



IC-07

Washback Effects on Implementing English Proficiency Tests as a Graduation Requirement at Thailand National Sports University (TNSU)

Ketsunee Taweekaew

Ph.D., Thailand National Sports University Bangkok Campus, Thailand

Email: ketsunee2012@gmail.com

Abstract

The influence that a test has on teaching and learning is commonly known as washback. This study seeks to explore the washback effect of Thailand National Sports University English Proficiency Test (TEPT), a high-stake compulsory university graduation test. The focusing area was on students as the main participants in this research, the issue of how university students of Thailand National Sports University (TNSU) perceived on the TEPT and their own self efficacy of the intended results in relation to the test was examined.

The findings suggested that the students' perceptions played a major role in mediating the washback effect of the TEPT, in which the test shaped their goals and consequently stimulated their use of language learning strategies when preparing for the test. However, with regard to the concern on the graduation requirements and passing the test, though it was the initial issue to implement this policy was ignored due to the failure to enforce the rule to use the language assessment (TEPT) as a gatekeeping device for a graduation.

In sum, it may be concluded that when a policymaker introduces a novel policy by using English proficiency test as a graduation requirement, the implementation should be clear, hence the practitioners would be able to take actions that align with the policy. In particular, to avoid negative unintended consequences, the policy should not be ambiguous as it may result in unexpected policy ineffectiveness, or neglect.

Keywords: washback, washback effect, high-stake test, language assessment

Introduction

Language assessment is becoming one of the key areas of study in the field of educational research due to its major influence on teaching and learning. It is used to obtain evidence and information concerning students' knowledge, skills and abilities in the tested language. Performance in carrying out language tasks in the test will be used as evidence of their mastery of the language. Unlike specific tests that measure expertise in a specific skill, language assessment is something that most people have to undergo regardless of their background. The results obtained from a language assessment or test, especially a high-stake language test, will to a certain extent affect someone's life as such tests are used to make many crucial decisions, for example determining who should have access to important opportunities in areas including employment, further education and so on. In other words, the results are often used as a gatekeeping device to filter and exclude those who are deemed to be unqualified.

This phenomenon thereafter has contributed to an intended top-down policy of educational changes in Thailand National Sports University (TNSU) starting the last academic year 2020 since then. According to the provision of the Ministry of Education, it announced that all educational institutes need to provide the resources to serve the need to enable the standard teaching and learning for the students. Using tests as a lever of educational change is commonly believed to have an impact on teaching, learning, education systems, and society, a phenomenon described as 'backwash' in general education or as 'washback' in language instruction, and the impact is often taken for granted. In theory, the assessment must cover the students' language skills in using the language. However, a question is raised whether such an assessment affects the teachers and students in their teaching and learning. So far, washback is associated primarily with 'high-stakes' tests, that is, tests used for making important decisions that affect different sectors, for example, determining who receives admission into further education, a requirement for graduation or employment opportunities (Chapman and Snyder, 2000). Chen (2019), a researcher on washback usually refers to those of high-stakes tests. It is commonly agreed that the higher the test consequence is, the more serious it will influence the students as well as the teachers in terms of their attitude, motivation, and teaching-learning process. The research on washback of alternative tests is also seen as an effort to investigate whether a new or revised form of test brings about changes in teaching and learning. Because the change of the test type might change how the teachers perceive their role in assessment. In general, washback refers to the effect of testing on teaching and learning (Cheng, 2015). The role of washback is required to estimate and ensure that the test gave appropriate influence to teaching and learning in the classroom decision or in the educational system. The influence of the test is asserted as a part of classroom practice when it could direct what will happen and occur in the classroom. It includes the teaching aspects, such as the teaching technique,



teaching contents, teaching material, teaching strategies, activity or time arrangement and the ways of assessing (Sukyadi & Mardiani, 2011). It was thought that the mechanisms of washback were uniquely connected to “the test” and the context. However, as pointed out by Alderson and Wall (2013), although it is widely known that there is a relationship between testing and teaching and learning practices, the complexity of the washback concept makes it difficult to identify how tests affect teaching and learning practices without taking into account other mediators or variables that may or may not contribute to the influence.

To date, as noted above, research on the washback effect, particularly in foreign language learning, has tended to focus on teachers’ perspectives as they are viewed as among the most important stakeholders in assessment, whereas there is little concern of washback studies on learners (Pan, 2014). Therefore, as test-takers, learners’ points of view need to be taken into consideration as they are the ones directly affected by these tests. It is unfair to learners and has proven rather problematic to employ a “one-size-fits-all” approach (Chu, 2009) in making judgments concerning the washback effect of high-stakes tests.

Hughes (2003), as one of the earliest scholars to give due attention to washback, developed a preliminary model of washback in which the “participants” are one of the three main components, together with “processes” and “products”. Based on Hughes’ tri-partite model (2003), positioning learners as the principal stakeholders in the washback phenomenon, it is important to take into consideration the variety of effects that tests can have upon learners and also other variables. For example, Tsai and Tsou (2009) found that learners’ negative opinions of a test will lead to a decrease in motivation to learn. It would be interesting to explore the relationship between learners’ perceptions of the test and how these perceptions influence their course of action, for instance in terms of their language learning strategies when preparing for the test. Washback on learners is associated with how students react towards a test, for example by amplifying or reducing their effort to learn because of their perceptions of the test. The current study thus seeks to explore how learners’ perceptions determine the different extent of washback they experience by looking at the language learning strategies they adopt when preparing for a high-stake English language test. Concerns about the lack of empirical studies aimed at determining the relationship between how the two most important stakeholders of testing, namely the teachers and learners, are affected have been raised. Thus, this study was prompted by the scarcity of washback studies focusing predominantly on the learners, but at the same time considering the perspective of the teachers as interlinked.

