Data-Driven Learning: Do Learning Tasks and Proficiency Make a Difference?

Shiauping Tian

(National Taiwan University of Science and Technology)

This paper examines the effectiveness of Data-Driven Learning on learning tasks with different instructional focuses and on students at different proficiency levels.   98 university students participated in the study.   A control/experimental group pre-test/post-test design was implemented. Statistical analysis of the results indicate that students in the DDL group improved significantly more on comparison of word usage and distinctive features of a text type but not on grammar. Differences in gain scores by students at different levels did not reach a significant level, although the data favored students with higher proficiency on two of the instructional focuses.

INTRODUCTION

Data-Driven Learning is a term coined by Tim Johns (1991) that refers to the learner sorting through large amounts of authentic language data to arrive at patterns or rules of language use.   The DDL approach has been widely acclaimed for its potential in language instruction, for it places the learner in the center of the learning process, encourages hypothesis testing and discovery, fosters learner autonomy, and allows the learner to develop important learning skills.   Over the years, researchers and language teachers have proposed and implemented DDL approach in the teaching of lexical collocations, affixes, polysemous words, grammatical patterns, and discourse features (Dyck, 1999; Ball, 1998; Tribble, 1997; Kettemann, 1995). A number of studies have also been conducted to provide empirical evidence to the instructional benefits of DDL in language classrooms (Hadley, 2002; Lee & Liou, 2003; Wang, 2002; Someya, 2000; Ilse, 1991; St. John, 2001).

In addition to the instructional benefits of the DDL approach, attention has also been directed to the factors that influence the effectiveness of such an inductive approach.   Researchers have discussed the impact of factors such as learner proficiency, learning style, and difficulty of the language pattern on the effectiveness of DDL.   In terms of learner proficiency, studies in the past did not seem to agree on who would benefit more from an inductive teaching approach (Carroll, 1964; Shaffer, 1989; Gross, 1991).   In a more recent study on the effectiveness of concordancing on learning of collocations, no significant difference was found between learners at different proficiency levels (Wang, 2002).

Another factor is the features of the language rules being taught; in other words, the complexity of the rules.   Researchers have proposed that an inductive approach was more

appropriate for learning of easy rules (Robinson, 1997; Reber, 1989).   In Wang's study, the factor of difficulty of collocation patterns was investigated, and the results indicate that the inductive approach is more effective for learning of easy patterns.

In a different study, Lee & Liou (2003) used DDL approach to assist vocabulary learning of high school students and investigated the effects of vocabulary level and preferred learning styles on the learning outcome.   Results of their study show that students at the low vocabulary level benefited more from the approach, indicating the potential of using this approach to help less proficient learners.   Another finding of their study is that, understandably, students who preferred inductive learning improved more after the DDL lesson.

Given the attention to the instructional benefits and factors involved in the effectiveness of DDL, this study was conducted to investigate the effectiveness of DDL on three different instructional focuses (grammar, comparison of word usage, and special features of a text type) and to attempt to answer the question about whether DDL has differential effects on learners at different proficiency levels.

PURPOSE

The purpose of this study was twofold—to see if the DDL approach was effective on each of the learning tasks and to see if the DDL approach had different effects on students at different proficiency levels.   The study was designed to answer the following research questions:

1. Is DDL an effective approach for learning tasks with different instructional focuses?
2. Does DDL work better with students at lower or higher proficiency level?
3. Does DDL work better with students at a particular proficiency level for any of the three instructional focuses?

METHOD

Participants

Participants of the study were 98 university students from two news media English classes taught by the researcher.   These students were all non-English majors from various departments in the university.

Length of study

The duration of the study was 5 weeks, not including administration of the pre-tests and proficiency test, with the class meeting 2 hours each week.

