to new contexts.

The third principle is of technical importance in computerized/technology-mediated environments. The goal of C-DA is to deploy dynamic assessments to large numbers of learners while minimizing the labor intensity of administering and scoring/rating a test. Therefore, the instrument must be in a format that automatically administers the test (i.e., displaying tasks/items but also providing contingent and graduated assistance) and also scores and records test-taker responses. Multiple-choice tests are one example. Although this principle limits the kinds of capacities we can assess through C-DA—open-ended written or oral responses cannot be easily scored, for instance—we hasten to remind the reader that we are not proposing C-DA as the solution to all problems. Instead, we advocate using C-DA as a means of conducting standardized assessments at scale while at the same time capturing ZPD-relevant information that may be missed in a non-dynamic administration of an assessment.

Why C-DA? Contribution to Technology-Enhanced Language Assessment

In addition to contributing to the general pragmatics and dynamic assessment literature, this study intersects with the growing field of technology-enhanced language assessment. As Chapelle and Voss (2016) have pointed out, computer and Internet technologies have the potential to make assessment processes more efficient as well as to lead to assessment innovations. The C-DA instrument used in the present study was developed with both of these contributions in mind.

As noted, a major argument in support of dynamic assessment in general has been that it provides an expanded evidential basis for understanding learner capability that includes what becomes possible with support. This means that we can glean insight into the learner's immediate future (i.e., Vygotsky's ZPD concept). However, conducting dynamically administered assessments in face-to-face contexts can be time consuming and labor intensive. By contrast, C-DA can increase efficiency because mediation is pre-programmed as part of the test and does not require a human administrator to be co-present (Poehner & Lantolf, 2013). The online delivery of the instrument also means that it is scalable (in principle, the number of test-takers is constrained only by bandwidth and server limitations) and time and place independent (it

can be administered to anyone at any time in any part of the world that has access to the Internet). Our own interests have focused on developing C-DA tools for teachers to use with their classes in a way that distributes assessment and instructional labor efficiently (e.g., assigning C-DAs as homework as part of a diagnostic process that informs the teacher's lesson planning and course adaptations).

C-DA technologies also certainly have innovative potential in comparison to non-computerized forms of assessment. For example, although we have not integrated this aspect into the current report, the C-DA instrument collected information about response processes. This included which responses were selected in which order for each test-taker and time elapsed for the test as a whole, as well as for each item and between first and subsequent responses to a single item (if the first was incorrect). Such information could be used by teachers or test administrators to gain insight into possible sources of difficulty (e.g., long intervals between response option selection on a particular item).

Methods

The study focuses on test-takers' performances on a C-DA of implicature comprehension in L2 Chinese. We reported independent and mediated performance scores as well as a so-called "learning potential score" (Kozulin & Garb, 2002) that represents test-takers' responsiveness to mediation during the assessment. Together, the scores allow us to ascertain the difference between what the test-takers could do independently (i.e., ZAD) and what became possible with support (i.e., ZPD).

Recruitment and Participants

For recruitment, the URL for the test was distributed via email to several instructors of Chinese (all personal or professional contacts of the first author) from nine private and public US universities with a request to forward the recruitment email to their students. Participants received a \$10 Amazon gift card upon completion of the study. The responses were anonymous, so respondents were not personally identifiable. The online test was open from June 03, 2016 to June 19, 2016.

In total, 85 individuals completed the assessment. Of these, nine were deleted because they were identified as repeated submissions from the same participant. Of the 76 remaining participants, seven reported having a first or dominant language other than English. They were excluded from the final analysis in order to minimize the potential influence of first/dominant language in the study. Two additional participants were excluded because they reported having no formal instruction in Chinese.

The remaining 67 participants (43 females and 24 males) included in the study were all adult university-level learners of Chinese. Their ages ranged from 18 to 30, with a mean age of 20.4 years. 32.8% ($n = 22$) had completed elementary level Chinese courses, 31.3% ($n = 21$) had completed intermediate level courses, and 35.8% ($n = 24$) had completed advanced level courses. In addition, 26.9% of the participants ($n = 18$) had never been to a Chinese-speaking region, while 73.1% ($n = 49$) had visited or lived in a Chinese-speaking region. Among these participants, five did not provide clear and countable information about their exact length of stay and thus were excluded from this part of the analysis. The remaining 44 participants' length of stay ranged from 0.5 to 84 months, with an average length of 15.5 months. We note that 13 participants reported having lived in a Chinese-speaking country for more than two years, three had visited or lived for one year, and the remaining 28 had stayed in a Chinese-speaking country for six months or less. 41.8% of the participants ($n = 28$) reported speaking or using Chinese (including Chinese dialects) at home. They were classified as heritage Chinese learners following He's (2011) definition of the concept as involving people who use Chinese at home at least some of the time (e.g., with family).

Instruments

Two instruments, a Chinese pragmatics listening DA test and a background questionnaire (see [Appendix A](#)), were developed and administered in an online computerized format. Participants first completed the

DA session, designed to evaluate their ability to comprehend indirect speech acts both independently and with the help of standardized assistance. They then filled out the background questionnaire, designed to collect information on learner variables including native language, institutional level, length of stay in Chinese-speaking countries, and heritage learning context.

