

Effectiveness of Deep, Blended Language Learning As Measured By Oral Proficiency and Course Evaluation

Francois Victor Tochon
University of Wisconsin-Madison

Abstract

This article explores the effectiveness of a newly-launched, technology-rich approach in the classes of a less commonly taught language, Turkish; it is one of the rare studies on the proficiency level of students of Turkish as a foreign language, and it provides valuable feedback on new ways of teaching and learning a second language that might help the development of future strategies in the field. Both proficiency and course evaluation significantly improved with the new, deeper approach compared to control groups. The study is a unique contribution and a real opening to approaches in which students are placed as curriculum builders and the language instructor plays a role of facilitator in the management of rich online resources, blending face-to-face interactions and learning support with online activities.

Introduction

This article describes the effects of the integration of online instructional material that takes a “Deep Approach” to Turkish language acquisition. This immersive, learner-centered, technology-rich, and project-based approach was designed for institutions of higher education that offer programs in Turkish language and culture, and Middle East studies (Tochon, Ökten, Karaman, & Druc, 2012). It creates opportunities for blended learning, defined as an education that combines classroom interactions with computer-mediated activities. One advantage is the connection with differentiated instruction: blended learning can be based on personal learning environments that are self-regulated (Tochon, Ökten, Karaman, & Druc, in press). Among the technologies used for blended learning in this research are a new hypertextbook, streaming videos, multimedia, and PowerPoints (Tochon, 2013), and the integration of current technologies into instructional modules, such as blogs, Skype, Livemocha, forums to create educative projects, etc. Larmer, Ross, and Mergendoller

(2009) define project-based learning as a teaching method in which the learners are in large part independent from the teacher—they demonstrate in-depth understanding and critical thinking, and engage in an inquiry process on authentic and complex issues that lead to quality public productions. In addition to these aspects, one key principle of the Deep Approach is that students choose their own projects (rather than having them imposed by the teacher) and create their own curriculum for the projects with their peers and guided by the teacher. This article explores the difference this integrated learning and Deep Approach makes in terms of proficiency growth, as demonstrated through the American Council on the Teaching of Foreign Languages' Oral Proficiency Interview, or OPI, and course evaluations.

Self-determined Projects, Deeper Learning, and Enhanced Motivation

Blended learning is particularly helpful to scaffold projects, as it supports autonomous problem solving (Delialioğlu, 2012). It places students in self-regulated environments in which they are able to evaluate their own progress (Lee & Lim, 2012). The ever-changing nature of communication in educational spaces and the existence of various literacies with social, cultural, and regional variation require new ways of conceptualizing language instruction. Computers and Internet connections can help access multiliteracies, yet they often lead to shallow learning. Carr (2011) suggests that a counterweight to fast and shallow learning is needed: educators need to study in what ways technologies can enhance rather than prevent deep learning.

Deep language learning is the target of the approach studied in this research project. It involves both cultural understanding and proficiency. A specific teaching methodology is required for deep learning. In Figure 1, Entwistle (2008, p. 7) indicates how knowledge can and should evolve away from authoritative knowledge and dualism toward reasoned commitment, intercultural respect and recognition, which implies parallel conceptions of learning evolving from acquisition to epistemic transformation. Understanding education as an identity-seeking project may transform our vision of what curricula should be and of the need for personalized and self-motivated pathways towards deep understanding. This is the intent of the Deep

Approach, which is relevant and applicable to both education and language learning.

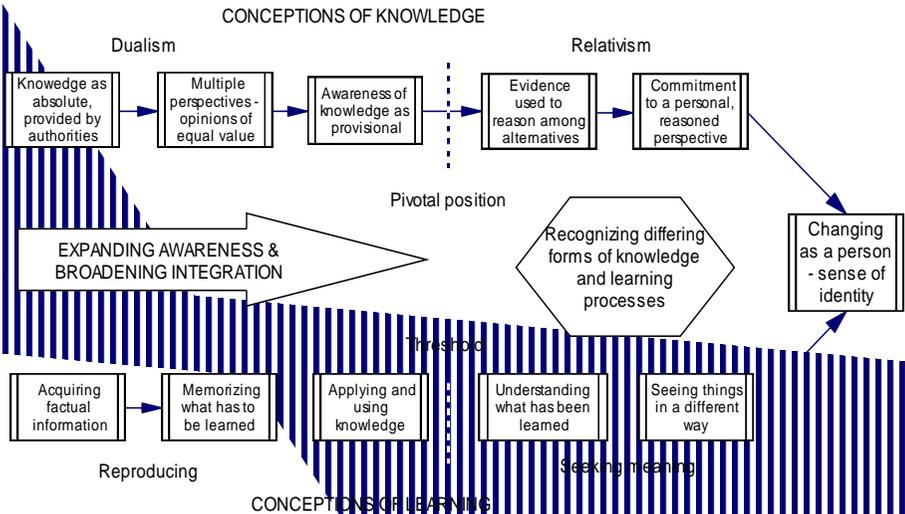


Figure 1. Why we need evolved conceptions of knowledge and learning (derived from Entwistle 2008, p. 7, with authorization of the author)

Deep learning is defined by behaviors related to remembering what is learned for a long time, applying knowledge to new situations, inferring new meanings and generating new ideas, associating concepts to daily experiences, establishing relationships between incidents and results, and examining principles of thought in discussions, while surface learning is defined as quickly forgetting what is learned, not holding discussions using correct principles of thought, understanding a limited amount of knowledge, memorizing only the required information to pass tests, and seeing learning as an extrinsic load. (Göçmençelebi, Özkan, & Bayram, 2012, p. 554)

While the conceptualization and operationalization of deep and surface processing may differ across studies depending on the theoretical frame utilized, the results from studies that meet stringent criteria for deep learning demonstrate that the levels of processing and performance are related, and confirm the assumption that deep processing promotes stronger learning outcomes, while surface processing promotes weaker learning outcomes (Dinsmore & Alexander,

2012).

Deep learning concerns the whole person, implying a sense of purpose, empowerment and transformation (Tochon, 2010a). Environment providing increased learner autonomy may be an excellent ground for a deepening of knowledge. Such are blended learning environments, which integrate e-learning and virtual education with face-to-face interaction (Thorne, 2003). Among the many conceptions of learning, blended environments can be used for project-based learning and peer work, which is conducive to deep learning (Mahoney & Schamber, 2011). They create a sense of community. Jaffee (2007) summarized the advantages of “interdisciplinary connections,” promoting “deeper understanding of content,” and facilitating active learning through application (p. 65). Deep learning emphasizes action, quality, relevance and purposefulness rather than rote learning. Learning a new language is understood as a process of intercultural accommodation, which connects to a variety of subtle meanings and situational elements that need to be related to catch the whole.

Blended with the use of new technologies, project-based learning can stimulate autonomous discovery of new content and direct contact with native speakers (Tochon, 2010b). While conceiving of computer-assisted environments may be extremely complex if the goal is to create virtual learning worlds, such as making a Second Life environment useful for learners (Wold, 2011), the task is easier when it comes to scaffold thematic resources in an environment that permits students to create personal and team projects. Studies on learning effectiveness indicate that students perform significantly better in blended learning environments than in environments that are exclusively online (Chen, 2012). Asynchronous environments are most frequently used for convenience: learners explore topics through emails, forums, blogs, and make their own team schedule without interaction with the instructor (Kruse, 2004). However, synchronous interaction is crucial for second language acquisition (Lee, 2002). Synchronous learning usually integrates the familiar classroom model, learners can receive feedback from their peers as well as the instructor, and can retrieve or create content quickly that otherwise would not be available in the normal classroom (Keegan et al., 2005), which helps students access complex subject matters (Pfister, 2005). Addi-

tionally, Ge (2012) showed that a blended approach in which language learners can interact with contents through streaming video and can contact native speakers through networks such as Livemocha brings significantly better learning results than the asynchronous approach.

