Positive Washback from Thai University Entrance Examinations

AMPORN JIANRATTANAPONG
WSC, Thailand

Bio Data:
Amporn Jianrattanapong is now pursuing an M-Ed (MTEIL) at SIU, her thesis is on the topic "The factors leading to success in learning English at leading high schools in Bangkok". She is now a test prep tutor at WSC.

Abstract
The paper presents weak points and certain strong washback effects from Thai university entrance examinations supporting hints to overcome some shortcomings from writing assessment and encourages the practices and classroom activities that enable graduates to have more competence in English communication particularly in writing skill.

Introduction

It is quite rare to find Thai graduates from local universities competent in English writing in spite of their over 10 year period of English language learning. One likely factor is that writing practice is rarely found in schools and universities (except English major graduates). The purpose of English learning in school and universities is to enable students to communicate in English, including writing, thus why do teachers not get their students to practice writing? One possible cause is washback from university entrance examinations which do not include direct writing tests. This essay looks at how positive washback can be generated by using direct writing tests along with possible ways of implementation.

In the Entrance Examination of 2008, there are three parts (100 items) including:

Part One: Language Use and Usage which consists of 1. Oral Expression, 2. Error Identification

Part Two: Writing which consists of 1.Sentence Level, 2. Paragraph Level (all are to choose an answer to complete each blank space.)

Part Three: Reading which consists of 1. Vocabulary (choose the word that best completes each blank in the passage) 2. Reading Passages (to read the passage and choose the best answer to each question that follows, total 4 passages)

All are multiple choices.

Positive Washback

Washback is the effect or the influence that a test has on teaching and learning. According to Gates (1995) there are two angles to look at washback which are strong-weak, and positive-negative. He states that "if washback is strong, students and teachers will tend to alter their classroom behaviors in order to achieve good marks in the test." (p. 101) Gates (1995) also mentions that monopoly is a factor that affects washback. That is "the less competitors an exam has, the stronger its washback." (p. 102) If we look into the entrance examinations at Thai universities, they certainly have strong washback due to the monopoly. Thus, it has formed the classroom behaviors, and we should aim it at the needed behaviors or improvement of student performances that we want to encourage. In doing this, we have to consider positive washback which happens when course and test objectives overlap or coincide. To study in a Thai state university, one has to pass the tests of several subject matters that s/he has to further the study there including English. If the entrance exams effectively test students by using direct writing tests, then high

school teachers will aim at practicing writing in order that students will pass the entrance exams. On the contrary, the current exams have negative washback, i.e. there is no item in the exams that students have to write. As a result, high school teachers do not get students to practice writing. At the same time, knowing that a practice of writing does not help them to pass the exams, students themselves are not willing to do it thinking that it wastes their time. To promote positive washback, Browns (2005, p. 245) lists a number of suggestions some of which that are applicable to Thai context and related to this topic are quoted as follows:

i)   Design the test on sound theoretical principles (Bailey 1996)

ii)   Use direct testing (Hughes 1989, Wall 1996)

iii)   Use variety of exam formats

The details of these suggestions will be mentioned further.

Writing Assessment

To promote positive washback, characteristics of good writing assessment would be a base factor to consider. A number of educational experts have a consensus that the direct method or composition test is the best way of writing assessment. Hughes (2007, p.3) writes "…if we want to know how someone can write, there is absolutely no way we can get a really accurate measure of their ability by means of a multiple choice test. Professional testers have expended great effort, and not a little money, in attempts to do it, but they have always failed." Another similar opinion of Harris (1969, p.69) is "It would seem obvious that the most direct way of measuring students' writing ability would be to have them write." According to Hughes (2007, p.83) to test writing ability directly, there are 3 parts of problems:

- Writing tasks have to be proper representative of the tasks that students are able to perform.

- The task should elicit valid samples of writing.

- It can be scored validly and reliably.

Hughes gives a sample of a test, which the Cambridge Certificates in Communicative Skills in English (CCSE) examiners choose for one of their tests in 2000 as follows:

An advertisement looking for volunteers to summer camps for children

Task 1 Write an application letter

Task 2 Fill in an application form

Task 3 Write a postcard to a friend (similar to a cloze passage recommended by Bailey, 1998)

Task 4 Write a note (apology)

Hughes (2007) believes that wide sampling can measure a person's writing ability more accurately and meaningfully. He writes "But if the result is going to be very important to candidates (…) then certainly more than one sample is necessary if serious justices are not to be perpetrated." (p.89) This concept is to share the risk of test-takers as they may not do well in some kinds of tasks but are better at others. Additional characteristic that Hughes emphasizes is to test only writing ability and nothing else (creative, imaginative, or even intelligent) (p.90). Concerning valid and reliable scoring, Hughes suggests that we should give no choices of tasks. However, Hamp-Lyons 1990, as cited in O'Malley & Pierce (1996) argue that research results are mixed on whether students write better with single or with multiple prompts. In

general classroom assessment, it would be better for each student to have an opportunity to choose to write the topic in which they are interested and have background knowledge. Nonetheless, in the national exam it is too serious and risky to score on the performances which tend to be various on topics, processes and genre, etc. if there are choices of topics. O'Malley & Pierce add that the performance of students given multiple prompts may be less than expected because they waste time deliberating over the topic to select when they should be investing effort in planning, writing and editing. (1996, p.140)