Test Instrument

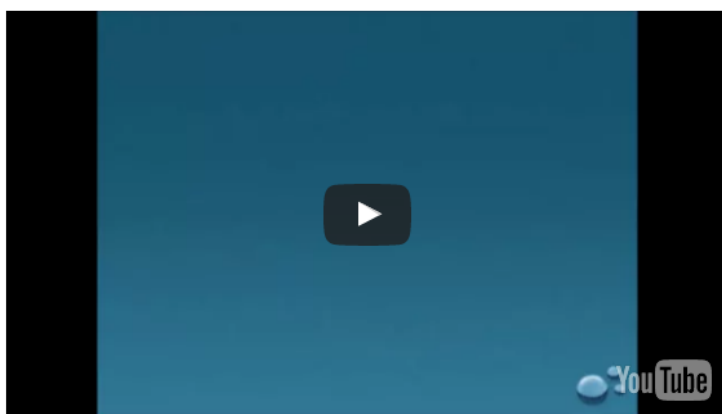
The C-DA instrument was designed following Poehner and Lantolf's (2013) model. Test items were adapted from Taguchi et al.'s (2013) study of indirect speech acts in L2 Chinese. The first part of the test focused on indirect refusals and acceptances (six items each for a total of 12). The second part of the test (i.e., transfer) focused on indirect opinions, of which half were negative opinions and half were positive opinions. Both parts of the test were multiple-choice format. In addition to the 12 target items in each test, we included four non-target items. The four non-target items in the first part were direct speech acts of apology, invitation, suggestion, and request. The four non-target items in the second part involved two direct positive opinions and two direct negative opinions. The point of having non-target items was to minimize the chances of participants guessing what the assessment was about (i.e., indirect speech acts). The two main parts of the test therefore provide some insight into test-takers' ability to transfer knowledge from one context to another (i.e., from acceptances and refusals to opinions). This is an important aspect of DA that aims to evaluate responsiveness to mediation (Feuerstein et al., 1979; Poehner, 2007; Poehner & Lantolf, 2013). A third part of the test involved open-ended responses to indirect speech acts. We will not consider this data in the present article as it is beyond the scope of our quantitative analysis.

During the test, items were presented in a randomized order. The test-taker would press a "play" button and then hear and read the context of the target speech act in English, followed by audio of the speech act in Chinese in oral form only. [Figure 1](#) shows a sample context description and audio prompt from the test.

Figure 1

Context Description and Audio Prompt

Mike plans to travel to Beijing. He knows his high school classmate, Lucy, is a graduate student there. Mike calls Lucy and asks if he can stay in her dorm for a few nights. Lucy says,



Note. The corresponding target utterance for this item was 我们宿舍人很多，而且已经住满了 “There are many people living in my dorm, and it is already fully occupied.” After listening to the audio file only once, participants were asked to move to the next page where the multiple-choice question (MCQ) was presented ([Figure 2](#)).

Figure 2

Sample Multiple-Choice Question and Response Options

*** Question 1: What does Lucy mean?**

- Mike can't stay in Lucy's dorm.
- Lucy's dorm is already fully occupied.
- Lucy invites Mike to visit Beijing.
- Lucy's dorm is very small.
- Lucy plans to have Mike stay in Beijing for a long time.

Each MCQ included five options in total: a correct choice and four distractors. One distractor was the literal meaning of the utterance (the second option in Figure 2), while the remaining three distractors were designed to appear to be reasonable responses if the test-taker had comprehended only parts of the utterance. It should be noted that while selecting the literal meaning would indicate comprehension of the locutionary force of the utterance (i.e., the words said), we did not count it as correct or give partial credit because indirect speech act comprehension requires one to infer the speaker's intent behind the words said (Taguchi et al., 2013, 2016). Thanks to the dynamic administration of the test, we were able to track changes across the test and account for response processes, including shifts from focusing on literal meaning to focusing on implied meaning, which would be a desirable outcome of the C-DA. All questions and response options were delivered in English to avoid a situation in which participants comprehended the meaning but encountered difficulties in understanding the questions and options in Chinese (see Poehner & Lantolf, 2013).

Approach to Mediation

The test followed a cake format of DA with standardized mediating prompts. In line with Aljaafreh and Lantolf (1994), mediation was designed to be graduated and contingent. That is, the assistance provided moved from the least explicit to the most explicit, and mediation was available continuously in the test only when test-takers encountered difficulty. The mediating prompts were operationalized following the C-DA model in Poehner and Lantolf (2013) as well as Davin's (2013) standardized prompts for classroom DA. Up to three prompts were available for each item, and the kinds of mediation were the same for all test-takers. If the test-taker chose the correct response on the first attempt, no prompts were presented. Instead, the test moved on to the next item. However, if the correct response was not chosen, prompts were presented to assist the test-taker. Table 1 shows the general design of the prompts.

Table 1

Prompt Design

Level of Explicitness	Prompt Design
Prompt 1	Allow participant to listen to the context and the target sentence again and then try a second time without a specific hint
Prompt 2	Allow participant to listen to the target sentence again with a hint highlighting the relevant content of the speech act (e.g., topic or vocabulary)
Prompt 3	Allow participant to listen to the target sentence again with a forced choice hint

For instance, in the multiple-choice question presented in Figure 2, the first prompt was "That is not the right answer. Listen to the clip and try again." If this failed to produce a correct response, the second prompt

focused on key vocabulary/topic: “That is not right either. Why does Lucy talk about her dorm?” If this also failed to produce a correct response, the third prompt was: “That is not right, either. Is Lucy going to allow Mike to stay with her?” Of course, we acknowledge the potential for guessing, especially as the degrees of freedom is reduced on each subsequent attempt. However, in our data, test takers tended to use fewer attempts as the test went on, suggesting that even if they had been guessing at the beginning, they eventually learned what the test was about and how to comprehend implicature. If this were not the case, we would expect to see several attempts on each item throughout the test with no change in the need for multiple attempts.