Blended environments can offer authentic, collaborative challenges over which learners have control and stimulate meaningful second language use. Real-world themes, issues, problems and actions create reflective situations to solve problems in their context, in a way that respects the autonomy of the student and is conducive to proficiency. Students then have choice, decision-making authority, and voice. They become engaged learners. The next section examines how a deeper approach to language learning could reach these goals. Then we will examine how in-depth projects can stimulate a form of apprenticeship. Such projects can lead to creation, action, and experience. Enjoying the journey is part of the goal.

How the Deep Approach Differs from Previous Methodological Trends

What is new in the approach studied here is the emphasis on the transformational dynamics of life learning—in which learners become social activists through the practice of values and practical wisdom pertaining to the studied cultures (Ikeda, 2010; Obelleiro, 2012). Thematic organizers are used to scaffold projects. They are flexible ways of guiding students' choices, more like personalized standards, without reifying the concepts enacted in the projects. Tochon (2013) summarizes what makes this approach distinct:

- The students are placed in charge of building their own curriculum and projects to achieve their desired expertise, using accountability measures through instructional agreements.
- The basis of the students' curriculum building is the teacher's provision of blended resources (literary, visual, aural, and online) organized adaptively. The teacher becomes expert in scaffolding and facilitating feedback.
- Knowledge is not a "thing" that can be taught as an object: it is understood as deep, subjective and intersubjective, inseparable from the identity process. Depth is defined in a way

that transcends the commodification and commoditization of knowledge. Educative projects are open and become ways of preventing knowledge crystallization and sedimentation. Rather, it is about situated knowledge in action.

- The focus is on deep, personalized processing, not the same standardized outcomes for all. There is room for diversity and flexibility, non-native speaker comfort, code switching, and unique perspectives.
- The approach targets transdisciplinary values for a more sensible and wiser world—this way language learning becomes the means toward conflict resolution, ending war and poverty, re-greening the planet, and turning to more humane politics through socially and even globally-situated projects and problem exploration. Yet, rather than a dualistic view, the principle of the included middle (or third space) is applied, through which two apparently opposed elements can be integrated at a higher (still relative) level.

The Deep Approach is a convergence of what worked best in earlier approaches. Its holistic scope allows for more student autonomy and works for the planet and society while working on the language. The instructional principles for teachers are to:

- go by the results of motivation research, and provide incentives for self-directed learning and self-determination;
- merely scaffold possibilities, thus making the landscape as flexible as possible for students to choose, select, and frame—on their own—the curriculum; in this they use their own literacy-based thematic units, indexing all language modalities to each other, and use online modules (rather than a textbook) or supplement with a large variety of multimedia resources for blended learning;
- emphasize process rather than outcomes; refer to “instructional organizers in forward planning” (Tochon, 2010c) rather than goals or outcomes in backward planning;
- encourage individualized, peer-oriented, and small group project-based learning, focusing on cultural content and social action;

- give primacy to text; consider grammar as storytelling about language; target extensive reading/viewing and intensive writing/recording;
- use deep formative feedback and empowerment evaluation; integrate self-evaluations and peer-evaluations; and,
- focus on value creation by highlighting critical issues related to the respect of other languages and cultures, language status and invisible or open discrimination, the colonial mindset versus principles of social justice, and linguistic human rights for peace building.

In summary, understanding the Deep Approach is seeing education and schooling in an “avant-garde” way, as learning languages and cultures constitutes a threshold towards global and transdisciplinary goals supported by intrinsic motivation and identity investment.

Projects Stimulate Apprenticeship and Learning by Doing

Three motivation theories legitimate project-based approaches: self-efficacy theory, attribution theory, and self-determination theory (Beckett & Miller, 2006; Larmer, Ross, & Mergendoller, 2009). This literature indicates that in-depth projects benefit from the students’ intrinsic motivational impulse, which increases classroom dynamics. Students like to know that they are in control of the factors of their success. Their source of accomplishment is inside, not outside. They develop a sense of ownership over their projects. Self-determination, effort-enhancing attributions, as well as a sense of self-efficacy and what Norton (2000) defines as identity investment, form the groundwork of the Deep Approach to language learning. Blended learning in self-determined projects meets students’ needs for instructional relationships, autonomy, and competence (Deci et al. 1991).

As mentioned by Beckett and Miller (2006), project methodology was initiated by David Snedden in science education and then developed and disseminated by a student of Dewey named William Heard Kilpatrick (1918). Project-based apprenticeship enhances the quality of student learning compared to other approaches; it affects positively problem-solving and decision-making capacities (Thomas, 2000). Projects tend to reduce learners’ anxiety and emulate positive attitudes toward the discipline (Boaler, 2002). When educative pro-

jects target interpersonal and social situations in the other language, situated modeling, scaffolding, collaboration, and coaching stimulate various forms of socialization that enhance knowledge, skills, and experiences within contexts genuinely and informally created by the learners through their collaborations (Brown, Collins, & Duguid, 1989; Ding, 2008). Project-based learning promotes expression, interpersonal exchange, individual thought, and personal apprenticeship.

However, mainstream project-based learning often implies that the instructor is the curriculum builder, which prevents students from choosing their own projects and determining on their own how they will accomplish them. When they are given such self-directed opportunity, deep learners engage in problem solving from the phases of conception, design, decision making, investigation, realization, and report. Exposing students to new habits of self-directed learning may help them take charge of their own learning and develop proficiency. They can choose to work by themselves or in groups, however there must be a negotiation on the type of process involved, the life situation explored and the ways it will be explored and accounted for. Heilman and Stout (2005) indicate possible stages that can help language instructors stimulate the creation of educative projects among their students:

(a) generate ideas together and outline the project by determining what groups will be formed and what will be the role of each (teachers should not accept duplicate projects);

(b) visualize anticipated projects and prepare possible scenarios;

(c) conduct an Internet search, multimedia exploration and strategic skimming of data; engage in inquiry and summary writing; and practice the interviews among peers—initial contact, warmup, interview and close—before the actual experience;

(d) refine projects for the report phase and prepare and rehearse presentations;

(e) present the individual, peer or group projects, which can be done using various media, and conduct self- and peer-assessment as preps for instructional assessment; and,

(f) engage in post-active reflection on the work done, whereby students reflect on what they learned, the amount they use the tar-

get language, and the strategies that could have improved their action.

To sum up, self-directed projects can blend with the use of online resources and new technologies. Motivation research suggests that it would increase opportunities for deep learning, which might have an impact on proficiency growth, the hypothesis underlying the present study. Increased student satisfaction may result from that deep process, which can also be measured through course evaluations.

Proficiency Growth

Language teacher effectiveness and student learning or achievement are in large part measured on the basis of the growing ability by the language learner to speak and interact with the language, as measured on a scale that the American Council for the Teaching of Foreign Languages (ACTFL) has taken years to perfect. The movement towards proficiency in language studies is often known through its instrument: The Oral Proficiency Interview (OPI) is a structured conversation between one student and one rater lasting 20 to 30 minutes (ACTFL, 1999). The ACTFL OPI is used worldwide by universities and public and private agencies for purposes such as placement, assessment, program evaluation, professional certification, hiring and promotion. Research in second language acquisition indicates that the quantity and variety of target language input affects student learning, which is a sound rationale to make sure language instructors have reached a reasonably high proficiency level (Chambless, 2012; Weyers, 2010), which was the case in the present study.

