Evaluating "I Think-I-Can-I-Think-I-Can"

Stephen A. Templin
Timothy C. Guile
Takanobu Okuma
Meio University

PAC3
at
JALT
2001

Conference
Proceedings

MENU
Text Version
Help & FAQ

International
Conference
Centre

Kitakyushu
JAPAN

November
22–25, 2001

It is claimed that people with high self-efficacy outperform people with low self-efficacy. This study researches the reliability and validity of a self-efficacy questionnaire and English test. It also asks whether or not a self-efficacy course leads to higher self-efficacy and higher English achievement. Two hundred and ninety-three Japanese university students participated in this research. The questionnaire and English test proved reliable at $p < .05$. A t-test was used to determine if there was any significant difference before the instruction began at the beginning of the semester and after the instruction was completed at the end of the semester. Students' self-efficacy and English test scores showed a statistically significant gain at the end of the semester.

自己効力感の高い人は自己効力感の低い人に比べ、課題遂行力が高いと言われている。本研究では自己効力感アンケート用紙と英語テストの信頼性及び妥当性を調査する。さらに自己効力感の向上を目指した課程が、より高い自己効力感やより高度な英語力の向上に寄与するか否かを検証する。本調査では293名の日本人大学生を被験者とした。アンケート用紙と英語テストは、95％の有意水準（p<.05）で信頼性が認められる。　当該学期の指導（自己効力感の向上を目指す）開始前とその指導後に統計的に有意な差が生じたか否かを確認するためt-検定を行った。　その学期

末における学生の自己効力感及び英語テストの得点は統計的に有意な差が生じたことを表している。

I-Think-I-Can-I-Think-I-Can

There is a story (Piper, 1930) about some toys that are trying to get over a mountain to be with the girls and boys on the other side. The toys ask a few big train engines if they can take them over the mountain. But the big train engines do not want to go over the mountain. Finally, the toys meet a small engine that agrees to take them over the mountain. As the little engine huffs and puffs up the mountain, she says, "I-think-I-can-I-think-I-can." Eventually, she gets over the hill and takes the toys to the eager girls and boys.

Self-efficacy is a person's belief in how well he/she can accomplish a task or group of tasks (Bandura, 1997; Templin, Shiroku, & Taira, 1999). Learners who believe they can learn a language are more likely to learn a language than learners who believe they cannot learn a language.

It is often asked, how does self-efficacy differ from self-concept, self-confidence, and self-esteem (Coopersmith, 1967; Shavelson, Hubner, & Stanton, 1976; Yule, Yanz, & Tsuda, 1985)? Self-efficacy research differs from the other self-phenomena research in five areas: (1) judgement of capabilities, (2) multiple dimensions, (3) contexts, (4) mastery-criterion, and (5) pre-task measurements (Zimmerman, 1995).

The first difference is that other self-phenomena researchers usually judge how people feel about themselves (i.e., personal qualities) whereas self-efficacy researchers judge how well people believe they can accomplish a task or group of tasks. Second, although self-phenomena researchers may not pay attention to task dimensions, self-efficacy researchers do. For example, self-efficacy researchers understand that learners may strongly believe they can read English but may not believe they can speak English—or a learner who believes she can speak English fluently may lack a belief in her ability to speak English using accurate syntax.

Third, other self-efficacy researchers tend to ignore information about how various contexts affect people's beliefs in their capabilities, but self-efficacy researchers understand that, for example, although introducing oneself at a party of friends and introducing oneself in a meeting of strangers are similar acts, the context itself can affect how well a person believes he can make a self-introduction.

Fourth, other self-phenomena researchers compare how people feel about themselves compared with how other people feel about themselves. In contrast, self-efficacy researchers look at how well people believe they can perform a task/tasks based on some criterion, usually measured by numerical values. Just because a person feels better about her English ability than others around

her feel about their own ability does not automatically mean that she has a strong belief in her own ability to perform an English task or group of tasks. Self-efficacy researchers are interested in measuring her performance and comparing it to the strength of her belief in her ability.

Fifth, other self-phenomena researchers measure the self-phenomenon before their subjects perform the task, after they perform the task, or the subjects may not perform the task at all. Self-efficacy researchers measure learners' beliefs in their capability to perform a task before they perform the task, and then they measure how well the person actually performed the task—this process is necessary in order to predict learners' performance on a task or group of tasks.

Predicting Performance

Although other self-phenomena researchers have failed to predict performance, self-efficacy researchers have successfully predicted that people with high self-efficacy exert more attention, effort, persistence, and strategies than those with lower self-efficacy, and when those with low self-efficacy fail, they often blame their failure on everything except their own shortcomings (Bandura, 1997). Also, people with high self-efficacy set more challenging goals for themselves than people with low self-efficacy. Because challenging goals lead to increased performance (Griffee, 1997a; Griffee & Templin, 1998;

Locke & Latham, 1990), people with high self-efficacy outperform people with low self-efficacy.