Scoring Procedures

The test generated three numerical scores: an actual score, a mediated score, and a learning potential score (LPS) (see Kozulin & Garb, 2002; Poehner & Lantolf, 2013; van Compernelle & Zhang, 2014). The actual score represented independent performance without assistance. It was calculated based on first responses only (correct answer = 4 points (pts); incorrect answer = 0 pts). The mediated score included points awarded for items correctly answered on the second (3 pts), third (2 pts), or fourth (1 pt) attempt in addition to the actual score. After the fourth attempt, the answer was provided to the test-taker and 0 points were awarded. The idea of including the actual score in the calculation of the mediated score was in line with the future-oriented ZPD concept in which learners’ mediated performance reflects their ZADs plus what becomes possible with support. The comparison of actual and mediated scores provides insight into the degree to which test-takers were able to respond positively to the graduated mediating prompts provided during the test. Following van Compernelle and Zhang (2014, p. 400), the mediated score was calculated as follows:

$$\text{Mediated Score} = \text{Actual Score} + (N_{2\text{nd}} \times 3) + (N_{3\text{rd}} \times 2) + (N_{4\text{th}} \times 1)$$

For example, a learner who answered 12 of the 24 items correctly on the first attempt (4 pts each) would receive an actual score of 48. They would then receive additional points toward a mediated score for the remaining 12 items, depending on the number of attempts needed (i.e., 3 pts for a correct answer on attempt 2, 2 pts for a correct answer on attempt 3, 1 pt for a correct answer on attempt 4, and 0 pts if attempt 4 was incorrect).

The LPS (Kozulin & Garb, 2002) reflects a learner’s responsiveness to mediation. An LPS is akin to an effect size for an individual because it accounts for the difference between what is possible with and without support. The higher the LPS, the more effective the mediating prompts were for the learner. The LPS was calculated as follows (adapted from Kozulin & Garb, 2002, p. 121; van Compernelle & Zhang, 2014, p. 400):

$$\text{LPS} = ((2 \times \text{Mediated Score}) - \text{Actual Score}) / \text{Max Score}$$

Results

Results are presented in three parts. First, we examine differences between test-takers’ actual and mediated performance. Second, we consider the relationship between actual and mediated test performance and three individual learner differences: enrollment level, heritage learner status, and time spent in a Chinese-speaking environment. Third, we report the quantitative findings on test-takers’ ability to transfer inference of indirect acceptances and refusals to the inference of indirect opinions.

Actual Versus Mediated Performance

Actual, mediated, and learning potential scores are plotted in [Figure 3](#). Actual scores ranged from 16 to 96 ($M = 72.8$, $SD = 16.3$). Mediated scores ranged from 43 to 96 ($M = 85.5$, $SD = 9.2$). LPSs ranged from 0.73 to 1.15 ($M = 1.02$, $SD = 0.07$). Note that LPSs have been multiplied by 100 for the sake of visualization in [Figure 3](#). For example, the mean LPS of 1.02 appears as 102. This makes it possible to present all three

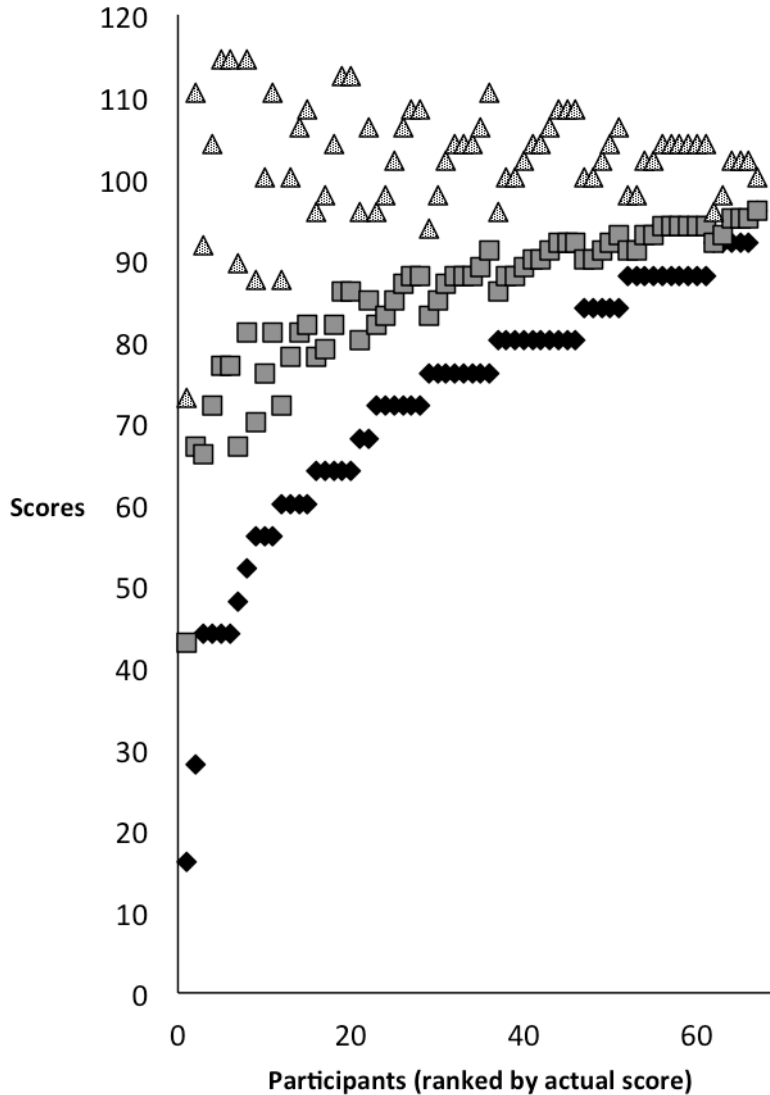
relevant scores in the same graph. The data points are arranged according to test-takers' actual scores in ascending order.

