

# Interlanguage Stretching within a Task-based Empirical Pedagogy

Tom Means

*Borough of Manhattan Community College  
City University of New York*

Larry Selinker

*New York University  
Research Production Associates*

## ABSTRACT

This paper is a reinterpretation of a study published on learning Italian as an L2 (Means, 2011). We look at how certain key variables in second language learning and teaching might intersect: (1) the interlanguages of learners—when they develop and when they are stuck; (2) task-based language teaching (TBLT) as an empirical pedagogy. We begin with some relevant definitions, describe the empirical study and discuss its implications in light of these variables. The study was a six-week, method-comparison, experimental study conducted with two groups: a control group learned intermediate Italian as a foreign language through traditional instruction (TI), and an experimental group learned through TBLT. Results showed that, for accuracy, the TBLT group marginally outperformed the TI group; for fluency, the TBLT group significantly outperformed the TI group. We then look at some qualitative interlanguage data and conclude that the results presented here suggest that TBLT does effectively stretch learners' interlanguages and may shed some systemic light on viable options for second language pedagogies.

Keywords: task-based language teaching; interlanguage; pedagogy; Italian

## 1. INTRODUCTION

In this paper, we are interested in trying to understand how certain key variables in 2<sup>nd</sup> language learning and 2<sup>nd</sup> language teaching might intersect. What we have worked on below are often presented as different endeavors, not necessarily affecting each other. We are interested in exploring the intersections of two sets of variables:

- the interlanguages of learners, which are notorious for getting stuck, the phenomenon known as “fossilization” (cf. Selinker, 1992; Han, 2004; Han & Odlin, 2006, amongst others)
- task-based language teaching as an empirical pedagogy, where input is heavy and well-defined (Willis, 1996; Means, 2008; Long, 2014)

We are motivated to study this intersection since results to date indicate that, within the domains examined, this empirical TBLT pedagogy (Means, 2008, 2011) does indeed effectively “stretch” learners’ interlanguages. We begin with some relevant definitions, describe an empirical study and discuss its implications.

## 2. DEFINITIONS OF KEY TERMS

In this section we attempt to define as clearly as possible the notions we will need in this work.

Definitions of “interlanguage” and “interlanguage stretching” in this study. We begin at the beginning (Selinker, 1972) where interlanguage is first conceived of as a partially independent linguistic system that is created when a learner attempts to express meaning in a second language in a meaningful performance situation.

We next define our understanding of “interlanguage stretching.” We first came across this concept in a very useful paper by Swain (Swain, 2000) explaining anecdotally a set of empirical data. In this paper, we wish to build on this concept by an initial attempt to create the definitions below and Swain (*ibid*) is not to be held responsible for these.

As a working definition we believe that interlanguage stretching involves the learner’s interlanguage learning conceived of in a particular way:

- the learner’s attention to interlanguage form at Time1 as compared with the learner’s attention to interlanguage form at Time2.
- the learner’s interlanguage (hopefully) restructured at Time2, (hopefully) as a result of clear pedagogical intervention, and (most hopefully) toward the desired target-language norm.

We find it useful to distinguish two types of interlanguage stretching:

- Interlanguage stretching of accuracy: students’ interlanguages come closer and closer to target-like production.
- Interlanguage stretching of fluency: students’ interlanguages are produced with more and more ease

## DEFINITIONS OF EMPIRICAL PEDAGOGY AND TBLT

Following Han & Selinker (1999) we define “empirical pedagogy” as curricula supported by empirical studies that can usefully provide guidance as to what viable pedagogical intervention should consist of. This perspective takes a lot of the guesswork out of pedagogical choice and Han & Selinker (*ibid*) offers empir-

ical evidence as to its efficacy in one context with a limited number of structures. We here add to that empirical base, adding to what others have called “research-based pedagogy” (cf., Norris, 2009).