The OPI is recognized by the American Council on Education for college credits, and more than 10,000 evaluations in 37 languages have been conducted through this testing program. The ACTFL guidelines for the evaluation of oral proficiency are associated to robust criteria. Its raters are submitted to thorough practical training that, in many cases, may last almost 1 year after the initial OPI training workshop (Swender, 2003). After an initial familiarization workshop, tester training for the OPI is organized in intensive 4-day sessions focusing on the techniques of administering and rating the proficiency interview. The rating scale and techniques for eliciting a ratable sample are highlighted in daily sessions and reinforced in language-specific sessions. Participants can observe demonstration

interviews and then conduct practice interviews with volunteer candidates under trainer supervision. Following this initial training, they will practice at all levels of the rating scales and send recordings for feedback and evaluation, until certified.

During their OPI training, OPI raters must submit recordings at the various levels of proficiency, which often require double the amount of interviews before selection of these levels and thresholds. Then an experienced rater evaluates the ratings, some feedback is provided and more often than not a second series of recordings is required before the final rater certification. The rater certification is temporary and its renewal depends upon additional supervised training and practice. The OPI scale is considered robust because—after much training—there is satisfactory reliability across testers and across tests by the same evaluator, and the criteria are robust enough to be applied and replicated without major bias. Yet the ACTFL OPI has been criticized for various reasons, some of which are its lack of sensitivity over the span of one semester (Young, 2001) and even over an 8-week intensive summer course associated with immersion abroad (Davidson, 2010a). Other criticisms relate to the artificial conversation in the interview, which simplifies interpersonal communication; moreover, when OPIs are done by phone, the lack of face-to-face interaction may lead one to question its external validity, since interviewees interact without the semiotic and behavioral clues typical of face-to-face conversations.

Therefore, there was some initial hesitancy at the time of defining the evaluation scheme of the present study as to whether the OPI scale should be used and would be sensitive enough to measure the results of Deep Approach program. Nonetheless the OPI is the most commonly used measure of proficiency in U.S. universities, and therefore there was not much choice in terms of the availability of raters. Most major Turkish programs have certified OPI raters, and a majority of language programs integrate the OPI in the normal student evaluations at the beginning and end of each semester, or at the beginning and end of the academic year. Thus the ACTFL OPI was integrated in the assessment model. Turkish OPIs are based on the Provisional Proficiency Guidelines for Turkish prepared in 1993 by the American Association of Teachers of Turkic Languages Proficiency Guidelines Working Committee; they are informed by the ge-

neric ACTFL proficiency guidelines. “The Guidelines afford the language teaching communities a means to describe proficiency levels that are language specific while at the same time providing linkage to the corresponding proficiency levels in other languages” (AATF, 1993, p. 1).

OPI Research Methodology

What is the effectiveness of the Deep Approach and of the use of this instructional material in terms of oral proficiency growth and course evaluations? To respond to this research question, ACTFL OPI and course evaluations were used to measure the effectiveness of the Deep Approach with its online resources. The OPI research protocol is presented first. The research protocol for course evaluations is presented after the OPI results. The ACTFL website mentions the characteristics of the OPI¹: it is “a valid and reliable means of assessing how well a person speaks a language. It is a 20–30 minute face-to-face or telephonic interview between a certified ACTFL tester and an examinee. The interview is interactive and continuously adapts to the interests and abilities of the speaker. The speaker’s performance is compared to the criteria outlined in the ACTFL Proficiency Guidelines.”²

OPI Data Collection

Testing sites: The tests were organized in the Turkish programs of five Big Ten and Ivy League universities.

Participants: The experimental group consisted of volunteer students from the Deep Approach programs in these universities. Following appropriate research procedures, students could decide that their ratings would be used for research or not. Twenty-four intermediate students and 21 advanced students for a total of 45 students OPI-tested at least twice. The control group consisted of 20 intermediate students and 11 advanced students for a total of 31 students OPI-tested at least twice.

¹ <http://www.actfl.org/professional-development/certified-proficiency-testing-program#opi>

² See the ACTFL website:
<http://www.actfl.org/sites/default/files/pdfs/ACTFLProficiencyGuidelines2012-Speaking.pdf>

OPI Testers: Each of the programs had an ACTFL certified OPI tester. All the raters were native speakers of Turkish, ACTFL certified and experienced OPI raters. Experience in Turkish teaching ranged from 7 to 10 years, and experience in OPI testing ranged from 5 to 11 years. The testers were among the most experienced of the Turkish instructor community in the United States.

Quasi Experimental treatment: The treatment was the integration of online instructional material specially created for Deep Turkish Learning, as described in the guidelines for the Deep Approach to languages and cultures (Tochon, 2013). In most language programs, an ACTFL trained lecturer is in charge of the OPI assessment, which is an integral part of the curriculum. Since the OPI is often the basis of some percentage of the students' grades, there is often a double check of oral recordings. One-rater OPIs are named "unofficial OPIs." Audio recordings were systematic in two universities, not in the others. Some programs include OPIs at the beginning and end of each semester, some only provide ratings at the beginning of the Fall semester and at the end of the Spring semester. It was not always possible to evaluate oral proficiency at the beginning and end of the same semester.

OPI Data Analysis

For the purpose of statistical analysis, the rating scale used numerical equivalents for each OPI level, as shown in Table 1. These numerical equivalents differ from Kenyon and Tschirner (2000); however, what is important is the order of these ordinal data rather than the value because non-parametric statistics are used, based on the median.

Table 1. Numerical Conversions of Proficiency Levels

Proficiency Rating	Numerical Equivalent
Novice Low (NL)	1
Novice Mid (NM)	1.35
Novice High (NH)	1.65
Intermediate Low (IL)	2
Intermediate Mid (IM)	2.35
Intermediate High (IH)	2.65
Advanced Low (AL)	3
Advanced Mid (AM)	3.35
Advanced High (AH)	3.65
Superior (S)	4

In the following sections, we explore the OPI data and then course evalua-

tions are compared.

OPI Results

The following two tables present the results of the OPIs for the experimental group and the control group in the Intermediate Course (Table 2) and the Advanced Course (Table 3).