As expected, mediated scores are higher than actual scores, and the difference is significant with a large effect size ($t_{66} = 12.37, p < .001, d = 2.7$). The largest differences between the two scores tend to be among those who had relatively low actual scores (left side of [Figure 3](#)). Those who had relatively higher actual scores made smaller gains because they answered a higher percentage of items correctly on the first attempt (i.e., they had few, if any, chances to gain additional points). This is why there is a tendency for actual and mediated scores to converge on the right side of [Figure 3](#).

LPSs are highly variable from one test-taker to the next. The lowest as well as the highest LPSs were obtained by test-takers who received low actual scores (left side of [Figure 3](#)). As actual scores increase, the range of LPSs decreases and they begin to converge around the mean score (102 in [Figure 3](#)). In other words, the LPS data show that there was a high degree of variability in responsiveness to mediation during the C-DA procedure, especially among those learners who struggled in terms of independent performance. While some made large gains with mediation, and therefore earned high LPSs, others continued to struggle in spite of the assistance provided, and they received lower LPSs. The LPS data helped to reveal differences among struggling students who may appear similar with regard to solo performance but are in fact at very different places in terms of their readiness to benefit from instruction. As we explain below, this is precisely the goal of DA: To ascertain not only a person's manifest level of performance but more importantly to infer capacities in the process of formation (i.e., the ZPD). We would also like to point out that while actual and mediated scores are strongly correlated ($r = .93, p < .0001$), no such relationship exists between actual scores and LPSs ($r = .13, p = .3$). This means that a learner's readiness to benefit from instruction would not be predictable based solely on what they were capable of doing independently.

Figure 3*Actual, Mediated, and Learning Potential Scores*

(Diamond = Actual Score; Square = Mediated Score; Triangle = LPS)

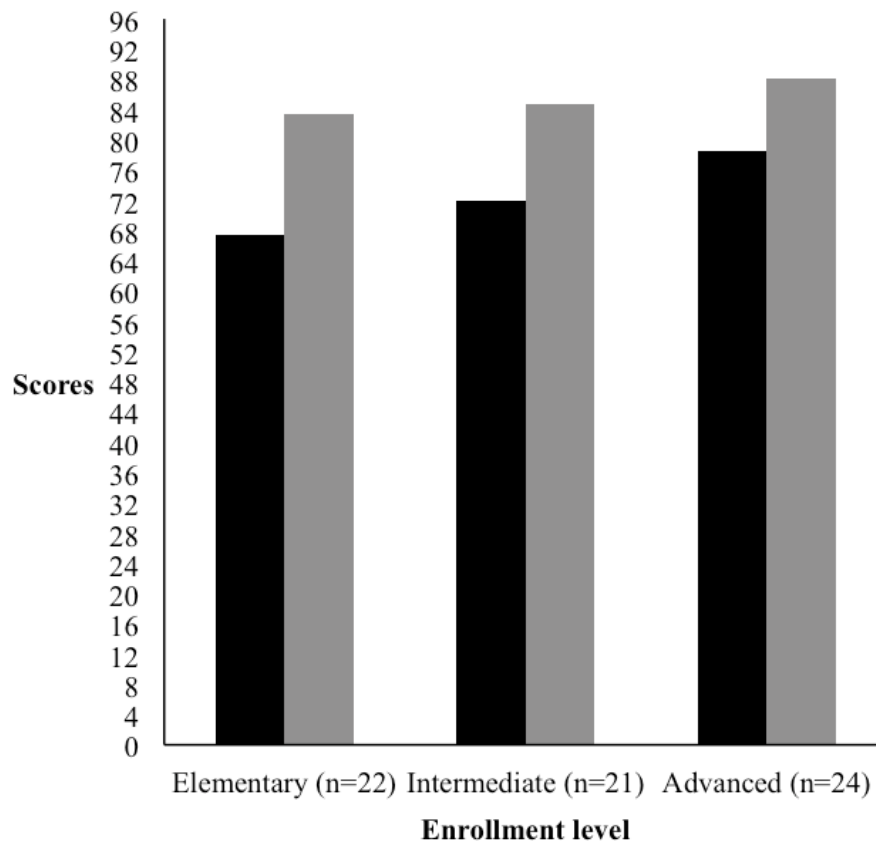
**Individual Variation**

As noted above, participants in the study were relatively diverse. They included learners at three instructional levels (i.e., reported enrollment in elementary, intermediate, or advanced Chinese course), heritage as well as non-heritage learners of Chinese, students who had never been to a Chinese-speaking country, and others who had studied or lived in one for some period of time, ranging from about two weeks to several years. Our analysis revealed that enrollment level and heritage learner status were related to test performance (see [Figure 4](#), [Figure 5](#)). However, no relationship was found between length of stay in a Chinese-speaking region and actual score ($r = -.11$, $p = .38$) or mediated score ($r = -.18$, $p = .16$).