Table 1 Instructional Focuses and Pre/Post-Tests

	Gram 1	Gram 2	Word usage	Headlines
Instructional focus	Subjunctive verbs	Reduction of adverb clauses to modifying adverbial phrases	*Contract* vs *infect*	Syntactic features of news headlines
Length of instruction (wks)	1	2	1	1
Item type in Pre-test	Fill in the blanks	Error correction	Error correction	Conversion
Item type in Post-test	Fill in the blanks	Error correction	Error correction	Conversion

Learning tasks

The instruction included three instructional focuses—grammar, comparison of word usage, and distinctive features of a text type.   The grammar points focused on were (1) subjunctive verbs, and (2) reduction of adverb clauses to modifying adverbial phrases.   The second instructional focus was comparison of the usage of two verbs— "*contract*" and "*infect*".   Since the course was on news media English, the third instructional focus was the special syntactic features of English news headlines, including tense, use of punctuation, and reduction of certain word categories.

Teaching approach

Students in one of the two classes were assigned to the control group (TRA) and those in the other class to the experimental group (DDL).   Two teaching approaches were implemented in the two groups.   Students in the TRA group were instructed with the conventional approach.   The instructor first introduced the instructional focus, laid out the rules or patterns, and then provided examples taken from grammar books or news media English textbooks to illustrate the rules.   On the other hand, students in the DDL group, after an initial introduction of the instructional focus, were given printouts of search results from news-related concordancers and websites and told to locate relevant sentences, observe the sentences to formulate patterns or rules, and then find more examples to confirm or modify the hypothesized rules when needed.   Due to the large class size, the researcher decided to use printouts from concordancing results instead of hands-on concordancing, which was more realistic for most instructional contexts with large classes.

Materials for the DDL group were made based on search results from English news concordancers and related websites.   Slight adaptations were made in terms of layout and task specifications to make the materials easier to read and the instructional focus more explicit.   For example, for headline features, headlines in the search results based on a particular feature were listed along with news leads so that students can see what a particular feature stands for more easily.

To measure the students' progress after the instruction, the students were given a pre-test and a post-test for each of the instructional focuses.   The item types for each test were listed in Table 1. The perfect score for each of the test was 100.

To determine students' level of proficiency, a General English Proficiency Test (High-Intermediate) was administered to the students before the study.   Students' scores on the reading comprehension section were calculated as the basis for proficiency grouping.   Students who responded correctly to 50% of the items and above were categorized as high proficiency students, and those with less than 50% correct responses were categorized as low proficiency students.

RESULTS

To answer the questions raised in the study, results will be presented in two parts—(1) improvement by group and instructional focus, and (2) improvement by proficiency and instructional focus.

Results by group and instructional focus

The students' scores on each of the tasks are presented in Table2.   For all of the instructional focuses, students in both groups improved significantly after the instruction.   However, it seems that the pre-test scores for comparison of word usage were very low for both groups, especially the DDL group, and the post-test scores for this focus (only 57.86 on average) were also the lowest among the three instructional focuses.   This shows that students in both groups had very little knowledge about the distinction between the two words in focus, despite their frequent appearance in the news.   It also means that comparison of usage for these two words was rather difficult for the students, and that further instruction was needed for students to gain a clearer sense of how to use these two verbs.

Another observation is that since students in neither group were familiar with the syntactic features of news headlines before the instruction, their pre-test scores were very similar.   In addition, since the conventions for headline formation were rather straight forward, students in both groups scored pretty high on the post-test (87.38 on average).

Table 2 Scores on Pre/Post-Tests

Measure		Pre-test		Post-test		N	T	Sig.
Focus	Group	Mean	SD	Mean	SD		(Paired)	
Grammar	DDL	29.5	12.26	80.52	12.2	50	24.46	0.000*
	TRA	22.92	10.5	67.39	27.13	48	10.90	0.000*
	Total	26.28	11.85	74.09	21.80	98	21	0.000*
Word Usage	DDL	10.5	17.56	57.4	20.88	50	15.15	0.000*
	TRA	23.72	24.09	58.33	26.54	48	7.98	0.000*
	Total	16.93	21.93	57.86	23.7	98	15.97	0.000*
Headline Features	DDL	28.62	3.07	90.56	5.493	50	66.36	0.000*
	TRA	30.29	2.74	84.06	19.39	48	19.06	0.000*
	Total	29.44	3.02	87.38	14.42	98	38.33	0.000*