Table 2. Pre-Fall Pre-Spring Post-Spring Intermediate Comparisons

INTERMEDIATE	EXPERIMENTAL GROUP		CONTROL GROUP		
	PRE-FALL	PRE-SPRING	POST-SPRING	PRE-FALL	PRE-SPRING
UNIVERSITY A			UNIVERSITY A		
IM	IH		NH	IL	IM
NH	IL		IL	IM	IH
IL		IH	IL	IM	AL
IL		IH	NH	IM	IH
	IM	IH	NH	IL	IM
	IM	IM	IL	IM	IM
	IM	IH	IM	IH	IH
IL	IH	AM	IL	IM	IH
NH	IM		UNIVERSITY D		
IM	IH		IM	IH	AL
IL		AL	IL	IM	AL
IL		IH	IL	IM	
UNIVERSITY B			NH	IL	IH
NM		IL	NH	NH	IL
NM		IM	IM	IH	AL
NM		IL	NH	IL	IM
NM		IL	NH	IL	IM
NM		AL	NH	IL	IM
NM		IH	UNIVERSITY E		
IL		IH		AL	AL
NM		AL		IH	AL
NM		IM		AL	AL
NM		IM			
UNIVERSITY C					
NH		IL			
IL		IM			

Table 3. Pre-Fall Pre-Spring Post-Spring Advanced Comparisons

EXPERIMENTAL					
ADVANCED	GROUP		CONTROL GROUP		
PRE-FALL	PRE- SPRING	POST- SPRING	PRE- FALL	PRE- SPRING	POST- SPRING
UNIVERSITY A			UNIVERSITY A		
AL		AM	IM	IM	IH
IH		AM	IM	IH	AL
	SUP	SUP	IH	IH	AL
	AH	AH	IH	AL	AM
	AH	AH-SUP	AL	AM	AM
AL	AM	AH	AM	AH	AH
IH	AL		AM	AH	AH
AM	AH		IH	AL	AL
AM		AH	IM	IM	AL
AL		AM	AL	AM	AM
UNIVERSITY B			AM	AM	AH
IM		AM			
IM		AL			
IM		AL			
IM		AM			
IM		AM			
IM		AL			
IM		AL			
AL		AM			
AL		AM			
UNIVERSITY C					
IM		IH			
IL		IH			

Progress across OPI levels and across OPI thresholds was noted in Table 4 (Intermediate Course, 2nd year) and Table 5 (Advanced Course, 3rd year). The results presented below differentiate proficiency growth and when a proficiency threshold had been passed. *Thresholds* represent progress from Novice (1) to Intermediate (2), from Intermediate to Advanced (3), or from Advanced to Superior (4). For example, if Laura progressed from IM (2.35) to AL (3), she gained two levels (IH and AL) and passed one threshold (the Advanced threshold): both are mentioned in each Table.

Table 4. Levels of Oral Proficiency Progress in Two Semesters

ACTFL Unofficial OPI	EXPERIMENT INTERMEDIATE One semester N = 7	EXPERIMENT INTERMEDIATE Two semesters N = 17	CONTROL INTERMEDIATE One semester N = 4	CONTROL INTERMEDIATE Two semesters N = 16
0 Progress	1 student	-	2 students	-
1 Level of Progress	5 students	2 students	2 students	3 students
2 Levels of Progress	1 student	7 students		9 students
3 Levels of Progress		4 students		4 students
4 Levels of Progress		2 students		-
5 Levels of Progress		2 students		-
1 OPI Threshold	2 students	10 students		13 students
2 OPI Thresholds	-	2 students		-

Table 5. Levels of Oral Proficiency Progress in Two Semesters

ACTFL Unofficial OPI	EXPERIMENT ADVANCED One semester N = 5	EXPERIMENT ADVANCED Two semesters N = 16	CONTROL ADVANCED Two semesters N = 11
0 Progress	2 students AH+S	-	-
1 Level of Progress	3 students	4 students	8 students
2 Levels of Pro- gress	-	6 students	3 students
3 Levels of Pro- gress		3 students	-
1 OPI Threshold		6 students	5 students

Discussion

What the results in Table 4 and 5 tell us is that the most engaged students using a Deep Approach, who were given a greater autonomy and the power to be curriculum builders, had the opportunity to progress faster towards higher levels of proficiency. Qualitative data helped understand how such results had been made possible (Tochon, Ökten, Karaman, & Druc, 2012). Here is an example. In one university, an intermediate group of students decided to organize their own extensive Reading Club. They found a series of short novels and decided to read a different one each week, which was far beyond what their instructor would have normally advised, but this was their decision as a group. Self-directed team decisions and extensive reading are aligned with the Deep Approach model (Tochon, 2010c, in press), so the instructor did not object. Each student would read a novel that would be different from the ones read by their peers, and would prepare vocabulary and pre-reading scaffolds for the future readers, and they would discuss their readings in their weekly Reading Club meetings. During the first few weeks, the students who chose to create an extensive Reading Club faced a situation similar to that of students who are suddenly immersed in a foreign language abroad. During the first weeks they were lost but kept up with the task because this activity was their own decision. After 4 to 6 weeks they were able to cope with their extensive readings and its fast pacing, moreover they could enjoy the pre-reading materials of their peers, as well as team discussions as scaffolds for forthcoming readings. Some leaped three to five OPI levels and transcended one threshold in one semester.

This is not to say that extensive reading alone made them move up so many OPI levels and one threshold: autonomous language learning with the Deep Approach was more likely the real motivator. It forced students to face their choice for the amount of reading, as well as book choices and collaborative work over the readings. They persevered because these had been autonomous choices. They would have resisted against such choices had they been imposed, taught, and evaluated by the instructor. Extensive reading might not provide the same results with another group that had not chosen on its own initiative to enact this recommended aspect of the Deep Approach.

The interpretation of the results has to be mitigated. Firstly the instructors did not report their detailed monthly activities, which had been initially requested. We learned afterwards that some of them had maintained 2 hours of direct instruction in grammar per week while dedicating the rest to the Deep Approach. To free instructors from grammar instruction, the instructional materials proposed grammar storytelling on video that students could consult when needed. It may be anticipated that within environments in which autonomous learning becomes more acceptable and instructors agree to attend intensive training on the new approach, the results would be even stronger. This aspect was emphasized in Tochon, Ökten, Karaman, and Druc (2012). Secondly, the control groups were particularly strong, as would be attested by comparisons with OPIs in other languages of similar difficulty such as German. It was very difficult to find control group data for the OPI, which led this author to include data from University D, whose instructor is knowledgeable of the Deep Approach and admitted she was using similar strategies in part of her courses.

In the following sections, descriptive statistics and inferential statistics are used to determine the significance of the OPI results and verify their generalizability and consistency, first in University A, where the results have reasons to be more consistent because of the presence of a single evaluator across years and across experimental and control groups, and then for all university sites.

OPI Statistical Analyses for University A

The first descriptive and inferential analyses focus on the results obtained in University A. It is the university in which the largest number of data was obtained, which allows comparisons across experimental and control groups. The way of assessing proficiency was most probably highly consistent because it operated on the same site by the same experienced evaluator over many years, with and without the Deep Approach and its instructional hyper textbook (Tochon, 2013). Table 6 indicates a progression in proficiency among students in both control and experimental groups, with an advantage in groups using the Deep Approach (Intermediate: compare 2.58 for the control group to 2.74 for the experimental group; Advanced: compare 3.24 for the control group against 3.60 for the experimental group in the

post tests).

Table 6. Descriptive Statistic of OPI Results in University A Means and Standard Deviations in the Intermediate Course and the Advance Course

OPI GROUP	EXPERIMENTAL Mean (<i>SD</i>)			CONTROL Mean (<i>SD</i>)		
	PRE-FALL	PRE-SPRG	POST-SPRG	PRE-FALL	PRE-SPRG	POST-SPRG
INTERMEDIATE	2.00 (.25)	2.42 (.23)	2.74 (.30)	1.91 (.25)	2.30 (.21)	2.58 (.23)
ADVANCED	3.00 (.29)	3.55 (.34)	3.60 (.24)	2.82 (.41)	3.03 (.48)	3.24 (.33)

The OPI test is on an ordinal scale, therefore non-parametric statistics were used first to explore the differences across groups and second to analyze the pre-post differences within each group. For University A, the U-Test of Mann-Whitney for independent groups indicated a significant difference between the experimental and control posttests in the Advanced Course ($U=18$; $U'=70$; $p=.026$).