Mean actual and mediated scores according to enrollment level are provided in Figure 4. The difference between actual and mediated scores was significant among all three groups, and the effect was large: Elementary $t_{21} = 7.56, p < .001, d = 2.86$; Intermediate $t_{20} = 8.37, p < .001, d = 2.97$; Advanced $t_{23} = 6.67, p < .001, d = 2.43$. Both types of scores increased on average with enrollment level: Advanced learners earned the highest scores, followed by intermediate learners, and then elementary learners. A significant difference was found between the actual scores of elementary and advanced learners, with a medium effect size: $t_{44} = 2.22, p = .03, d = .65$. The difference between the actual scores of elementary and intermediate learners ($t_{41} = .88, p = .38$), as well as between intermediate and advanced learners ($t_{43} = 1.61, p = .11$), did not, however, reach statistical significance. With respect to mediated scores, no differences were found between any of the groups: Elementary vs. Intermediate ($t_{41} = .42, p = .67$); Elementary vs. Advanced ($t_{44} = 1.63, p = .11$); Intermediate vs. Advanced ($t_{43} = 1.48, p = .15$).

Figure 4

Actual vs. Mediated Scores According to Enrollment Level
(Black = Actual Score; Gray = Mediated Score)



As shown in Figure 5, heritage learners tended to outscore non-heritage learners. A significant difference in actual scores with a very large effect was found: $t_{64.51} = 3.42, p = .001, d = .82$. Mediated scores were also significantly different, but the effect was somewhat smaller: $t_{62.92} = 3.06, p = .003, d = .73$. To be sure, both groups made significant gains with mediation (i.e., actual vs. mediated scores), and the effect sizes were large: Heritage $t_{27} = 7.56, p < .001, d = 2.45$; Non-heritage $t_{38} = 10.76, p < .001, d = 2.90$.

Although enrollment level and heritage learner status appear to be related to test performance, we would like to point out that there was a good deal of individual variation. [Table 2](#), for example, presents the actual, mediated, and learning potential scores of a subset of participants who earned between 44 and 60 points independently (i.e., actual score). This represents about the lower third of actual scores from [Figure 3](#). Included are learners from all three instructional levels, three of which were heritage learners of Chinese. As we see, different individuals responded to the test and mediation differently.

Figure 5

Actual vs. Mediated Scores According to Heritage Learner Status
(Black = Actual Score; Gray = Mediated Score)

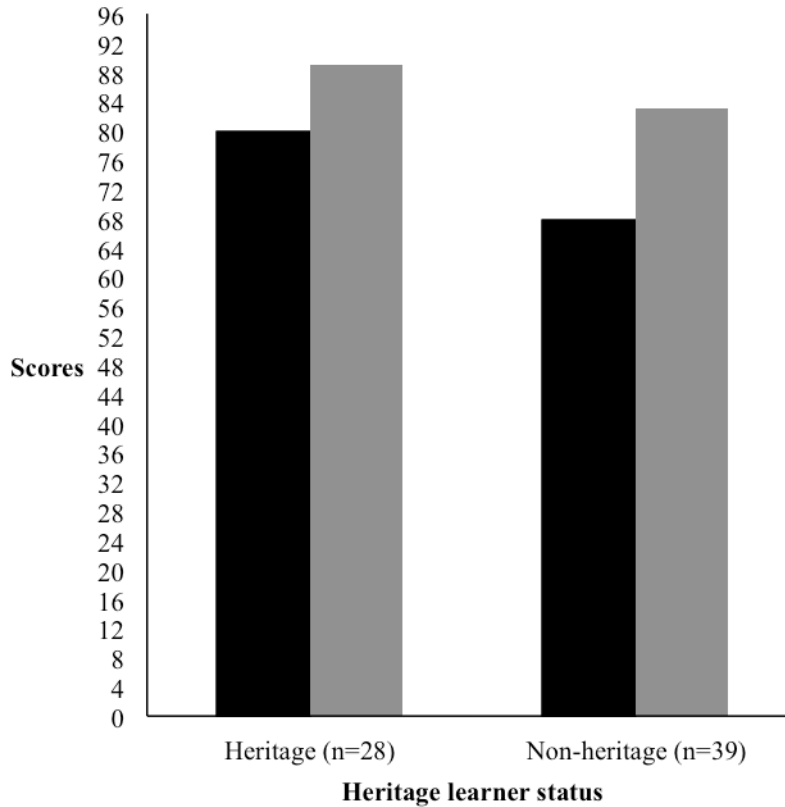


Table 2*Comparison of the Actual, Mediated, and Learning Potential Scores of a Subset of Participants*

ID	Enrollment	Heritage Status	Actual Score	Mediated Score	LPS
P10	Elementary	No	44	77	1.15
P49	Intermediate	No	44	72	1.04
P73	Elementary	No	44	77	1.15
P76	Advanced	No	44	66	0.92
P9	Intermediate	No	48	67	0.90
P58	Advanced	No	52	81	1.15
P16	Intermediate	Yes	56	70	0.88
P72	Intermediate	No	56	76	1.00
P75	Intermediate	No	56	81	1.10
P2	Advanced	Yes	60	81	1.06
P8	Elementary	Yes	60	82	1.08
P22	Elementary	No	60	78	1.00
P36	Advanced	No	60	72	0.88

For example, P10 and P73, both elementary-level students, responded very well to the mediation (LPS = 1.15), whereas P76, who was an advanced learner, did not make as much progress. This is to say that despite their identical actual scores (i.e., 44), P10 and P73 happened to be able to make more progress than P76. We can also see that while P58 had a lower actual score than P36 (i.e., 52 vs. 60), P58 responded much better to the mediation (i.e., mediated score = 81 vs. 82; LPS = 1.15 vs. 0.88). In other words, P58 was probably farther along in terms of development than P36, even though P58 earned fewer points independently than P36. This information, especially responsiveness to mediation as evidenced by the mediated score and LPS, could be useful for test administrators as well as teachers who are interested not only in group-level means but in individuals' abilities, including those whose abilities are continuing to emerge (i.e., the ZPD concept).