The version of empirical pedagogy, which we present here, TBLT, is “input-heavy” because, as we show, this version of TBLT combines copious (collaborative) input and opportunities for meaningful output. This TBLT model is based on a three-phase task cycle of pre-task, during-task, and post-task. This approach is an adaptation of J. Willis’ (Willis, 1996) three phases of pre-task, task cycle, and language focus. Emphasis is placed on *input-heavy* because the study dealt with Italian as a foreign language being learned in the United States; accordingly students had little exposure to Italian outside of the classroom (similar to most foreign-language learning environments). Means (2011) attempted to make up for this paucity of naturalistic exposure to Italian by flooding the TBLT group with in-class, task-relevant input. Following is an abbreviated outline of the model of TBLT used in the study:

Pre-task phase:

- tasks are explained (if needed, in both native and target languages)
- task demonstration performed by teacher
- copious task-relevant input is provided, and output activities are performed
- pre-task planning is performed individually and/or in pairs

During-task phase:

- learners, under some form of time pressure, perform the task
- performance is recorded (if oral)

Post-task phase:

- learners prepare transcript, version A, of performance (if oral)
- learners receive feedback on performance
- learners work on relevant focus-on-form exercises
- learners revise their transcripts (if oral) or written performances and hand in version B

Across the three phases, conditions were hopefully created for natural learning to occur: there was challenging exposure to extended input, opportunities for spontaneous output, and there was a focus on more formal, planned language. As TBLT has a rich research base (Long, 2014), we are confident referring to it as an empirical pedagogy.

### 3. THE STUDY

The study was a six-week, method-comparison, experimental study conducted with two groups: a control group learned intermediate Italian through traditional instruction (TI), and an experimental group learned intermediate Italian through task-based language teaching (TBLT). Participants were randomly assigned to either the TI or the TBLT group.

#### 3.1. Research Design

Participants (n = 22) were divided into two groups: the first received TI of intermediate Italian (n = 10); the second received TBLT of intermediate Italian (n = 12). A battery of four tests (two controlled-production measures, and two spontaneous-production measures) was used in a pre-, post-, and delayed posttest design (testing for both immediate and sustained effect). The same teacher taught both groups.

##### 3.1.2. Materials

Three task cycles were used for the TBLT treatment in the study (approximately one cycle per two weeks). The first task performance was “present your partner’s daily routine;” the second task was “present how your partner spent his/her last three birthdays;” the third task was a spot-the-difference task. All students were assigned a different partner for each task cycle.

Three textbook chapters were used for the TI treatment in the study. All three were taken from the intermediate-level text, *Ponti* (Tognozzi, 2004). The chapters treated (2, 4, 6) had the following respective grammar and cultural themes: Chapter 2: Present perfect/journalism in Italy; Chapter 4: Direct object pronouns/Italian contemporary music; Chapter 6: Adjective agreement/Italian holidays and superstitions.

It was predicted that both types of instruction would have beneficial effects on learner performance, but that the TBLT group might display more accuracy and/or more fluency after treatment. The research design was a true experimental group/control group design. The TBLT model was based on the identification of variables identified by SLA research to produce optimal learning effects, and categories described by Ellis (2003), Lee and VanPatten (2004) and Willis (1996). The TI approach was based on observations of Italian classes at a large public U.S. university and categories described by Sheen (2003). Both groups’ treatments

lasted for six weeks, with two 90-minute lessons per week; both groups met for a total of nine weeks, including the data collection of a 21-day delayed posttest.

The independent variable of this study is type of instruction, and it has two forms: TBLT and TI. There are two dependent variables: 1) learners scores on grammatical gender (GG) agreement accuracy in the noun phrase (NP) and 2) oral and written fluency counts, with all scores collected across three times: pre-test, posttest, and delayed posttest. One student level is included, intermediate Italian as a foreign language.

### 3.2. Central Hypothesis

The central hypothesis for this study was: TBLT will better promote interlanguage stretching in terms of accuracy and fluency as defined above than will TI for intermediate Italian as a foreign language.

This was assumed to be true because of:

- a. the nature of the copious input of this type of TBLT, and
- b. its primary focus on research-based priority of promoting meaning-making before form development.

We come to this assumption based on our experiences that in the pre-task stage, especially, we believe we are giving students a real reason to focus on the input which they are going to need for their upcoming performances, performances which have content communication as their main goal (noted in point b above). Previous research (Benati, 2004; Van Patten, 2003; Long, 2014) and our experiences tell us that there is an input advantage which can result in long-term interlanguage restructuring seen as interlanguage stretching in the sense described in this paper.