Another issue in direct writing test is to create reliability of scoring. One may argue that the assessment of a direct writing test is subjective and it is unsuitable for high stake exams. Reliability can be created by having several trained raters score essays. There is evidence that essays scored by five raters have the approximate scorer reliability of 0.92 while the essays scored by only one reader have only 0.25 reliability rate. (1961 College Board study, Godshalk, Swineford, and Coffman, op. cit., p.48 as cited by Harris, 1969, p.70) Also, to reduce subjectivity in marking, Qi (2005, p.9) gives an example that "the writing is provided through drawings or instructions in Chinese (L1). The Marking scheme then prescribes the key points that should have been covered in a writing script."

Weak Points of Thai University Entrance Examinations

Looking back to the assessment in Thai University Entrance Examinations, writing part, it still lags. That is, only multiple choice test or indirect test is available. It is suitable for a diagnostic or a progress test, but it does not measure writing ability of students. Bailey (1998) mentions her observations which reinforces this believe that

"… when I put a wrong sentence on the board, most of the students can tell me what's wrong with it, but they make the same mistakes in their own writing all the time." (p.84) This is similar to what my friend who was an undergraduate in Canada told me. He said one of his classmates (an Asian girl) got a score of TOEFL as high as 600 up, and it was about 17 years ago when TWE was not required and there were only multiple choices in TOEFL including the writing part. The problem was that she always and obviously wrote and spoke with incomplete grammar. The negative washback from this kind of exams is to encourage students to have recognition but they never develop their performance in writing. In other words, they know complex grammatical features, but still make mistakes in writing and we are able to claim that most of high school students or even graduates cannot write English well.

Concerns in Direct Tests

From the above benefits and possibilities of direct writing test, it leads to the question why it is still not applied in Thai University Entrance Examinations. There may be a few limitations that are obvious. One concern is that it consumes a lot of time and money in scoring and training teachers. In view of education, Hughes (2007) points out "When we compare the cost of the test with the waste of effort and time on the part of teachers and students in activities quite inappropriate to their true learning goals (…), we are likely to decide that we cannot afford not to introduce a test with a powerful beneficial backwash effect." (p.56) From the entrance exam of 2008, it can be seen that grammar is the main substance existing in both Part One: Language Use and Usage: Error Identification and Part Two: Writing. Students spend time concentrating on decontextually use of grammar and that is

rote-learning both in classrooms and tutorial schools. I remember myself trying to remember grammatical rules before taking the entrance exam and when I started the writing subject in the university, my instructors were shocked to find a lot of mistakes in my essays.

To find out how writing is currently taught in high schools, I asked 2 high school students and 2 newly graduates what they did in the writing class. It is found that there are some writing practiced, e.g. writing essays, letters, etc. However, they are not emphasized or done properly. That is, students write essays, but there is no feedback for them to improve their writing skill. Moreover, one of the answers is that "I had to write a letter in an exam, but the content was not so important, I just needed to remember the form of a letter correctly, that's all." There is no doubt that the classroom activities did not achieve the true learning goal to increase writing ability.

Furthermore, there are indirect costs, e.g. losses from people lacking English competency. Thai people often complain that "If I were more competent in English, I would have more progress in my career." Also, one can save much time and energy if s/he can write an email in English without doing it in Thai and ask someone else to translate it into English. Another concern is that one may claim that in writing compositions, students can cover up weaknesses by avoiding problems they find difficult. Such evasion is impossible with well-prepared objective tests. (Harris 1969, p.70) This is one weakness in direct test of writing. Harris (1969) proposes that the ideal practice is undoubtedly to measure writing skill with a combination of the two types of test, and it is recommended that the procedure be followed whenever

conditions permit. (p.71) Cloze passages can be another choice that is better than multiple choice tests.

How to Implement Direct Writing Tests?

Due to large-scale exams, it is quite difficult to implement direct writing test. However, there are some possible ways which can be separated into 3 aspects. To begin with, a process and procedures should be developed. It is necessary to create reliability by scoring each script by at least 2 raters. Then, criteria/ rating scales and the descriptors for each scale need to be set clearly for raters to follow. The six-point scoring system of TOEFL or that of IELTS can be a guideline. After scripts have been scored, there should be a group of administration to inspect the consistency of the scores which may be at variance with one another due to raters' bias, leniency, characteristic and rating condition. According to Taylor and Falvey (2007) as cited by Shaw & Weir (2007) IRT methods should be used for calibrating scores as it is accepted in this field. "Although this has cost implications, test fairness demands it." (p.253)