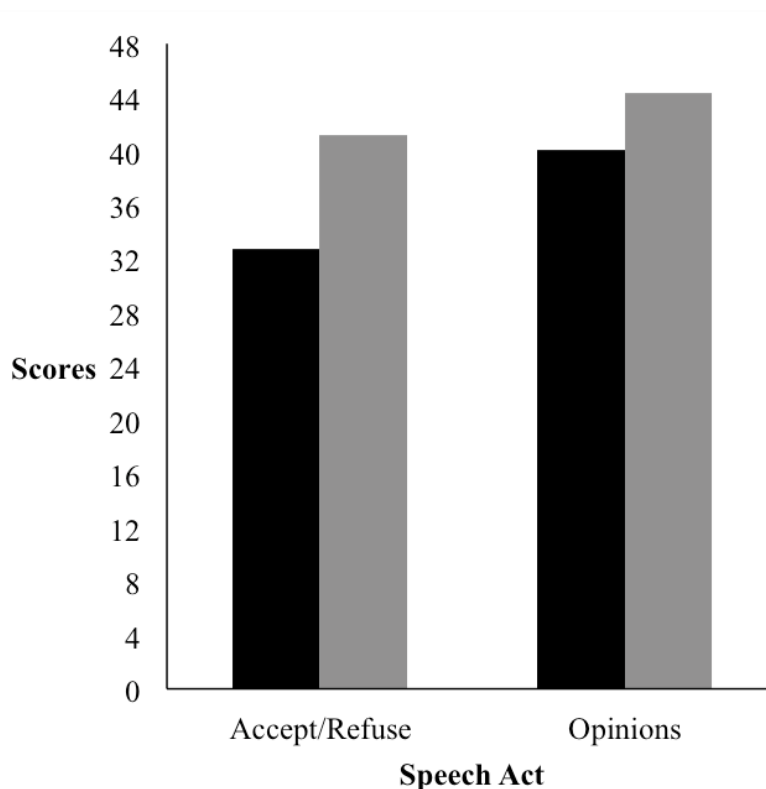
Transfer

As noted earlier, the C-DA instrument included transfer inasmuch as part one of the test centered on indirect acceptances and refusals, while part two involved indirect opinions. The reader will recall that indirect opinions are supposed to be more difficult since they follow less conventionalized linguistic patterns (Taguchi, 2005, 2008; Taguchi et al., 2013).

Figure 6 shows group-level means for actual and mediated scores in the two parts of the test. Generally speaking, test-takers struggled more with acceptances and refusals (part one) than with opinions (part two). Indeed, actual scores for opinions were significantly higher than for acceptances and refusals, and the effect size was large: $t_{66} = 8.52, p < .001, d = 1.15$. Likewise, mediated scores for opinions were significantly higher than for acceptances and refusals, and the effect size was moderate: $t_{66} = 6.07, p < .001, d = .75$. In other words, on average, test-takers significantly improved in terms of their ability to comprehend indirect speech acts over the course of the test, including across two different kinds of speech acts.

Figure 6

Actual vs. Mediated Scores According to Speech Act
(Black = Actual Score; Gray = Mediated Score)



Discussion and Conclusion

Our data have shown that test-takers were able to achieve higher scores with assistance (i.e., actual vs. mediated scores). This finding is unsurprising, of course, but it is an important starting point for understanding test-takers' responsiveness to mediation and, by extension, interpreting their readiness to learn through further instruction. Gaining an understanding of this readiness is the point of many implementations of DA.

Responsiveness to mediation, as reflected by the LPS, does not correlate with one's actual score. Rather, because the LPS is derived from the gains one makes as a result of mediation in relation to a maximum score, it is possible to have a very high LPS even if one's actual score is quite low. By contrast, a high actual score does not necessarily predict a high LPS; if one makes little progress, a low LPS will be calculated. The reason the LPS still proves useful is because it provides additional information about the extent to which test-takers were able to learn during the assessment, which is not revealed by independent performance. The LPS therefore has the potential to "be used for the development of individual learning plans for students with different learning needs" (Kozulin & Garb, 2002, p. 123). For instance, we would predict that a test-taker with a high LPS would continue to benefit from, and make further progress with, the kind of standardized graduated prompts used in the test. By contrast, a low LPS would suggest that another form of mediation might be necessary. The lowest LPS in our study (0.73) was obtained by the person who also earned the lowest actual and mediated scores. We speculate that this person made very little progress with mediation because of a lack of Chinese vocabulary and/or general proficiency, which is in contrast to the second lowest actual score earner, who made large gains and earned a relatively high LPS.

While we cannot draw many more conclusions from our data, interviews and/or stimulated recalls with test-takers could reveal response processes and specific challenges faced by individuals. Indeed, Qin (2018) has extended this work and included posttest interviews in a follow-up study. Such qualitative information (i.e., about response processes, not only scores) has the potential be used to develop learning plans for some test-takers that center on practicing indirect speech acts, while for others, learning plans might be more remedial in nature and focus on building vocabulary and grammar because the task was beyond their ZPD (van Compernelle & Zhang, 2014; Zhang & van Compernelle, 2016). We would also like to note that test-takers with LPSs around the mean and who also had very high actual scores (see the far right side of Figure 3) should be given additional consideration. Since the LPS is calculated in relation to the maximum score, there is a potential ceiling effect for high actual scorers, as Kozulin and Garb (2002) cautioned. In other words, for certain test-takers, it may be the case that the test is not sensitive to their ZPDs because it does not push them beyond what they are already capable of doing alone. Consequently, it is possible that high scorers are simply ready to move on to more challenging tasks.