*p<0.05

Table 3 Gain Scores by Group and Instructional Focus

Group	DDL		TRA		T	Sig.
Focus	Mean	SD	Mean	SD	(independent)	
Grammar	51.02	14.75	44.51	28.39	1.43	0.155
Word Usage	46.9	21.9	34.41	30.28	2.35	0.021*
Headline Features	61.54	7.35	53.6	19.44	2.69	0.008*

* $p < 0.05$

To assess the effectiveness of the DDL approach, gain scores of the two groups were compared on each of the instructional focus.   Table 3 shows that a significant difference was found in two of the three instructional focuses.   Students in the DDL group improved significantly more than those in the control group in their learning of word usage and headline features.   In tasks of grammar, although students in the DDL group seemed to make more progress than the control group, the difference did not reach a significant level.   These results indicate that the DDL approach is effective in teaching comparison of word usage and special features of a text type.

Results by proficiency and instructional focus

To determine whether the DDL approach has differential effects on students at different proficiency levels, the students were divided into high proficiency and low proficiency groups according to their performance on GEPT.   Tables 4-6 present scores on the pre-test and post-test for each of the instructional focuses according to group and proficiency level.   Gain scores for each of the focuses by group and proficiency level can be found in Table7.

Table 4 Scores on Grammar

Grammar		Pre-test		Post-test		T	Sig.
Group	Proficiency (N)	Mean	SD	Mean	SD	(paired)	
DDL	H (27)	31.94	13.71	83.17	13.07	16.45	0.000*
	L (23)	26.63	9.83	77.41	10.52	18.42	0.000*
TRA	H (25)	26.00	9.49	68.35	29.04	6.94	0.000*
	L (23)	19.57	10.71	66.35	25.51	8.59	0.000*
Total	H (52)	29.09	12.14	76.05	23.23	13.92	0.000*
	L (46)	23.1	10.78	71.88	20.09	16.09	0.000*

* p<0.05

Table 5 Scores on Comparison of Word Usage

Word Usage		Pre-test		Post-test		T	Sig.
Group	Proficiency (N)	Mean	SD	Mean	SD	(paired)	
DDL	H (27)	11.73	21.09	56.85	20.76	12.53	0.000*
	L (23)	9.06	12.54	58.04	21.47	9.25	0.000*
TRA	H (25)	30.2	28.29	62.6	29.44	4.67	0.000*
	L (23)	16.67	16.35	53.7	22.73	7.13	0.000*
Total	H (52)	20.61	26.27	59.62	25.22	10.03	0.000*
	L (46)	12.86	14.92	55.87	21.97	11.39	0.000*

*p<0.05

Table 6 Scores on Headline Features

Headline Features		Pre-test		Post-test		T	Sig.
Group	Proficiency (N)	Mean	SD	Mean	SD	(paired)	
DDL	H (27)	29.07	3.15	91.41	5.21	53.42	0.000*
	L (23)	28.09	2.95	89.57	5.76	40.44	0.000*
TRA	H (25)	30	3.16	86.08	19.44	14.48	0.000*
	L (23)	30.6	2.21	81.87	19.52	12.38	0.000*
Total	H (52)	29.52	3.16	88.85	14.10	29.87	0.000*
	L (46)	29.35	2.88	85.72	14.75	24.4	0.000*

*p<0.05

From tables 4-6, we can see that students in both groups at both proficiency levels improved significantly on all of the three instructional focuses after instruction. To answer the question of whether DDL works better for students at a particular proficiency level, we need to look at the gain scores for each proficiency group on each instructional focus as presented in the table below.