Additionally, the nature of this model of TBLT is such that an inherent pressure is built into the pedagogy. Research has shown that time pressure can enhance learners' output (Hulstijn & Hulstijn, 1984). The data in this paper hint that the TBLT students' burgeoning Italian may have fed off this pressure.

### 3.3. TI Treatment

Having already described the TBLT treatment above, here we will detail TI. TI, also commonly cited as PPP (Presentation, Practice, Production), is generally presented as a three-part cycle in the literature (Willis & Willis, 1996; Sheen, 2003, 2005; Swan, 2005).

The basic order of TI is to: 1) *present* a discrete grammatical or cultural item; 2) *practice* that discrete grammatical point, or discuss the cultural item and then 3) *produce* that grammatically- or culturally-relevant language in a relatively free context, with learners interacting in pairs or groups, often to complete a written exercise and/or an oral communication task.

To establish an empirical base for the TI group’s treatment, two sections of intermediate Italian were observed across a semester at the same university where the study took place, recording the activities and interactions of the classes. Means (2011) was able to replicate the way intermediate Italian is taught at this large public university, and this replication did, indeed, mirror the ‘three Ps’ frequently found in the TI literature (Sheen 1994, 2003, 2005; Swan 2005). The TI group’s treatment was largely based on these recorded observations. The TI treatment was intended to accurately reflect the way Italian is currently taught in a university setting in the United States. Based on the collected data, an empirically-based treatment for the TI group was created that consisted of four phases: Cultural readings; Vocabulary exercises; Grammar point presentation; Individual and group work.

### 3.4. Battery

The battery of four tests (Appendix) developed for this project consists of 4 production tests (two controlled-production tests, two spontaneous-production tests). The controlled-production tests were designed to be partial to the TI group’s treatment, as most of their instructional time was spent with similarly controlled activities. The spontaneous-production tests were designed to be partial to the TBLT group’s treatment, as most of their instructional time was spent decoding and/or producing relatively spontaneous language. A split-block design was followed across the three testing times, using three balanced versions of the test.

### 3.5 Data Collection Procedure

The battery of four tests was developed and administered to both groups to measure the effects of the two treatments. There were three versions of the battery (version A, B, and C). Version A was used for the pretest for both groups, version B was used for the posttest for both groups after the 6-week, 3-hour-per-

week treatment, and version C was used for the delayed posttest, 21 days after the posttest.

### 3.6. Scoring Conventions

#### 3.6.1. Scoring for GG accuracy:

For GG coding, following Doughty (2003), the scoring is structured in an interlanguage-sensitive fashion, by rewarding partially target-like tokens with positive points. The GG agreement feature (Franceschina, 2001, 2005; White et al., 2004) is analyzed across all instances of article-noun agreement and noun-adjective agreement collected in the three data collection sessions. The scoring for GG accuracy was based on a system that ranged from 3 points, for target-like use of GG, to 0 points for target-deviant use of GG. Along the continuum were intermediate points of GG use that were awarded 1 point.

#### 3.6.2. Scoring for fluency:

Fluency counts were only possible for 3 of the 4 tests: Written fluency was counted on Test 2, Narrative Writing Summary; oral fluency was counted on Test 3, Narrative Retell and Test 4, Family Tree. (Test 1, Create Complete Sentences, was not amenable to a fluency count.) Oral Fluency in Tests 3 and 4 (Narrative Retelling and Family Tree, respectively) were scored as spoken syllables per minute (Ellis and Yuan, 2004); both tests were time-pressured oral tests lasting one minute. Fluency scores on Test 2, Narrative Writing Summary, were scored as total words per summary (Chenoweth and Hayes, 2001). Participants were given 6 minutes to write the summary of a film clip across all three data collections.

### 3.7. Results

#### 3.7.1 Accuracy results

Results showed that for the accuracy measure, GG agreement in the noun phrase, effects were beneficial for both treatments but a marginal advantage was noted for the TBLT group. Table 1 displays the descriptive data for the GG agreement accuracy in the spontaneous-production test, Test 4, Family Tree. The maximum possible score obtainable was 100. Here we see that although the TBLT group did improve more than the TI group did, it was not at a significant level; therefore we cannot say that the advantage displayed by the TBLT group was not due to chance. Statistical significance was tested using a t-test: Paired

Two Sample for Means, to test for statistically significant improvement for overall GG agreement in both groups, across all three times.