The potential for ceiling effects, as well as the lack of a relationship between actual scores and LPS, highlights the importance in DA of identifying individual variation in relation to responsiveness to mediation. As reported, group-level analyses revealed significant differences in test performance (i.e., actual and mediated scores) according to enrollment level and heritage learner status. However, as we saw in Table 2, there was quite a bit of variation within groups. Some elementary-level test-takers performed, or at least responded to mediation, better than some advanced-level test-takers, and some non-heritage learners outperformed heritage learners. Developing learning plans for students, as Kozulin and Garb (2002) have advocated, has to be done on an individual basis, and not in relation to group norms. While this could certainly pose challenges for teachers, we believe adding information such as the LPS could help them to create tasks or task components that are individualized but also useful in whole-class formats. For instance, teachers who use our C-DA instrument and found individual variation among their students could design a broad learning task focused on implicature comprehension, but tailor their feedback and interactions with individual learners in ways that are sensitive to those individuals' needs. A learner with a high LPS, for example, may only receive graduated prompts as done in the test (see Davin, 2013 for an example of classroom-based DA), while another with a low LPS might receive additional explicit instruction from the teacher during the task.

The finding that test-takers generally performed better in the second part of the test (i.e., indirect opinions) than in the first part (i.e., indirect acceptances and refusals) suggests that they learned during the test and were able to transfer their knowledge to a subsequent task. We suspect that the lower scores in the first half reflect a need to learn what was being assessed and how to interpret indirect speech acts. After the initial learning phase, the test-takers were able to continue to perform even when the target speech act—indirect opinions—became more challenging. This kind of information may be useful in assessment contexts focused on responsiveness to mediation because it suggests that improvement was more than a simple training effect resulting from experience with a single kind of speech act. To be sure, this is an empirical question that requires additional research. For example, we might wish to compare the order in which the speech acts are presented (i.e., acceptances/refusal followed by opinions vs. opinions followed by acceptances/refusals) in order to see whether, or to what extent, speech act type, or conventionality of speech act forms, is a factor.

Further research in this area is certainly needed. For example, in the current test, the mediated score is derived from the actual score. Consequently, the two scores are not independent of one another. A modification of the design might include an unassisted pretest and an unassisted posttest with a mediation phase in the middle, which has been referred to as sandwich format DA by Lantolf and Poehner (2004), following Sternberg and Grigorenko (2002). This design has been successfully implemented in face-to-face L2 DA, most recently by Zhang and van Compernelle (2016) for testing L2 Chinese grammar, but it has yet to be used in a C-DA format or in the domain of pragmatics.

These are some of the directions that our own continuing research is taking currently. We hope that the current article can serve as a call for other researchers in pragmatics, but also in language testing more generally, to take up work in C-DA. We are particularly encouraged by the potential of C-DA to provide insights into individuals' readiness to benefit from instruction while it can also work at scale, something that is not possible in one-to-one interactive DA formats. For example, additional research that correlates LPS with the outcomes of an individual learning plan derived from C-DA, as suggested by Kozulin and Garb (2002), could help us to determine whether C-DA can function as a predictive form of learning potential assessment.

References

- Aljaafreh, A. & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *Modern Language Journal*, 78(4), 465–483.
- Antón, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals*, 42(3), 576–598.
- Bouton, L. F. (1992). The interpretation of implicature in English by NNS: Does it come automatically without being explicitly taught? *Pragmatics and Language Learning*, 3, 53–65.
- Bouton, L. F. (1994). Can NNS skill in interpreting implicature in American English be improved through explicit instruction? A pilot study. *Pragmatics and Language Learning*, 5, 88–108.
- Chapelle, C.A., & Brindley, G. (2002). Assessment. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 267–288). Edward Arnold.
- Chapelle, C. A., & Voss, E. (2016). 20 years of technology and language assessment in Language Learning & Technology. *Language Learning & Technology*, 20(2), 116–128. <http://lt.msu.edu/issues/june2016/chapellevoss.pdf>
- Davin, K. J. (2013). Integration of dynamic assessment and instructional conversations to promote development and improve assessment in the language classroom. *Language Teaching Research*, 17(3), 303–322. <https://doi.org/10.1177/1362168813482934>
- Davin, K. J., Troyan, F. J., & Hellmann, A. L. (2014). Classroom dynamic assessment of reading comprehension with second language learners. *Language and Sociocultural Theory*, 1(1), 1–23. <https://doi.org/10.1558/lst.v1i1.1>
- Feuerstein, R., & Rand, Y. (1974). Mediated learning experiences: An outline of the proximal etiology for differential development of cognitive functions. *International Understanding*, 9/10, 7–36.
- Feuerstein, R., Rand, Y., & Hoffman, M. B. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques*. University Park Press.
- Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (pp. 41–58). Academic Press.
- He, A. W. (2011). Heritage language socialization. In A. Duranti, E. Ochs, & B. B. Schieffelin (Eds.), *The handbook of language socialization* (pp. 587–609). Wiley-Blackwell.
- Kozulin, A., & Garb, E. (2002). Dynamic assessment of EFL text comprehension. *School Psychology International*, 23(1), 112–127.
- Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the past into the future. *Journal of Applied Linguistics*, 1(1), 49–72.
- Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice divide*. Routledge.