The pre-Fall vs. post-Spring comparison with the Wilcoxon Sign Rank test in each group indicated that both the experimental group and the control group in both the Intermediate Course and Advanced Course were demonstrating significant progress in proficiency, with a minimal probability of error of $p=0.002$. The Friedman ANOVA for related groups confirms this progress across the three testing periods of pre-Fall, pre-Spring and post-Spring (Table 7). Note: Missing data had to be replaced by column means to operate the Friedman ANOVA.

Table 7. OPI Progress in University A Measured with the Friedman ANOVA

OPI	EXPERIMENTAL GROUP PROGRESS Chi-Square (<i>Prob</i>)	CONTROL GROUP PROGRESS Chi-Square (<i>Prob</i>)
INTERMEDIATE	23.04 ($p<.0001$)	22.17 ($p<.0001$)
ADVANCED	18.17 ($p=.0001$)	15.88 ($p=.0004$)

For an ultimate verification, an ANCOVA was used to confirm the differences for each testing time and group, comparing pre-Falls, pre-Springs, and post-Springs for both experimental and control groups. The rationale for using an ANCOVA is that no equivalent measure exists in the non-parametric arena. The results indicate $F=12.11$ for the Intermediate Courses ($p=.0002$) and $F=3.71$ for the Advanced Courses ($p=.037$). These results can be considered food for thought, confirming the prior results of significant pre-post differences and differences across groups.

The Alpha post hoc reliability of the University A measures was verified with Hoyt's analysis, which indicated an F-ratio of 58.94 between items ($p<.0001$) and an F-ratio of 2.33 between cases ($p=.01$), which were highly significant.

OPI Statistical Analyses for All Sites

The second set of analyses focuses on the results obtained in all university settings, including University A. Various evaluators were involved in this measure, described in Table 8.

Table 8. Descriptive Statistic of OPI Results in All Sites Means and Standard Deviations in the Intermediate Course and the Advance Course

OPI GROUP	EXPERIMENTAL Mean (<i>SD</i>)			CONTROL Mean (<i>SD</i>)		
	PRE-FALL	PRE-SPRG	POST-SPRG	PRE-FALL	PRE-SPRG	POST-SPRG
INTERMEDIATE	1.73 (.35)	2.42 (.23)	2.55 (.38)	1.90 (.27)	2.36 (.35)	2.65 (.32)
ADVANCED	2.64 (.40)	3.55 (.34)	3.31 (.24)	2.82 (.41)	3.03 (.48)	3.24 (.33)

For all university sites, the pre-Fall vs. post-Spring comparison with the Wilcoxon Sign Rank test in each group indicates that both experimental and control groups in both the Intermediate Course and Advanced Course were demonstrating significant progress in proficiency, with probabilities of error ranging between $p=.003$ and 0.0002 . Note: The reason why the Advanced experi-

mental group has better results in pre-Spring comes from the small number of ratings obtained at that time from University A. See Table 2. The Friedman ANOVA for related groups confirms progress across the testing period with a probability of error inferior to .0001 (Table 8).

Table 8. OPI Progress in All University Sites Measured with the Friedman ANOVA

OPI	EXPERIMENTAL GROUP PROGRESS Chi-Square (<i>Prob</i>)	CONTROL GROUP PRO- GRESS Chi-Square (<i>Prob</i>)
INTERMEDIATE	41.17 ($p < .0001$)	47.57 ($p < .0001$)
ADVANCED	43.56 ($p = .0001$)	45.39 ($p < .0001$)

As an ultimate verification, ANCOVAs confirm the differences for each testing time and group, comparing pre-Falls, pre-Springs, and post-Springs for both experimental and control groups, which non-parametric statistics do not offer. $F=16.03$ for the Intermediate Course ($p < 0.0001$) and $F=9.59$ for the Advanced Course ($p = .0003$). Thus there are significant differences across groups and across time. The Alpha post hoc reliability measures were verified with Hoyt's analysis, which indicated an F-ratio of 128.06 between items ($p < .0001$) and an F-ratio of 2.58 between cases ($p < .0001$), which were highly significant.

Discussion

Comparisons with other corpora are possible (see Table 9). As a test, the ACTFL OPI is well documented in undergraduate language courses and comparisons abound—while we must admit that they are very rare for the Turkish language, a language that presents special difficulties to students because of its morphological cases, and its reversed and agglutinative structure. Tschirner and Heilenman (1998, p.156) compared 20 students at the end of their fourth semester and found results ranging from Novice High to Intermediate Mid, which was consistent with seven prior university studies in French, German, and Russian: “It was previously assumed that the median OPI rating

of the fourth semester students of German is IM. The results ... seem to indicate that OPI level expectations need to be lowered by one sublevel". This issue relates to program expectation. The median results in the present study were higher (Table 9).

Table 9. Oral Proficiency Range and Median in the present study

Course Level	Pre-Fall	Post-Spring
Intermediate	NH	IH
Advanced	IH	AM

To sum up, the oral proficiency results indicate that both experimental and control groups were successful in significantly improving the students' proficiency across the semesters of both the Intermediate and the Advanced courses (2nd and 3rd year of study). Furthermore, there are significant differences between the experimental and the control groups at both levels, which means that students developed their skills differently. The results are reliable, consistent, and have been confirmed by inferential statistics over time and across groups. What then is the difference for the Deep Approach group? The differences in progress pacing indicate that the best students improved more swiftly and in more important ways with the Deep Approach, to the point of reporting results superior to study abroad programs (Davidson, 2010ab). While average results are better at the Intermediate level, the Deep Approach makes a stronger difference in terms of proficiency growth at the Advanced level. These results were obtained in conditions in which not all instructors had been fully trained to the new approach, and in which the control groups were particularly strong. Project-based learning is an environment of blended interactions in which learners are curriculum builders, using online resources for their project, and provides valid, successful and reliable proficiency growth.

Course Evaluations Methodology

All universities use standardized forms to assess the success of instruction on the basis of the evaluation made by the students. It is not easy to enter into the logics of the tests used by various institutions, some focusing on content effectiveness and planning, others

on pedagogy, rapport, and fairness. Not only were the scales different from university to university, they might vary from year to year in the same university. They may focus on the students' own perception of various dimensions of the instruction, such as the articulation of course goals, the emphasis on language skills and culture, the value of homework and assignment, course quality and rapport with the class, appropriateness of activities and course pacing, and usefulness of the approach and challenges in terms of language abilities. Many of the items permit an evaluation of the program's approach and thus can illuminate some aspects of the study. The control groups' comparisons focus on courses given by the same experienced instructors between 2007 and 2012 with various classes. Descriptive and inferential statistics are used to study the responses to similar items. Thus the methodology used here was to find items that would (1) correspond across scales, and (2) would be relevant to aspects of the present study.

Course Evaluations by Students

In this section, the course evaluations of different classes with or without the Deep Approach are compared. The methodology to compare course evaluations and allow statistical analysis was complicated by the presence of competing scales across universities (0 to 4 or 1 to 5), and of questionnaires that were different across institutions and even across years at the same institutions. The statistical differences are examined from a descriptive and then from probabilistic perspective: (1) across scales presenting high levels of similarities, such as similar questions, see Tables 10 and 11; and, (2) across three universities having different scales for which semantically similar questions could be found; a selection of ten questions presenting a high level of similarity across universities has been done, and the 0–4 scales were adjusted to 1–5 to ease comparison (Table 12).

Intermediate Level. The first comparison was made at the intermediate level (Table 10).