Another aspect to look at is training, we need to train scorers who have to read and score a certain amount of scripts to have experiences covering the varieties of essays. There are approximate 100,000 test-takers of the national exam annually (The Nation, March 27, 2008), so at least 300-400 scorers are needed. To organize trainings, many methods can be integrated including remote training, eLearning, VDO, etc. Concerning organizations, e.g. the National Institute of Educational Testing Service (NIETS) should be able to organize trainings which prospect raters can attend. It can also be broadcasted for those in remote areas to participate in two ways, e.g. can ask

questions if they do not understand. In addition, eLearning can be provided on a website for rater trainees to learn and test by themselves, and VDO/CD can be an option for training access. The use of VDO is convinced by the cases in the United States and Canada of in 1991 when a study was conducted that addressed both the impact of the original research on the field of assessment, and the impact of the video training package on teachers. Results of interviews with members of the research community and follow-up with trainers using the materials indicated extensive use and impact on practice. Several hundred packages are distributed each year. At the time of the follow-up study in 1991, over 4,000 facilitators had trained more than 10,000 teachers. (http://www.ed.gov/pubs/ triedandtrue/class.html) To keep a good standard, after the training, the trainees have to be certified before starting scoring.

The last aspect to mention is to use technologies. E-rater developed by ETS Technologies can provide one of the two scores as it has been proven highly effective and accurate in scoring more than 800,000 examinee responses to essay questions of GMAT Analytical Writing Assessment. (Princeton, N.J.,Business Wire, 2001) We can use Electronic Script Management (ESM) which is an on-screen marking of candidate responses captured in electronic format through the scanning of paper scripts (Shaw & Weir, 2007, p.300) to facilitate the process for more accuracy and speed. The factor of huge test population has been proven not to hinder the implementation of direct writing test as it has been applied in China, where there are 7.23 million test-takers in 2004 (Qi, 2005, p.9).

Apart from the implementation, to enable high school teachers to teach by getting students to practice writing with peace of mind, there are some suggestions for them that self-assessment and peer-assessment can be an alternative. Students can use a scoring rubric to rate their peers' papers. In doing this, when students learn to evaluate the work of their peers, they are extending their own opportunities to learn how to write. (Cramer, 1982 as cited in O'Malley & Pierce 1996, p.156) By these options, teachers spend scoring time only for important test, e.g. midterm and final exams.

Summary

As long as Thai University Entrance Examinations still have strong effects to the students, positive washback should be promoted so as to encourage the practices and classroom activities that enable graduates to have more competence in English communication particularly in writing. And the positive washback that can be generated is to apply direct tests in the writing part of the exams. There are ways that we can overcome the shortcomings like subjective and large-scale scoring, for instance, to have effective rating scales, to have well trained and certified raters, to use scoring technologies which have proven their reliability by standardized tests like TOEFL, IELTS. Li, 1990; Li et al., 1990 as cited by Qi (2005) confirms that positive washback is effective as she writes "In the early years after the test was introduced, teaching to its content induced intended changes in that teachers began to teach reading and writing." The change of writing assessment of SAT college entrance exam in 2005 into a 20-30-minute essay is another example that shows the development trend in writing assessment (Woolfolk, 2004, p.539). Thus, this mission

is possible and it is a hope for Thai people to become competent in written English. Hence it is worth doing in spite of time and cost.

References

At least 161 cheat in police entrance exams, (2008, March 27). *The Nation.* Retrieved

    December 30, 2008, from

    http://www.nationmultimedia.com/search/adsearch.php?keyword=entranc

    e+exam

Bailey, M. K. 1998. *Learning About Language Assessment.* Boston, MA: Heinle & Heinle

    Publisher.

Brown, J. D. 2005. Language test validity. *Testing in Language Programs.* New York:

    McGraw-Hill.

Gates, S. 1995. Exploiting washbacck from standardized test. *JALT Applied Materials*

    Language Testing in Japan, 101-102.

Hughes, A. (2007). *Testing for Language Teachers.* Cambridge: Cambridge University

    Press.

Harris, D. P. (1969). *Testing English as a second language.* New York: McGraw-Hill.

Northwest Regional *Educational* Laboratory (NWREL), *Classroom assessment video*

    training workshops. Retrieved December 31, 2008, from

    http://www.ed.gov/pubs/triedandtrue/class.html

O'Malley, J. M. & Pierce, L. V. (1996). *Authentic assessment for English language*

    learners. New York: Addison-Wesley Publishing.

Princeton, N. J. (2001). *ETS technologies offers TOEFL writing topics as part of criterion*

    online writing evaluation. Retrieved December 22, 2008, from

    http://www.accessmylibrary.com/coms2/summary_0286-10783954_ITM

Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a

    high-stakes test. *Sage Journal Online.* Retrieved December 22, 2008, from

    http://online.sagepub.com/cgi/searchresults?andorexactfulltext=and&fullte

    xt=stakeholders%27+conflicting+aims+undermine&src=hw

Shaw, S. D. & Weir, C. J. (2007). *Examining writing*. Cambridge: Cambridge

    University Press.

Woolfolk, A. (2004). *Educational psychology*. Boston, MA: Allyn & Bacon.