- Poehner, M. E. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *Modern Language Journal*, 91(3), 323–340.
- Poehner, M.E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Springer.
- Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, 9(3), 233–265.
- Poehner, M. E., & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during Computerized Dynamic Assessment (C-DA). *Language Teaching Research*, 17(3), 323–342. <https://doi.org/10.1177/1362168813482935>
- Poehner, M. E., Zhang, J., & Lu, X. (2015). Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing*, 32(3), 337–357. <https://doi.org/10.1177/0265532214560390>
- Qin, T. (2018). *Computerized dynamic assessment of L2 Chinese implicature comprehension* [Doctoral dissertation, Carnegie Mellon University].
- Roever, C. (2005). *Testing EFL pragmatics*. Gunter Narr.
- Ross, S., & Kasper, G. (Eds.) (2013). *Assessing second language pragmatics*. Palgrave Macmillan.
- Sternberg, R. J., & Grigorenko, E. I. (2002). *Dynamic testing: The nature and measurement of learning potential*. Cambridge University Press.
- Taguchi, N. (2005). Comprehension of implied meaning in English as a second language. *Modern Language Journal*, 89(4), 543–562.
- Taguchi, N. (2008). Pragmatic comprehension in Japanese as a foreign language. *Modern Language Journal*, 92(4), 558–576.
- Taguchi, N. (2008b). Cognition, language contact, and the development of pragmatic comprehension in a study-abroad context. *Language Learning*, 58(1), 33–71.
- Taguchi, N. (2011). The effect of L2 proficiency and study-abroad experience in pragmatic comprehension. *Language Learning*, 61(3), 904–939. <https://doi.org/10.1111/j.1467-9922.2011.00633.x>
- Taguchi, N., Li, S., & Liu, Y. (2013). Comprehension of conversational implicature in L2 Chinese. *Pragmatics & Cognition*, 21(1), 139–157. <https://doi.org/10.1075/pc.21.1.06tag>
- Taguchi, N., Gomez-Laich, M. P., & Arrufat-Marques, M. J. (2016). Comprehension of indirect meaning in Spanish as a foreign language. *Foreign Language Annals*, 49(4), 677–698. <https://doi.org/10.1111/flan.12230>
- van Compernelle, R. A. (2013). From verbal protocols to cooperative dialogue in the assessment of second language pragmatic competence. *Intercultural Pragmatics*, 10(1), 71–100. <https://doi.org/10.1515/ip-2013-0003>
- van Compernelle, R. A. (2014). Profiling second language sociolinguistic development through dynamically administered strategic interaction scenarios. *Language & Communication*, 37, 86–99. <https://doi.org/10.1016/j.langcom.2014.01.003>
- van Compernelle, R. A., & Kinginger, C. (2013). Promoting metapragmatic development through assessment in the zone of proximal development. *Language Teaching Research*, 17(3), 282–302. <https://doi.org/10.1177/1362168813482917>

van Compernelle, R. A., & Zhang, H. (2014). Dynamic assessment of elicited imitation: A case analysis of an advanced L2 English speaker. *Language Testing*, 31(4), 395–412.

<https://doi.org/10.1177/0265532213520303>

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Zhang, H., & van Compernelle, R. A. (2016). Learning potential and the dynamic assessment of L2 Chinese grammar through elicited imitation. *Language and Sociocultural Theory*, 3(1), 99–119.

<https://doi.org/10.1558/lst.v3i1.27549>

Appendix A. Background Questionnaire

(1) Gender: _____

(2) Age: _____

(3) Dominant language: _____

(4) Other languages: _____

(5) What level of Chinese classes have you already completed at your university?

Elementary Intermediate Advanced Other (specify) _____

(6) What level of Chinese class are you currently taking at your university?

Elementary Intermediate Advanced Other (specify) _____

(7) Number of years of formal Chinese instruction: _____

(8) List the following information for any previous Mandarin studies (e.g., college, high school, intermediate/elementary school, Chinese language school, private language institute, private tutor, etc.). Please also include the current study program.

1) School: _____ (school name) in _____ (country name)

Start year: _____ End year: _____ Hours of Mandarin class per week _____

2) School: _____ (school name) in _____ (country name)

Start year: _____ End year: _____ Hours of Mandarin class per week _____

3) School: _____ (school name) in _____ (country name)

Start year: _____ End year: _____ Hours of Mandarin class per week _____

4) School: _____ (school name) in _____ (country name)

Start year: _____ End year: _____ Hours of Mandarin class per week _____

(9) Have you visited/lived in a Chinese-speaking country?

No Yes (if yes, see below)

At what age _____ Length of stay _____ Location _____

At what age _____ Length of stay _____ Location _____

(10) Do you speak or use **Chinese** (including **Chinese dialects**) at home? _____

(11) Do you identify yourself as Chinese? _____

(12) Check if your parents, grandparents, or anyone else in your immediate/extended family is a native speaker of Mandarin Chinese or a Chinese dialect.

Mother Father Maternal grandparent(s) Paternal grandparents(s)

Other(specify) _____

About the Authors

Tianyu Qin is Assistant Professor of Chinese in the Department of Modern and Classical Languages at the University of North Georgia. Her primary research interests include second language pedagogy, assessment, and interlanguage pragmatics. She approaches her work from the perspective of Vygotsky's sociocultural theory.

E-mail: tianyu.qin@ung.edu

Rémi A. van Compernelle is Associate Professor of Second Language Acquisition and French and Francophone Studies at Carnegie Mellon University. His work extends Vygotskian psychology to L2 education, with particular focus on concept-based instruction, dynamic assessment, and classroom discourse and interaction.

E-mail: vancomp@cmu.edu