Table 1  
Descriptive Data for Accuracy Findings in Test 4, Family Tree

Group	Pretest		Posttest		Delayed Posttest	
	Mean	SD	Mean	SD	Mean	SD
TBLT	65.4	26.8	70.4	24.4	73.4	17.2
TI	75.0	13.4	69.7	20.3	67.4	31.7

### 3.7.2. Fluency results

For fluency, on all three tests that were amenable to such a measure, the TBLT group demonstrated significantly greater gains at posttest. At delayed posttest this significant gain was maintained on two of the three tests amenable to such a measure (at delayed posttest the TI group matched the TBLT group in significant gains on the written fluency test, Narrative Writing Summary).

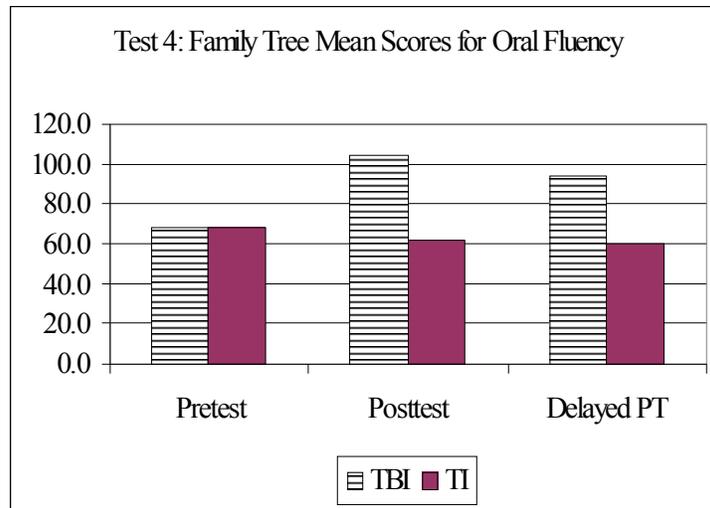
Table 2 displays the descriptive data for the oral fluency measures in the spontaneous-production test, Test 4, Family Tree. The numbers reflect spoken syllables per minute.

Table 2  
Descriptive Data for Oral Fluency Findings in Test 4, Family Tree

Group	Pretest		Posttest		Delayed Posttest	
	Mean	SD	Mean	SD	Mean	SD
TBLT	68.4	16.9	104.8	26.3	93.9	24.9
TI	68.1	14.8	61.7	17.6	59.8	24.9

These results show a significant increase of the mean scores for the TBLT group between the time of the pretest and posttest, and between the time of the pretest and the delayed posttest (significance was tested using a t-test: Paired Two Sample for Means). Figure 1 presents these findings in a bar chart.

Figure 1  
Family Tree Test, Mean Scores for Oral Fluency



#### 4. DISCUSSION AND CONCLUSIONS

This final section will integrate the above quantitative work with some of the interlanguage data gathered for the study, presented qualitatively. Our overall purpose here is an attempt to make some progress integrating the two major sets of variables defined above, which we repeat here:

- 1) the interlanguages of learners—when they develop and when they are stuck;
- 2) TBLT as an empirical pedagogy that helps guide the teacher into making pedagogical decisions relevant to particular learners.

First, we ask, was the central hypothesis retained as a result of the empirical work above? We repeat the hypothesis here for easy reference: TBLT will better promote interlanguage stretching in terms of accuracy and fluency as defined above than TI for intermediate Italian as a foreign language.

For interlanguage oral fluency, yes, the central hypothesis was strongly retained; for written fluency it was partially retained.

For interlanguage accuracy, the central hypothesis was marginally retained. We next explore these results.

4.1. First we will draw inferences from the data as to why fluency better improved through the TBLT treatment. To arrive at these inferences, we will explore qualitatively a small subset of the interlanguage data, viz. how one TI sub-

ject and one TBLT subject responded in the same fluency test at pretest and at delayed posttest experiences, leading us to infer what, precisely, the TBLT treatment did that the TI treatment did not do, in order to explain these improvements. The examples are from Test 3, a narrative retelling of a video clip.