Too much self-efficacy can get a person killed in dangerous activities such as rock climbing, but a lack of self-efficacy in less dangerous activities can lead to a lifetime of regret: "educational opportunities forsaken, valued careers not pursued, interpersonal relationships not cultivated, risks not taken, and failures to exercise a stronger hand in shaping one's life course" (Bandura, 1997, p. 71).

Self-Efficacy Course

Psychology researchers have shown that first language (L1) learners with high self-efficacy in speaking (Zimmerman & Martinez-Pons, 1990), listening (Schunk & Rice, 1984), writing (Zimmerman & Bandura, 1994), and reading (Schunk & Rice, 1993) outperform L1 learners with low self-efficacy. These researchers were also able to raise students' L1 performance by raising their self-efficacy. It seems odd that similar studies in L2 acquisition are scarce; thus, research such as our study is needed to fill the gap.

In an intensive summer English class, Templin, Shiroku, and Taira (1999) piloted a self-efficacy syllabus to raise second language (L2) learners' English ability via raising their self-efficacy. This led to the creation of a self-efficacy course for freshman Japanese students studying English as a foreign language (Templin &

Guile, 2000). Similarly, we created a self-efficacy curriculum in our English I course.

We intend to measure our students' self-efficacy improvement and evaluate their achievement in relation to the goals and objectives of English I. In order to measure our students' self-efficacy and achievement in the course, we want to create a reliable and valid English test and self-efficacy questionnaire.

The Study
Research Statement

We guided our research by asking ourselves the following questions: What is the reliability and validity of the self-efficacy questionnaire? What is the reliability and validity of the English I test? If the self-efficacy questionnaire and English I test are reliable and valid, have our students' self-efficacy and English ability significantly increased by the end of the English I course?

Method
Participants

The 293 students who participated were mostly 18 year-old freshmen (about half male and half female) enrolled in English I, a required course. Although the enrollment for this course was 345 students, we dropped 52 students from the study who were not present on both days the pre- and posttests and questionnaire

were administered. The students came from three departments: International Cultural Studies, Business and Information Systems, and Tourism. The classes were one and a half hours long and convened twice a week for fifteen weeks. English proficiency scores were not available for the students.

Instruments

We created two instruments: an English I test (contact authors for a copy) and a self-efficacy questionnaire (contact authors for a copy). For both instruments, we looked at available theories describing our constructs (Griffee, 1997b) and examined similar instruments created by other researchers (see Templin, 1999, for a discussion of self-efficacy instruments). We felt our students would be able to understand English I test items written in English, but we did not feel they could understand the self-efficacy questionnaire items written in English, so we decided to write the English test in English and the self-efficacy questionnaire in easy-to-understand Japanese. We chose a seven-point Likert scale for the questionnaire and brainstormed which items to include in the questionnaire from the English I goals and objectives.

Next, we wrote numerous items for each instrument and held frequent meetings to make revisions. Then we piloted the instruments. After piloting, we analyzed the English I test using item analysis. We scrutinized

items with low point biserials, eliminating eight items that did not discriminate well between students who had achieved the course objectives, and students who had not. We also re-worded one item to improve clarity. Then we modified the test further by giving 60 percent of the points possible to the four skills section (which we wanted to emphasize the most) and 40 percent to the other areas (e.g., syntax). For the piloted self-efficacy questionnaire, we correlated the average of each item with the overall average and eliminated items with low correlations and re-worded one item.

We also created a content validity judgement scale for the English I test and the self-efficacy questionnaire (see example in Brown, 1996, pp. 235-236), rated each item, and compared our ratings. At this point, we felt satisfied with the English I test and the self-efficacy questionnaire and decided not to make any more changes.

Reliability of the Instruments
The reliability of the self-efficacy pre-questionnaire and post-questionnaire was calculated using the Pearson correlation and Cronbach alpha reported in Tables 1 and 2.

Table 1: *Pearson Correlation (One-Tailed) and Cronbach Alpha of Self-Efficacy Pre-Questionnaire*

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Q1												
Q2	.61*											
Q3	.53*	.46*										
Q4	.47*	.40*	.64*									
Q5	.48*	.43*	.48*	.56*								
Q6	.41*	.39*	.38*	.53*	.57*							
Q7	.44*	.36*	.50*	.72*	.55*	.58*						
Q8	.47*	.36*	.62*	.56*	.53*	.47*	.75*					
Q9	.52*	.52*	.41*	.45*	.58*	.49*	.47*	.48*				
Q10	.42*	.39*	.40*	.45	.54*	.53*	.50*	.48*	.54*			
Q11	.41*	.39*	.34*	.38*	.36*	.38*	.36*	.35*	.42*	.38*		
Q12	.46*	.37*	.38*	.35*	.34*	.34*	.38*	.44*	.43*	.39*	.50*	
Q13	.39*	.40*	.37*	.42*	.41*	.42*	.38*	.39*	.41*	.42*	.37*	.38*

* Correlation is significant at the .05 level (one-tailed).