Table 10. Descriptive comparison of course evaluation means in Intermediate courses of one University

Intermediate Course Evaluations in one University across years	Non-Deep Approach 12 students	Deep Approach 10 students
1. Course goals clearly articulated	3.17	3.72
2. Appropriate emphasis on language skills	3.57	4.00
3. Useful readings & culture materials for language acquisition	3.67	3.92
4. Valuable homework as reinforcement of classroom work	3.17	3.53
5. Exams consistent with assignments, method & materials	3.42	3.81
6. Challenged linguistic abilities in reading, writing, speaking and listening	3.50	3.83
7. Course pacing	2.17	1.92
8. Weekly hours spent outside class	1.84	.95
9. Overall course quality	3.59	3.92
10. Instructor was able to stimulate interest in material	3.92	3.92
11. Appropriate activities to encourage oral skills	3.83	3.92
12. Care for students learning	3.92	3.92
13. Rapport with the class	3.92	4.00
14. Instructor attitude towards the course	3.92	4.00
15. Instructor effectiveness in presenting & explaining course materials	3.42	4.00
16. Instructor availability outside of class	3.92	4.00
17. Fairness in grading process	3.83	3.92
18. Overall quality of instructor	3.67	3.92
19. Survival immersive skills learned in the course	3.88	4.00

Analysis

The comparison of the means presented in Table 10 is in favor of the experimental group in these Intermediate level courses. The students perceived the goals were clearly articulated with the Deep Approach, which was possibly related to the online modules presenting the re-

lated 5C standard goals. The template used for projects emphasized the language skills appropriately. The readings and culture materials were perceived as more useful for second language acquisition, and the homework was more valued. Evaluations were perceived as more consistent and the abilities of the students were challenged in a higher fashion. However pacing (question 7 in Table 10) was perceived as somewhat quieter (from 2.17 down to 1.92) and these intermediate students had the perception of spending less time working outside class (which is contradicted by the results of question 8 in advanced courses as shown in Table 11). Overall intermediate courses with the Deep Approach were perceived as having higher quality, yet the terms “course quality” in the questionnaires may suggest a reference to directed courses, which was not the case here. Teacher rapport, attitude, effectiveness and availability outside of class were perceived as higher than in usual courses. The Deep Approach provided the survival skills for immersion abroad.

The statistical differences in Table 10 were subjected to statistical analysis to verify whether they were significant. For that purpose, the U-Test of Mann-Whitney was used, which is the appropriate non-parametric measure for two independent groups. The U-Test indicates that the differences between control groups and experimental groups are significant in favor of the Deep Approach, with an error probability threshold of $p=.0068$ ($z\text{-value}=-2.71$) allowing generalization.

Advanced Level. The second comparison was made at the advanced level (Table 11).

Table 11. Descriptive comparison of course evaluation means in advanced courses of one university

Advanced Course Evaluations in one University across years	Non-Deep Approach 9 students	Deep Approach 15 students
1. Overall quality of instructor	3.88	4.00
2. Overall course quality	3.90	3.78
3. Instructor ability to communicate the subject matter	3.75	3.88
4. Instructor ability to stimulate student interest	4.00	3.88
5. Instructor accessibility and willingness to discuss course content & any problems	4.00	4.00
6. Value of assigned readings	3.75	4.00
7. Course difficulty	2.10	3.63
8. Amount of work required for the course	2.45	3.63

Analysis

In the advanced courses of one university, the quality of the instructor was highly noted in both groups. However, its grading was higher in the Deep Approach, with a higher note for the instructor's accessibility and willingness to discuss course contacts and any problems that might pop up, which is typical of a Deep Approach. The factors of communication about the subject matter and course quality were perceived as lower than in other courses, possibly due to the nature of self-directed projects. Therefore, course difficulties and effects were perceived as more important at the advanced level. Advanced students used to systematic teaching might have some difficulty adapting to a regimen in which they are in charge of their own learning and projects. The results shown in Table 11 were obtained in one university only, and the U-Test statistic indicates that they are *not* significant ($p=.52$; $z\text{-value} = -.64$). The next table, Table 12, presents the overall results of intermediate and advanced students in three classes for 10 factors that are common to all settings.

Table 12. Descriptive comparison of course evaluation means on ten questions presenting a high level of similarity in three universities

Questions	Non-Deep Approach 52 students	Deep Approach 49 students
1. Clear & effective presentation of subject matter	4.65	4.75
2. Instructor's ability to help clarify course readings & materials	4.57	4.75
3. Instructor's responsiveness to students' questions, opinions & criticisms	3.79	4.85
4. Instructor's ability to stimulate intellectual curiosity	4.76	4.79
5. Instructor's feedback outside class	4.62	4.82
6. Appropriateness and fairness of grading	3.67	4.77
7. Overall value of readings for second language acquisition	3.68	4.72
8. Overall value of assignments & homework	3.36	4.62
Questions	Non-Deep Approach 52 students	Deep Approach 49 students
9. Contribution to knowledge of the subject matter	4.86	4.74
10. Contribution to interest in the discipline or subject matter	4.76	4.79

Analysis

Table 12 shows the comparison of similar questions in three universities with a larger number of students with and without the Deep Approach. The item means indicate a superior appreciation for all items for the Deep Approach, with very high values (between 4.62 and 4.85 contrasting with much lower appreciations in the control groups). The only exception is question 9, as the contribution to subject matter knowledge was somewhat lower than 4.86; yet 4.74 remains an excellent course evaluation. Some students might perceive that they get less content when they explore the subject matter on their own, rather than getting the "master's voice" from the instructor. Yet here the difference was minimal. The U-Test of Mann-Whitney indicates that the difference is significant with a probability of error of $p=.031$ (z -value=-2.16). Thus, compared to control groups with the same

instructors using a different approach, the Deep Approach program had significantly better course evaluations.

Discussion

Course evaluations based on student responses introduce a series of biases that are well studied in that literature. Variables other than the program approach might account for the discrepancies between the two sets of data, such as the amount of experience held by the teacher at the time of evaluation or the proficiency level of the students. These possibly parasite variables could be controlled in a future study. Moreover, we kept here some items that might at first sight seem superfluous, such as grading satisfaction or instructor's ability to answer questions. However, these items actually are dependent upon aspects of the program chosen: grading satisfaction may be higher when self-assessment and peer-assessment are accounted for, and the ability to free time for personal questions at least partly depends upon the teaching approach in the program. The results of course evaluation are an indication that the orientation is positive and successful; they provide food for thought and they confirm and triangulate the subjective perceptions of the language instructors, analyzed in other studies (Tochon, Ökten, Karaman, & Druc, 2012 and in press).

To sum up, the strongest results of this section indicate that students perceive that Deep Approach-oriented courses with blended resources and self-determined projects offered clear and effective presentations of subject matter; the instructor was able to clarify course readings and materials; the instructor was responsive to questions, opinions, and criticisms; and the instructor stimulated the students' intellectual curiosity. The instructor provided enough feedback outside class. The students valued the assignments and homework highly, which contributed greatly to their interest in the discipline and subject matter. Grading was perceived as highly appropriate and fair (compare 4.77 to 3.67 for the control group).

These results are significant and can be generalized. In addition, Hoyt's analysis provides a high coefficient of reliability of these data (.84) for the course evaluations and indicates internal consistency and a high level of generalizability, with high reliability between items ($p < .0001$) and between cases ($p < .0001$). Course evaluations are signif-

icantly better with a Deep Approach program.