Example 1a. Data from TI participant, narrative retelling task, pretest

TI subject, Example 1a:

*Pretest, Test 3, Narr Retell, TI-group subject # 5080 (80 syllables in 1 minute)*

RECORDED INTERLANGUAGE DATA:

Nel video il ragazzo ha visto un amico ...um... nel background italiano musica e l'uomo non è contento con il ragazzo ...um ...poi il ragazzo \*ha partito la sorella ha visto il ragazzo e la sorella ha detto, « Michele scendi « ...um.. Michele non \*d'accordo

[The \* indicates an ungrammatical Italian construction in the grammatical domain under consideration.]

Example 1b. Data from same TI participant as above, narrative retelling task, delayed posttest (i.e., 9 weeks later)

TI subject, Example 1b:

*DelayedPosttest, Test 3, Narr Retell, TI-group subject # 5080 (82 syllables in 1 minute)*

RECORDED INTERLANGUAGE DATA:

Nel film il ragazzo \*ha arrivato con \*il bicicletto nella piazza, \*un altre ragazze hanno visto le scarpe nell'automobile um...il padre di Michele non e' contento e Michele \*correndo, nell'albero Michele \*realizante c' è la sorella e finalmente Michele risponde con 'che c' è \*la' mangiare'?

As we can see, comparing examples (1a) and (1b), the TI subject's fluency remained constant after nine weeks of treatment and delay. That is, from these data, we cannot trace any improvement in fluency from the TI treatment. This may be due to TI's continuation of teacher frontedness: one of our assumptions is that the teacher-centered, form-centered TI treatment promotes excessive monitoring and that this manifests itself as an obstacle to fluency. Our essential observation on this front is that TI subjects did not have the same level of (good) pressure (discussed above in Section 3.2) that TBLT subjects did throughout their

treatment, and this explains their lack of improvement on fluency scores. Note that, importantly, the qualitative interlanguage data also show continued ungrammaticality in comparing both samples.

Next, we look at what an evenly-matched TBLT subject did in a comparative time frame on the same test. We refer to them as evenly-matched because fluency scores were similar at pretest.

Example 2a. Data from TBLT participant, narrative retelling task, pretest

TBLT subject, Example 2a:

*Pretest, Test 3, Narrative Retelling, TBLT-group subject # 3262 (58 syllables in 1 minute)*

RECORDED INTERLANGUAGE DATA:

Con la sua bicicletta, il ragazzo va al centro per un po' e...la sua mamma  
\*pregunta quanto \*coste le scarpe e...il ragazzo incontra \*il suo padre e \*ha ...  
incontra \*la sua fratella.

Example 2b.

Data from same TBLT participant as above, narrative retelling task, delayed post-test, i.e., 9 weeks later

TBLT subject, Example 2b:

*Delayed Posttest, Test 3, Narr. Retell, TBLT-group subject # 3262 (107 syllables in 1 min.)*

RECORDED INTERLANGUAGE DATA:

Un ragazzo si chiama Michele cammina al centro di una campagna con una  
bicicletta ... \*suo bato ... \*hai \*molte gente e una donna che prende una scarpa  
di una machina di un uomo che vende le scarpe. Parla con le sue amiche .... della  
scarpa ... il ragazzo incontra \*il suo padre e c' è un problema ... cammina  
\*scende un albero

As we can see, comparing Examples (2a) and (2b), the TBLT subject's fluency nearly doubled. One of our assumptions is that the type of real-world, time-pressured communication utilized in the TBLT treatment appears to pay large dividends in fluency improvements. The type of training involved in a TBLT course, especially if the tasks are oral—as all three tasks were in the TBLT group's 6-week

treatment—should promote fluency development. According to these data, for this small population, such fluency progress may have been the direct result of TBLT.