Cronbach alpha = .91

Table 2: *Pearson Correlation (One-Tailed) and Cronbach Alpha of Self-Efficacy Post-Questionnaire*

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Q1												
Q2	.45*											
Q3	.48*	.38*										
Q4	.46*	.37*	.67*									
Q5	.42*	.31*	.44*	.44*								
Q6	.46*	.28*	.47*	.43*	.57*							
Q7	.47*	.30*	.55*	.63*	.42*	.42*						
Q8	.39*	.26*	.65*	.59*	.38*	.43*	.69*					
Q9	.29*	.34*	.39*	.33*	.50*	.43*	.37*	.40*				
Q10	.40*	.44*	.39*	.39	.38*	.37*	.44*	.35*	.33*			
Q11	.33*	.27*	.28*	.28*	.38*	.42*	.19*	.26*	.25*	.36*		
Q12	.33*	.31*	.33*	.33*	.44*	.44*	.32*	.30*	.33*	.39*	.41*	
Q13	.31*	.24*	.33*	.35*	.33*	.43*	.33*	.32*	.32*	.37*	.48*	.42*

* Correlation is significant at the .05 level (one-tailed).

Cronbach alpha = .89

All questions in both tests exhibited significant correlations at the .05 level (one-tailed). The Cronbach alpha for the pre-questionnaire was .91 and the post-questionnaire was .89. In other words, the questionnaire appears reliable.

The English pretest and posttest reliability was calculated using the K-R20 shown in Table 3.

Table 3: *Descriptive Statistics of English Pretest and Posttest (A and B Versions)*

Statistics	Pretest A	Posttest A	Pretest B	Posttest B
N	293.00	293.00	293.00	293.00
k	36.00	36.00	36.00	36.00
M	18.71	27.55	17.88	27.33
Low	2.00	14.00	3.00	17.00
High	34.00	36.00	30.00	35.00
SD	5.35	4.53	4.91	4.31
K-R20	.77	.75	.74	.74

Pretest version A, Posttest version A, Pretest version B, and Posttest version B showed .77, .75,74, and .74 reliability. These results exhibit moderate reliability.

Because of the subjectivity possible in grading the writing and speaking portions of the English test, intrarater reliability had to be calculated. Each of the four instructors graded the writing and speaking portions of the English tests twice, the second grading occurring about one week after the first grading (contact authors for speaking and writing scoring guides). The speaking interviews were recorded and played back only once during the second grading. The first grading and second grading were compared using the Cronbach alpha (see tables 4 and 5).

Table 4: *Descriptive Statistics and Intrarater Reliability for First Scoring of English Test Writing and Speaking Sections*

Statistics	Writing Pretest	Writing Posttest	Speaking Pretest	Speaking Posttest
N	293.00	293.00	293.00	293.00
k	3.00	3.00	5.00	5.00
M	.95	2.66	1.01	3.95
Low	.00	1.00	0.00	0.00
High	3.00	3.00	5.00	5.00
SD	1.11	.63	1.50	1.44
Cronbach Alpha	.96	.91	.92	.95

Table 5: *Descriptive Statistics and Intrarater Reliability for Second Scoring of English Test Writing and Speaking Sections*

Statistics	Writing Pretest	Writing Posttest	Speaking Pretest	Speaking Posttest
N	293.00	293.00	293.00	293.00
k	3.00	3.00	5.00	5.00
M	.94	2.61	1.01	3.91
Low	.00	1.00	.00	.00
High	3.00	3.00	5.00	5.00
SD	1.08	.66	1.50	1.44
Cronbach Alpha	.96	.91	.92	.95

The reliability for the writing portions of the pretest and posttest was .96 and .91, respectively. The speaking pretest and posttest were .92 and .95. The intrarater reliability for the writing and speaking portions of the English test was high.

Statistical Analysis

To analyze the data, descriptive statistics were calculated for the self-efficacy questionnaires and English tests (see tables 3, 6, and 7).