Conclusion

Studies on Turkish learners in the United States are rare. The approach used to deep language learning is “new.” Therefore, this study is unique in the data it provides *and* the results it demonstrates, which support a deep, blended approach to language learning. The data showed a link between learning with online resources used in self-determined thematic projects and high student language performance in Turkish in unofficial OPI measures. Course evaluations were also significantly higher. Measures of reliability indicate that the results are consistent. Furthermore, they were triangulated by qualitative data (Tochon, 2013): the connection with higher levels of proficiency was noticeable for the Turkish instructors who used various forms of assessment current in their programs, such as conversations, formative and summative evaluations, individual and group comparisons across years, drills, and examinations, which helped them ground their professional judgment. Self-directed projects blend with the use of online resources and increase opportunities for deep language learning and proficiency growth. Increased student satisfaction resulted from that deep process, which was measured through course evaluations.

Blended learning is gaining primacy in language learning as instructors and students try to integrate new technological environments into classroom activities, which places instructors as facilitators of learning, providing the learners with opportunities to work through both face-to-face and online interactions (Duhaney, 2012). Yet, as recently mentioned in two online journals, “blended learning begins with, and relies upon, skilled teachers” (Morrow, 2012), and “teachers must give up some control” (Fortson, 2012). These issues are crucial and require professional development. Another issue that appears so important when designing blended learning environments is to scaffold deep learning designs: a conceptual basis is needed for sound design, applicable across personal technological environments. Deep learning design shapes the possibilities for “technology to most effectively enhance learning” (Boyle & Ravenscroft, p. 1224). The Deep Approach website created conditions for such deep linguistic and cultural exploration. As was noted at the beginning of this article,

conceptual clarification is essential if we are to establish a stable and deep discipline of technology-enhanced learning. The technology is alluring; this can distract from deep design in a surface rush to exploit what the new technology affords.

One limitation of the study is the use of a speaking test (the ACTFL OPI) to evaluate the effectiveness of the Deep Approach as a methodology, which targets the improvement of all basic skills. There is no single, reliable, well-established, easy-to-implement proficiency test for Turkish language assessing all language skills. European-based comprehensive proficiency tests assume different proficiency guidelines. The Reading and Listening tests of the National Middle East Language Resource Center could be incorporated into future research for more complete measures. Life in a less than ideal world does not always allow researchers to obtain exactly the neat set of data that experimentalists might wish for. Instructors are not subservient to researchers and may refuse or have difficulty providing the needed comparison groups, which attests to a participant's freedom that is explicitly encouraged by Institutional Review Boards.

Regarding the OPI, the literature provides comparative measures (Malone, Rifkin, Christian, & Johnson, 2003; Tschirner & Heilenman, 1998). Reliability coefficients provide only one perspective and should be viewed within the larger system of reliability analysis of the OPI literature. The idea that the numerical values of OPI levels could be compared in an experimental design with a control group emerges from the conviction that such tests are comparable. It should be remembered that the concept of comparability rests upon a probability that settings were comparable, testers had comparable grading strategies, and that variability can be accounted for by the probability thresholds of the statistics. In this research of an exploratory nature, we adopt a more modest stand and perceive the comparison as food for thought. The statistics must be understood as a reflective tool.

Another aspect that might be questioned in the present study is whether using the course evaluations of different years with the same instructors might neglect the growth of experience that may increase student course ratings; however, these instructors already had more than 7 years of experience teaching in the initial courses used in the comparison and the effort, for them, to learn how to

handle the new approach in an institutional context in which self-determined learning is not welcome could have produced opposite evaluations, and could have led to a decrease in student course ratings.

The type of online resources that provide for a blended, Deep Approach to language learning represents a new orientation and an important contribution to less commonly taught language teaching and learning, as this is a field in which textbooks are rare and sometimes obsolete. We need more research on the challenges that face teachers to help their students initiate self-determined projects for a deeper learning of world languages and cultures. Indeed, teachers face a certain number of challenges when they help their students implement projects, namely adapting students' pacing and work to class periods and the administrative pressure for them to cover the master study plan of their foreign language department.

Inquiry-based projects take time in a class schedule compared to so-called teacher-centered efficiency, but depth is a sounder choice than breadth with surface learning. The ability to stimulate various groups at the same time in the same class is to be developed in higher education. Another challenge is that, as many resources are online, teachers must continually update their technology practices (Boss, Krauss, & Conery, 2008; Nissen & Tea, 2012). Students might want to build up their projects online in electronic portfolios and linguafolios that must be pre-structured by the teacher (Gulbahar & Tinmaz, 2006). In this context, teachers continue to play a crucial role in providing resources, enhancing motivation, and empowering the students to become curriculum builders. Teachers then stimulate, not stifle; they support learners' accomplishments by encouraging their efforts. The creation of free online resources for less commonly taught languages can motivate an increasing number of learners to start learning critical languages, and actually, the Deep Approach would work for any language.

Acknowledgements

The project, titled "A Deep Approach to Turkish Teaching and Learning", was funded by a 2009–2012 Title VI grant from the U.S. Department of Education. This study has been approved by the Institutional Review Board. We are grateful to the instructors who

shared their experiences with us, to Isabelle Druc, Project Director, and to the project assistants on this project, in particular Esra Alagoz, Yasin Tunc, and Mukaddes Şahin.

References

- American Association of Teachers of Turkish (1993). *Provisional Proficiency Guidelines for Turkish*.
- American Council on the Teaching of Foreign Languages. (1999). *ACTFL proficiency guidelines—speaking*. Retrieved from <http://www.actfl.org/files/public/GuidelinesSpeak.pdf>
- Beckett, G. H., & Miller, P. C. (2006). *Project-Based Second and Foreign Language Education. Past, Present, and Future*. Charlotte, NC: Information Age.
- Boaler, J. (2002). Learning from teaching: Exploring the relationship between reform curriculum and equity. *Journal for Research in Mathematics Education*, 33(4), 239–258.
- Boss, S., Krauss, J., & Conery, L. (2008). *Reinventing Project-Based Learning: Your Field Guide to Real-World Projects in the Digital Age*. Eugene, OR: International Society for Technology in Education.
- Boyle, T., & Ravenscroft, A. (2012). Context and deep learning design. *Computers & Education*, 59, 1224–1233.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18, 32–42.
- Carr, N. (2010). *The shallows: What the Internet is doing to our brains*. New York: Norton.
- Chambless, K.S. (2012). Teachers' oral proficiency in the target language: research on its role in language teaching and learning. *Foreign Language Annals*, 45(S1), S141-S162.
- Chen, W.-F. (2012). An investigation of various types of blended learning environments on student achievement: An experimental study. *International Journal of Instructional Media*, 39(3), 205-212.
- Davidson, D. E. (2010a). Study abroad: When, how long, and with what results? New data from the Russian front. *Foreign Language Annals*, 43(1), 6-26.
- Davidson, D. (2010b, October). *L2 gain, time-on-task, and language use in Arabic and Russian overseas Flagship programs*. Lecture given at the Language Institute, College of Letters and Science, University of Wisconsin-Madison on October 26, 2010 for the American Councils for International Education.