Table 7 Gain Scores by Proficiency, Group, and Instructional Focus

Proficiency		High			Low			T	Sig.
Focus	Group	Mean	SD	N	Mean	SD	N	(independent)	
Grammar	DDL	51.23	16.18	27	50.78	13.22	23	0.10	0.916
	TRA	42.46	30.64	25	46.74	26.22	23	5.18	0.607
	Total	47.01	24.39	52	48.76	20.64	46	0.38	0.705
Word Usage	DDL	45.12	18.72	27	48.99	25.41	23	0.62	0.54
	TRA	32	34.84	25	37.03	24.91	23	0.57	0.571
	Total	38.81	28.17	52	43.01	25.60	46	0.77	0.445
Headline Features	DDL	61.59	7.53	27	61.47	7.29	23	0.05	0.957
	TRA	55.76	19.19	25	51.26	19.86	23	0.8	0.429
	Total	58.79	14.53	52	56.37	15.67	46	0.79	0.430

*P<0.05

The figures in Table7 show that for all of the instructional focuses, proficiency level made no significant difference in the students' gain scores.　As to the effectiveness of the DDL approach, the raw figures seemed to favor high proficiency students for two of the three learning focuses. Students in the high proficiency group seemed to improve more in learning of grammar and headline features.　On the other hand, the conventional approach seemed to favor the low proficiency group in learning of grammar and word usage.　However, since none of these differences in scores reached a level of significance, no claim can be made about the differential effects of either approach on any of the instructional focuses.

CONCLUSION

Data-driven learning has been frequently cited as an effective approach in language teaching, and there have been discussions about the factors that influence the effectiveness of the approach. In light of such discussions, the present study aims to determine whether the DDL approach is effective for teaching of different linguistic aspects and whether learner proficiency influences the effectiveness of the approach.　Three instructional focuses were investigated in the study (grammar, comparison of word usage, and special features of a text type) to see if the DDL approach was effective for all of these focuses.　In addition, in an attempt to answer the question of whether the DDL approach can work with students at all proficiency levels, the participants' proficiency level was taken into consideration to see whether the approach had differential effects on students at different proficiency levels.　Results of the study indicate that DDL is an effective instructional approach for teaching of word usage and features of a text type, and that language proficiency does

not seem to affect the effectiveness of the approach.

Based on results of this study, data-driven learning can be employed in real instructional contexts as an effective approach to illustrate differences in usage between similar words and special features of a text type.   In addition to learning the language aspects in focus, students are also placed in a better position to benefit from exposure to authentic language materials of the particular text type in concern (such as authentic English news for this study) and to develop inductive reasoning ability which can be immensely helpful for their future learning.   Results of the present study can only support inferences made on the limited amount of evidence gathered in the study. To make more solid and general claims about the effects of data-driven learning, further studies can be designed to take more factors into consideration, such as difficulty of the grammatical features in focus, a larger number of words in comparison, and diverse features of a text type.

As to the question of whether data-driven learning works better with students at higher or lower proficiency, the present study does not find any evidence to give a definite answer.   Future studies with larger and more heterogeneous groups of participants might be more effective in answering the question.

REFERENCES

Ball, C. N. (1996).   Tutorial notes: Concordances and corpora.   Retrieved March 24, 2003 from http://www.georgetown.edu/cball/corpora/tutorial.html.

Dyck, G. N. (1999).   Concordancing for English teachers.   Paper presented at TESL Manitoba, Canada.   Retrieved March 21, 2003 from http://home.cc.umanitoba.ca/~gdyck/conc.html

Gross, D. (1991).   *A practical handbook of learning teaching*.   London: Cassell.

Hadley, G. (2002).   Sensing the winds of change: An introduction to data-driven learning.   *RELC Journal*, 33(2), 99-124.

Ilse, W. R. (1991).   Concordancing in vocational training.   *ELR Journal*, 4, 103-113.

Johns, T. (1991).   From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning.   *ELR Journal*, 4, 27-45.

Kettemann, B. (1995).   On the use of concordancing in ELT.   *TELL & CALL*, 4, 4-15.

Lee, C. Y., & Liou, H. C. (2003).   A study of using web concordancing for English vocabulary learning in a Taiwanese high school context.   *English Teaching and Learning*, 27(3), 35-56.

Shaffer, C. (1989).   A comparison of inductive and deductive approaches to teaching foreign languages.   *Modern Language Journal*, 73, 395-403.