Table 6: *Overall Descriptive Statistics of Self-Efficacy Pre-questionnaire and Post-questionnaire*

Statistics	Pre-questionnaire	Post-questionnaire
N	293.00	293.00
k	13.00	13.00
M	2.90	3.58
Low	1.00	1.00
High	7.00	7.00
SD	1.18	1.14
Cronbach alpha	.91	.89

Table 7: *Descriptive Statistics for Self-Efficacy Pre-Questionnaire and Post-Questionnaire Items**

Questions	N	Low	High	Mean	Std. Deviation
Q1	312	1.00	7.00	2.89	1.08
Q1*	312	1.00	7.00	3.27	1.02
Q2	312	1.00	7.00	3.26	1.14
Q2*	312	1.00	7.00	4.01	1.06
Q3	312	1.00	7.00	3.00	1.03
Q3*	312	1.00	7.00	3.59	1.00
Q4	312	1.00	7.00	2.60	1.28
Q4*	312	1.00	7.00	3.17	1.08
Q5	312	1.00	7.00	2.82	1.09
Q5*	312	1.00	7.00	3.56	1.07
Q6	312	1.00	7.00	2.85	1.22
Q6*	312	1.00	7.00	3.67	1.18
Q7	312	1.00	7.00	2.21	1.17
Q7*	312	1.00	7.00	3.10	1.21
Q8	312	1.00	7.00	2.69	1.23
Q8*	312	1.00	7.00	3.58	1.32
Q9	312	1.00	7.00	3.30	1.22
Q9*	312	1.00	7.00	4.05	1.16
Q10	312	1.00	7.00	3.21	1.15
Q10*	312	1.00	7.00	3.56	1.11
Q11	312	1.00	7.00	3.00	1.28
Q11*	312	1.00	7.00	3.72	1.35
Q12	312	1.00	7.00	2.96	1.21
Q12*	312	1.00	7.00	3.79	1.21
Q13	312	1.00	7.00	3.31	1.27
Q13*	312	1.00	7.00	3.54	1.16

*Pre-Questionnaire questions have no asterisk, and post-questionnaire questions have an asterisk.

Differences were measured by a paired t-test, with an alpha level of .05.

Results

The differences between the pre-questionnaire and post-questionnaire scores were significant at $p < .05$ (see tables 8 and 9).

Table 8: *Matched T-Test of Pre-Questionnaire and Post-Questionnaire*

Statistic	Pre-Questionnaire and Post-Questionnaire
Mean	-0.60
Standard Deviation	1.09
Standard Error of Mean	6.60
Lower Confidence Interval of Difference (.05)	-0.73
Upper Confidence Interval of Difference (.05	-0.47
df	276.00
t	-0.91*

*p<.05

Table 9: *Matched T-Test of Pretest and Posttest*

Statistic	Pretest and Posttest
Mean	-10.23
Standard Deviation	6.54
Standard Error of Mean	0.38
Lower Confidence Interval of Difference (.05)	-10.98
Upper Confidence Interval of Difference (.05)	-9.48
df	291.00
t	-26.73*

*p<.05

Also, the differences between the English pretest and posttest scores were significant at p < .05.

Discussion

The self-efficacy questionnaire and English test were reliable and valid. Also, students' self-efficacy and English ability increased over the period of instruction. The results are in agreement with self-efficacy research that shows students' self-efficacy and achievement can be raised through teaching. Future research should examine the correlation and casual relationship between self-efficacy and L2 achievement.

References

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.

Brown, J.D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.

Griffee, D.T. (1997a). Using goals and feedback to improve student performance on vocabulary homework. *The Language Teacher, 21* (7), 19-25.

Griffee, D.T. (1997b). Validating a questionnaire on confidence in speaking English as a foreign language. *JALT Journal, 19* (2), 177-197.

Griffee, D.T., & Templin, S.A. (1998). Goal-setting affects task performance. In Visgatis, B. (Eds.) *On JALT97: Trends and transitions* (p. 21-26). Tokyo: The Japan Association of Language Teaching.

Piper, W. (1930). *The little engine that could*. London: Platt & Munk.

Schunk, D.H., & Rice, J.M. (1984). Strategy self-verbalization during remedial listening comprehension instruction. *Journal of Experimental Education, 53,* 49-54.

Schunk, D.H., & Rice, J.M. (1993). Strategy fading and progress feedback: Effects on self-efficacy and comprehension among students receiving remedial reading services. *The Journal of Special Education, 27,* 257-276.

Templin, S.A. (1999). The relationship between self-efficacy and language learners' grades. *JALT Journal, 21* (1), 112-124.

Templin, S.A., & Guile, T.C. (2000). *Dream Quest.* Ashiya, Hyogo, Japan: Seido Language Institute.

Templin, S.A., Shiroku, S., & Taira, K. (1999). Self-efficacy syllabus. *The Language Teacher, 23* (4), 9-12.

Zimmerman, B.J. (1995). Self-efficacy and educational development. In Bandura, A. (Ed.),*Self-efficacy in changing societies* (pp. 202-231). Cambridge: Cambridge University Press.

Zimmerman, B.J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal, 31,* 845-862.

Zimmerman, B.J. & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology, 82,* 51-59.