- Deci, E. L., Vallerand, R., Pelletier, L., & Ryan, R. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist*, 26, 325-346.
- Delialioğlu, Ö. (2012). Student Engagement in Blended Learning Environments with Lecture-Based and Problem-Based Instructional Approaches. *Educational Technology & Society*, 15(3), 310-322.
- Ding, H. (2008). The Use of Cognitive and Social Apprenticeship to Teach a Disciplinary Genre. *Written Communication*, 25(1), 3-52.
- Dinsmore, D., & Alexander, P. (2012). A Critical Discussion of Deep and Surface Processing: What It Means, How It Is Measured, the Role of Context, and Model Specification. *Educational Psychology Review*, 24(4), 499-567.
- Duhaney, D. C. (2012). Blended learning and teacher preparation programs. *International Journal of Instructional Media*, 39(3), 197-203.
- Entwistle, N. (2008, April). *Taking stock: teaching and learning research in higher education*. International symposium on Teaching and Learning Research in Higher Education. Guelph, Ontario, April, 25-26, 2008.
- Fortson, K. (2012). Using iPads With Mixed-Ability Students, Teachers Must 'Give up Some Control'. *THE Journal, Transforming Education through Technology*, December 9, 2012. Retrieved from: <http://thejournal.com/articles/2012/09/12/using-ipads-with-mixed-ability-students-teachers-must-give-up-some-control.aspx?admgarea=Features1&m=2>
- Ge, Z.-G. (2012). Cyber Asynchronous versus Blended Cyber Approach in Distance English Learning. *Educational Technology & Society*, 15(2), 286-297.
- Göçmençebe, S. I., Muhlis Özkan, M. & Bayram, N. (2012). Evaluating Primary School Students' Deep Learning Approach. *International Online Journal of Educational Sciences*, 4(3), 554-562.
- Gulbahar, Y., & Tinmaz, H. (2006). Implementing project-based learning and e-portfolio assessment in an undergraduate course. *Journal of Research on Technology in Education*, 38(3), 309-327.

- Heilman, J., & Stout, M. (2005). Putting projects into practice. In K. Bradford-Watts, C. Ikeguchi, & M. Swanson (Eds.), *JALT 2004 Conference Proceedings* (p. 587-591). Tokyo: JALT.
- Ikeda, D. (2010a). *Soka education: For the happiness of the individual*. Santa Monica, CA: Middleway Press.
- Jaffee, D. (2007). Peer cohorts and the unintended consequences of freshman learning communities. *College Teaching*, 55(2), 65–71.
- Karaman, C., Ökten, C., & Tochon, F. V. (2012). Learning the Deep Approach: language teachers' voices. *Porta Linguarum*, 18, 79-95.
- Keegan, D., Schwenke, E., Fritsch, H., Kenny, G., Kis-mihók, G., Bíró, M., Gábor, A., ó'Suilleabháin, G. & Nix, J. (2005). Virtual classrooms in educational provision: synchronous e-learning systems for European institutions. Retrieved from Fern Universitaet, Central Institute for Distance Education Research website: <http://www.fernuni-hagen.de/ZIFF/synchronous.pdf>
- Kilpatrick, W. H. (1918). The project method. *Teacher's College Record*, 19(4), 319-335.
- Larmer, J., Ross, D. & Mergendoller, J.R. (2009). *PBL Starter Kit*. Novato, CA: Buck Institute for Education.
- Lee, L. (2002). Synchronous online exchanges: A study of modification devices on non-native discourse. *System*, 30(3), 275–288.
- Lee, H.-J., & Lim, C. (2012). Peer Evaluation in Blended Team Project-Based Learning: What Do Students Find Important? *Educational Technology & Society*, 15(4), 214–224.
- Mahoney, S, & Schamber, J. (2011). Integrative and Deep Learning through a Learning Community: A Process View of Self. *JGE: The Journal of General Education*, 60(4), 234-247.
- Malone, M., Rifkin, B., Christian, D., & Johnson, D. (2003, January 23). Attaining high levels of proficiency: Challenges for language education in the United States. Paper presented at the Conference on Global Challenges and U.S. Higher Education, Duke University, Durham, NC.
- Merrow, J. (2012). Three fears about blended learning. *Washington Post*, September 22, 2012. Retrieved from: <http://www.washingtonpost.com/blogs/answer-sheet/post/three-fears-about-blended->

learning/2012/09/22/56af57cc-035d-11e2-91e7-2962c74e7738_blog.html

- Nissen, E., & Tea, E. (2012). Going blended: new challenges for second generation L2 tutors. *Computer Assisted Language Learning*, 25(2), 145-163.
- Norton, B. (2000). *Identity and language learning: Gender, ethnicity, and educational change*. New York: Longman/Pearson.
- Obelleiro, G. (2012). A moral cosmopolitan perspective on language education. *Critical Inquiry in Language Studies*, 9(1-2), 33-59.
- Pfister, H.-R. (2005). How to support synchronous Net-based learning discourse: Principles and perspectives. In R. Bromme, F. W. Hesse, & H. Spada (Eds.), *Barriers and biases in computer-mediated knowledge communication* (pp. 39-57). New York: Springer.
- Swender, E. (2003). Oral proficiency testing in the real world: Answers to frequently asked questions. *Foreign Language Annals*, 36, 520–526.
- Thomas, J. W. (2000). *A review of research on project-based learning*. San Rafael, CA: Autodesk Foundation.
- Thorne, K. (2003). *Blended learning: how to integrate online and traditional learning*. New York: Kogan Page.
- Tochon, F. V. (2010a). Deep Education. *Journal for Educators, Teachers and Trainers* (JETT), 1, 1-12. <http://www.ugr.es/~jett/articulo.php?id=1>
- Tochon, F. V. (2010b). A Deep Approach to Language Multimedia and Evaluation: For a more Colorful Future. Invited Keynote Speech. *Proceedings of the Fourteenth international conference of APAMALL and ROCMELIA* (pp.73-92). Kaohsiung, Taiwan: National Kaohsiung Normal University.
- Tochon, F. V. (2010c). Chapter 5. Forward Planning: Instructional Organizers for Deep Apprenticeship. In *Guide to a Deep Approach to World Languages and Cultures*. Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research.
- Tochon, F. V. (2013). *Online Instructional Hyper Textbook as Personal Environment for Deep Language Learning* (WCER Working Paper No. 2013-4). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research website:

- <http://www.wcer.wisc.edu/publications/workingPapers/papers.php>
- Tochon, F.V. (in press). *Help Them Learn a Language Deeply! Why You Need a Deep Approach to World Languages and Cultures*. Blue Mounds, WI: Deep University Press.
- Tochon, F.V., Ökten, C. E., Karaman, A. C., & Druc, I. C. (2012). *The Deep Approach to Turkish Teaching and Learning. Final Report to the U.S. Department of Education*. Blue Mounds, WI: Deep University Press.
- Tochon, F.V., Ökten, C. E., Karaman, A. C., & Druc, I. C. (in press). *Teaching Turkish Deeply. The Deep Approach to Language and Culture*. Blue Mounds, WI: Deep University Press.
- Tschirner, E., & Heilenman, L. K. (1998). Reasonable expectations; Oral proficiency goals for intermediate-level students of German. *Modern Language Journal*, 82, 147–158.
- Weyers, J. R. (2010). Speaking strategies: meeting NCATE oral proficiency standard. *Foreign Language Annals*, 43(3), 384-394.
- Wold, K. A. (2011). Blending theories for instructional design: creating and implementing the structure, environment, experience, and people model. *Computer Assisted Language Learning*, 24(4), 371-382.
- Young, R. F. (2001, February). *The entextualization of talk*. Paper presented at a joint symposium on “Defining and Assessing Speaking Ability” held at the annual meeting of the American Association for Applied Linguistics and the Language Testing Research Colloquium. February 24, 2001, St Louis, Missouri, USA.