We also believe that the heightened and salient presence of copious collaborative input in the TBLT treatment proves to be a methodological advantage, in that this type of interpretation exercise, in which students are asked to listen to long passages of task-relevant target language, repeatedly, and to identify, transform or confirm content in the passage, better captures the TBLT learners’ attention. Such means of attempting to force attention appears to be more effective than traditional exposure to input that TI groups receive. The TI group’s input was primarily provided by the textbook (for written input) and by the teacher (for audio input) and was complemented by Italian songs and brief movie clips.

In the TBLT treatment, on the other hand, input that appears to be processed immediately before the structured output inherent in TBLT performances, seems to have altered the TBLT learners’ access to memory-based language system and their ability to articulate such language, thus stretching the interlanguage in the sense we intend.

Furthermore, it may also be the case that such a TBLT methodology better brings out the working memory potential in learners, also because of the training it provides with practicing spontaneous speech. In this view, memory usage is concerned primarily with retrieval and with the way language elements appear to be stored in the brain (cf. Temple, 2000). Following Skehan’s cognitive model (1998)—one theoretical framework for this study—these elements are probably stored redundantly and formulaically, the emphasis here being on how one’s memory can efficiently retrieve items in real time to handle the demands of spontaneous speech production. Thus, the data collected, we argue, illustrate the different fluency-building effects that the respective treatments had.

4.2. Next, we look at why the TBLT treatment enjoyed greater success (though at a marginal level) at developing accuracy. Again, we look at examples from the qualitative interlanguage data, comparing how two evenly-matched participants performed across all three data collections. The two participants are taken from the respective treatments as described above. They are referred to as evenly-matched because they had similar accuracy scores on GG agreement at pretest.

Table 3

Recorded Interlanguage Data TI-group, subject #5080 Accuracy score at Pretest: 2.4

Test 1, Controlled-production, Create Complete Sentences	Pretest	Posttest	Delayed PT
Token 1	*mese	*mese	*le mese
Token 2	la forma	la forma	*forma

Table 4

Recorded Interlanguage Data, TBLT-group, subject #3262, accuracy score at pretest: 2.6

Test 1, Controlled-production, Create Complete Sentences	Pretest	Posttest	Delayed PT
Token 1	*il spettacolo	*il spettacolo	lo spettacolo
Token 2	*il articolo	*il articolo	un articolo

Comparing Tables 3 and 4 we can see that the TI subject, #5080, did not reach the target-like GG agreement form, *il mese*, for the first token in question. At both pretest and posttest she omitted the definite article completely. Then she seems to overgeneralize a masculine singular article in the delayed PT phase, *\*le mese*, one that appears to have surfaced as a type of interlanguage transfer (Selinker & Baumgartner-Cohen 1995; Selinker & DeAngelis 2001; Cenoz, Hufeisen & Jessner, 2001, *passim*), i.e. the use of interlanguage knowledge gained from her previous study of French, this fact was revealed in her biographical data.

Interestingly, with the following token, *la forma*, another interlanguage phenomenon might be revealed here as well, backsliding (Selinker, 1972, 1992, and Cook, 2003, *passim*), At pretest and at posttest she created the correct token, but

at delayed posttest, she omitted the definite article completely and produced just *\*forma*. She had been following a similar omission strategy with the previous token *\*mese*. If our analysis is correct, and this can only be settled by in-depth qualitative interlanguage study, by delayed posttest she had backslid to the omission strategy she had employed at pretest and posttest for the first token, *\*mese*.

In the context of this small study, we take this as evidence that TI can cause students to over-edit and to over-monitor their production, possibly because of the overriding focus on forms which is an inherent feature of TI (cf. the overriding focus on meaning-making in TBLT).

As a counterpoint, we turn to Table 4 to observe how the TBLT subject, #3262, gradually moved toward a target-language norm, improving her agreement accuracy and by delayed posttest she displays a target-like token with both examples tracked above: *lo spettacolo* and *un articolo*. She seemed to overcome her tendency to overgeneralize with the masculine, singular ‘default’ article of *il* in achieving this target-like agreement at delayed posttest.

TBLT seems to have restructured her interlanguage here, probably altering her production strategy, possibly through the collaborative input mentioned above, such that she was providing target-like agreement by the delayed posttest. So, our faith in this type of TBLT has a modest qualitative foundation as well.