Someya, Y. (2000).   Online business letter corpus KWIC concordancer and an experiment in data-driven learning/writing.   Paper presented at the 3rd Association for Business Communication International Conference, Dashisha University, Kyoto, Japan.   Retrieved July 10, 2003 from http://www.kamakuranet.ne.jp/~someya/DDW_Report.html

St. John, E. (2001).   A case for using a parallel corpus and concordancer for beginners of a foreign language.   *Language Learning & Technology*, 5(3), 185-203.

Tribble, C. (1997).   Improvising corpora for ELT: Quick and dirty ways of developing corpora for language teaching.   In B. Lewandowska-Tomasczczyk & J. Melia (Eds.), *Proceedings of the First International Conference on Practical Applications in Language Corpora* (pp. 106-117).   Retrieved March 24, 2003 from http://web.bham.ac.uk/johnstf/palc.htm

Wang, L. Y. (2002).   *Effects of inductive and deductive approach on EFL learning collocation patterns by using concordancers*.   MA thesis, Yun Lin University of Technology.

APPENDIX I

Segment of DDL Instructional Materials (Grammar—subjunctive verbs—*demand*)

Observe the following sentences and find rules about how these verbs (*demand, insist, admit, etc.*) are used in sentences.

1. MSF <u>demands</u> that Putin help release volunteers.
2. A gay group <u>demands</u> that God be left out of EU Constitution.
3. When affirmative action was challenged in 1998-99 students stood up and <u>demanded</u> that minorities be given equal opportunities.
4. All the Opposition members <u>demanded</u> that a House committee be constituted to probe into the matter and suggest ways how to protect the state's interest.
5. More than 30 developing countries <u>demanded</u> that workers be granted the "right" to have their families join them.
6. The Russian foreign ministry has <u>demanded</u> that the USA take measures to save the living quarters of Baghdad where the Russian embassy is located from air strikes.
7. "It's unfair that the [Commission] has <u>demanded</u> that the ballots be separated," Wang later said.
…

APPENDIX II

Segment of DDL Instructional Materials (Headline Features)

Observe the following news headlines and lead sentences and write down what you find about the conventions of English news headlines.

BBC trio to face questions about scientist's suicide

BBC director general Greg Dyke, chairman Gavyn Davies and head of news Richard Sambrook are expected to face legal questioning following the suicide of Ministry of Defense scientist David Kelly.

Boeing to add up to 500 jobs

The Boeing Co. will increase its Puget Sound-area military business by up to 500 jobs over the next few months, the company's defense and space boss said yesterday.

Air safety boss to quit

The head of the aviation safety watchdog will leave within weeks as the authority moves to a new structure and the government searches for his replacement.

Indian MPs to discuss AIDS

A forum of Indian MPs from several political parties has announced that a national conference to plan strategies for the fight against AIDS and HIV will be held this weekend.

State to study rates of cancer, birth defects

State health officials, responding to residents' concerns about industrial solvents polluting some 470 homes or business, will study birth outcomes and cancer rates in the area to determine whether they reflect a broader public health problem.

CDC: Southerners, blacks more likely to die of stroke

The first country-by-country atlas of U.S. stroke deaths confirmed in graphic detail Thursday that Southerners and blacks are more likely to die from strokes than other Americans.

Syria, Iran warned by Bush

President Bush on Monday issued a stern warning to Syria and Iran, telling them to stop protecting and aiding terrorists or face the consequences.

Britain <u>accused</u> of distorting UN weapons report

British Prime Minister Tony Blair's Iraq crisis deepened on Thursday as ministers were accused of distorting the findings of the chief UN weapons inspector to support Britain's claims about former president Saddam Hussein's weapons program.

Two US soldiers <u>killed</u> as helicopter crashes in Iraq

In section World News Two soldiers were killed when a US Army reconnaissance helicopter crashed into a river in Iraq, as unknown attackers killed a top police officer and a member of former Iraqi president Saddam Hussein's Baath party in separate incidents.

Man <u>killed</u> in anti-Chavez rally

One person was killed on Monday in protests demanding a recall vote to remove President Hugo Chavez, as electoral authorities postponed announcing whether a recall referendum will be held until midday yesterday.