### 4.3

In this study, by using both quantitative measurements on learners’ interlanguages, integrated with close attention to qualitative interlanguage data, we have seen learning occur in both treatment groups. This learning, conceived of as interlanguage stretching, appears to have been more significant in the TBLT treatment, within an empirical version of such. Thus, this study identified fruitful interaction between second language learning—a particular type of learning we are calling interlanguage stretching—and second language teaching conceived of in a particular way. We have here compared (input-heavy) TBLT, which our practical experiences have led us to favor, with classical TI with its attention on forms—and which, as noted in our empirical base for the TI treatment, is still the way Italian is most commonly taught as a foreign language in the United States.

The fruitful interaction we have been seeking appears to have been generated most effectively within the TBLT treatment. Therefore, we are cautiously optimistic that this study adds evidence to TBLT’s viability as an effective empirical pedagogy. We have seen the interlanguages of the TBLT participants stretch convincingly from Time1 to Time2. Through the healthy balance of copious input and meaningful output espoused in TBLT, this approach seems to have led to a desirable restructuring of the participants’ interlanguages.

The usual limitations of generalizability apply (teacher, learner, school, and educational environment factors, at the very least, will vary greatly from situation to situation) but this study shows in its limited context that TBLT does indeed improve learners fluency (and accuracy, albeit marginally) better than TI.

#### REFERENCES

- Benati, A. 2004. ‘The effects of processing instruction and its components on the acquisition of gender agreement in Italian.’ *Language Awareness* 13, 2.
- Cenoz, J., B. Hufeisen & U. Jessner (Eds.). 2001. *Cross-linguistic influence in third language acquisition: psycholinguistic perspectives*. Clevedon: Multilingual Matters.
- Chenoweth, N. Ann and J. Hayes. 2001. ‘Fluency in writing.’ *Written Communication* 18.1.
- Cook, V. (Ed.). 2003. *Effects of the Second Language on the First*. Clevedon: Multilingual Matters.
- Doughty, C. 2003. ‘Instructed SLA.’ In C. Doughty and M. H. Long (eds.), *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell.
- Ellis, Rod. 2003. *Task-Based Language Learning and Teaching*. Oxford: OUP
- Ellis, R. and F.Yuan. 2004. ‘The effects of planning on fluency, complexity and accuracy in second language narrative writing.’ *Studies in Second Language Acquisition* 26: 59-84.
- Franceschina, F. 2001. ‘Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals.’ *Second Language Research* 17.3.
- Franceschina, F. 2005. *Fossilized Second Language Grammars: The acquisition of grammatical gender*. Philadelphia: John Benjamins
- Han, ZH. & Odlin, T. 2006. Introduction. In Z. Han, T. Odlin (eds.) *Studies of fossilization in second language acquisition*. Clevedon, U.K.: Multilingual Matters, 2006, pp. 1-20.
- Han, ZH. & Selinker, L. 1999. ‘Error Resistance: Towards an Empirical Pedagogy.’ *Language Teaching Research* 3.3: 248-275.
- Han, ZH. 2004. *Fossilization in Adult Second Language Acquisition*. Clevedon: Multilingual Matters.
- Hulstijn, J. and W. Hulstijn. 1984. ‘Grammatical errors as a function of processing constraints and explicit knowledge.’ *Language Learning* 34: 23-43.

- Lee, J and B. VanPatten. 2004. *Making Communicative Language Teaching Happen*. New York: McGraw Hill.
- Long, M.H. 2014. *Second Language Acquisition and Task-Based Language Teaching*. Malden, MA: Wiley-Blackwell.
- Means, T. 2008. ‘Task-based instruction of intermediate Italian: A research-supported model.’ In Occhipinti, E. (ed.) *New Approaches to Teaching Italian Language and Culture: Case Studies from an International Perspective*. Cambridge: Cambridge Scholars Publishing.
- Means, T. 2011. *A Comparative Study of the Effectiveness of Task-based versus Traditional Instruction of Intermediate Italian: Findings on Accuracy and Fluency*. Lewiston, New York: Mellen Press.
- Norris, J. 2009. ‘Evaluating task-based language programs.’ Paper delivered at the 3<sup>rd</sup> Biennial Task Based Language Teaching Conference, Lancaster, UK.
- Selinker, L. 1972. Interlanguage. *International Review of Applied Linguistics*. 10: 209-231.
- Selinker, L. 1992. *Rediscovering Interlanguage*. London: Longman.
- Selinker, L. & B. Baumgartner-Cohen. 1995. Multiple Language Acquisition: ‘Damn it, why can’t I keep these two languages apart?’ *Language Culture and Curriculum* (special issue on multiple language acquisition), M. Ben-Soussan & I. Berman (eds.). 8.2, 1-7.
- Selinker, L. & G. De Angelis. 2001. Interlanguage transfer and competing linguistic systems in the multilingual mind. In: J. Cenoz, B. Hufeisen & U. Jessner (Eds.) *Cross-linguistic influence in third language acquisition: psycholinguistic perspectives*. Clevedon: Multilingual Matters
- Sheen, R. 1994. ‘A critical analysis of the advocacy of the task-based syllabus.’ *TESOL Quarterly*, 28, 1.
- Sheen, R. 2003. ‘Focus on form—a myth in the making.’ *ELT Journal* 57: 225-233.
- Sheen, R. 2005. ‘Focus on formS as a means of improving accurate oral production.’ in P. Jordens (ed.), *Investigations in Instructed Second Language Acquisition*. New York: Mouton de Gruyter.
- Skehan, P. 1998. *A Cognitive Approach to Language Learning*. Oxford: OUP
- Swain, M. 2000. ‘The output hypothesis and beyond: Mediating acquisition through collaborative dialogue.’ In J. Lantolf (ed.) *Sociocultural Theory and Second Language Learning*. Oxford: OUP. 97-114.
- Swain, M. 2005. ‘The Output Hypothesis: Theory and Research.’ in E. Hinkel (ed.), *Handbook of Research in Second Language Teaching and Learning*. Mahway, New Jersey.
- Swan, M. 2005. ‘Legislation by Hypothesis: The case of task-based instruction.’ *Applied Linguistics* 26.3: 376-401.
- Temple, L. 2000. ‘Second Language Learner Speech Production.’ *Studia Linguistica*, 54.2. 288-297.

Means & Selinker • “Interlanguage Stretching within a Task-based Empirical Pedagogy”

VanPatten, B. 2003. *From Input to Output: A teacher's guide to SLA*. New York: McGraw Hill.

White, L., E. Valenzuela, M. Kozłowska-MacGregor, and Y. Leung. 2004. 'Gender and number agreement in nonnative Spanish.' *Applied Psycholinguistics* 25: 105-33.

Willis, J. 1996. *A Framework for Task-Based Learning*. Harlow: Longman.

Willis, J. and D. Willis. (eds.). 1996. *Challenge and Change in Language Teaching*. Oxford: Heinemann Macmillan.

## APPENDIX

Battery of tests. Tests 1 & 2 are controlled production measures; Tests 3 & 4 are spontaneous production measures.

### TEST 1—CREATE COMPLETE SENTENCES

[For brevity’s sake, following are only three examples— all related to the tokens presented above—the actual test contained 12 prompts.]

Create complete sentences in the present perfect and use appropriate articles and prepositions (if necessary):

1. Noi/perdere/spettacolo.
2. Voi/partire/mese scorso.
3. Noi/decidere/forma.

### TEST 2—NARRATIVE WRITING SUMMARY

You will watch a 3-minute video. The video will be shown only once. After watching the video you will have 6 minutes to write a complete summary of the scene.

Use the back of this page to write the summary. If you have any questions, please ask them now.

### TEST 3—NARRATIVE RETELLING

You will watch a 3-minute video. The video will be shown only once. After watching the video you will have one minute to plan an oral summary of the scene. After the one minute of (silent) planning, you will each be recorded for 1 minute as you describe, in Italian, a summary of the scene.

You cannot write any notes. If you have any questions, please ask them now.

### TEST 4—FAMILY TREE

You will be paired with one of your classmates and you will have a total of 3 minutes to tell each other about your families (siblings, parents, aunts, uncles, cousins, etc.). After the 3 minutes are up, you will each be recorded for 1 minute as you describe, in Italian, your partner’s family tree.

You cannot write any notes. If you have any questions, please ask them now.