The target examination chosen for investigation in this study was the TEPT, a high-stake language test of Thailand National Sports University. As TNSU is newly established and thereafter dependent on sports business for all levels, English is deemed to be an indispensable tool for success in a requirement for graduation. There has been a growing concern for national standards of educational achievement in a competitive global world, then it has been reflected in tertiary institutions, where the problem of graduates’ English proficiency has been addressed. The concern emerges as a consequence of mounting dissatisfaction with graduates’ linguistic ability, both internally from the teaching staff of educational institutions, and externally from the public, especially graduates’ employers. Innovation has been proposed in order to ensure higher standards of English proficiency among students graduating from TNSU. Therefore, the innovation of passing English proficiency tests as a graduation requirement was expected to have an impact on what and how teachers taught and what and how students learned throughout their study curriculum. It was a deliberate attempt to influence the teaching of English in a positive direction, by moving away from the traditional, teacher-dominant classroom towards a more active learning context, where students had sufficient opportunity to practice the skills ready for the English proficiency tests. Accordingly, TNSU just used an evaluation and testing to be the engine for implementing educational policy. In this study, it was aimed to contribute to more discussion about washback effects of English Proficiency Test in TNSU for the requirement for graduation of the academic year 2020/2021 of students, by investigating the context of teachers’ descriptions of how their practice has or has not been affected by this standardized testing. Therefore, this study should determine to what degree does the TNSU English Proficiency Test exert a washback on students’ English skills.

Purposes

1) To investigate the phenomena of the washback effect in the light of using English proficiency tests as a gatekeeping device.

This subject deserves further investigation on the washback on non-English major students, because all of them have been asked to pass a certain level of English proficiency tests as one of their degree requirements, even though they do not major in English and have less English input throughout their curriculum.

The TNSU tests were developed in-house by a particular group of English teachers of the universities. In fact, as indicated by the President of TNSU, the students are required to achieve above the B1 level defined in the Common European Framework of Reference for Languages (CEFR, 2001).



2) To listen to how the main participants within the TNSU context reacted to the implementation of English proficiency tests as a graduation requirement.

3) To explore the nature and scope of the washback effect on aspects of institutional policies, teachers' and students' perceptions, and teachers' behaviors, within the context of the English graduation regulation.

Research Methodology

The main purpose of this research was twofold. First, a reliable and valid scale was designed to collect data from university students on their attitudes towards the backwash effect of TEPT and to check if these attitudes are positive or negative. In order to obtain more detailed information, a three-case study was employed as a case-oriented analysis, looking more closely into a particular case. The present researcher used the quantitative data that had been collected to corroborate and support the evidence with the qualitative data which was most useful for understanding the rationale or theory underlying relationships, as explained in the following chart.



Figure 1 Sequential Mixed Method Design
Source: (Creswell, 2003)

The above figure indicates that to obtain a more comprehensive information on the subject of this study, a qualitative method was used, in which the case study strategy was chosen as another choice. The original source materials were consulted and analyzed to seek out and extract evidence from original sources from teachers of English in TNSU.

1. Data Collection

1.1 Questionnaire

In accordance with the research objectives, two groups of participants were involved in this study: 1) English language instructors, and 2) senior students attending the TEPT. In order to examine the role of language proficiency level in mediating the washback of the Proficiency of English Test on Washback Effects (PEWE), the students were categorized into three groups based on their faculty members: Faculty of Education (ED1), Liberal Arts (LB2) and Sports and Health Science (SC3). Data of this study were collected from the final-year TNSU students with the contribution of students from different 17 campuses across the country. The instruction prescribed in the questionnaire mentioning that all participants' identities would be protected, and they would remain anonymous in the reporting of the research. The questionnaires of this study were launched online in March 2021 when the senior students returned to TNSU to sit for the TEPT. The convenience sampling method was chosen to collect the maximum amount of data from voluntary participants from different campuses in TNSU. The returned questionnaires were used to screen PEWE's scale & factor reliability to confirm the factor structure of the instrument. To be able to collect data from participants, it was first aimed to develop an attitude scale to examine the washback effect of TEPT considering a preliminary version of the washback scale was developed. Next, the new scale was piloted in order to do its exploratory and confirmatory factor analyses. Third, the scale's internal reliability was measured and proven to be a valid and reliable tool to collect data including university students' views towards the washback effect of the university exam on their attitudes.

The questionnaire consisted of three parts. The first section dealt with the demographic characteristics of the students, namely gender, faculty, number of English courses in their entire curriculum, and tutorial courses for TEPT. Section 2 comprised 50 items within 8 categories addressing: (1) teaching methods, (2) teaching resources, (3) class activities, (4) learning and assessment in relation to preparing students for the TEPT, (5) English practice outside class, (6) benefits of learning English, (7) effects of English scores during the course of study, and (8) perceptions toward TNSU English Proficiency Test. A five-point Likert-type scale was adopted, which addressed aspects of classroom teaching and learning of English and the TEPT, where 1 = lowest, 2 = low, 3 = moderate, 4 = high, and 5 = highest. In the last part, the researcher used an open-ended questionnaire to permit more opinion and feedback from the participants.



In the processes, firstly, the objectives, the participant group and the required time were defined. Before the item writing process, the researcher tried to gather some preliminary ideas about the possible backwash effects of TEPT on students' academic lives in university by interviewing a small group of students who came to sit for the first TEPT in academic year 2020 at Bangkok Campus. The effects of the test commonly mentioned by these students were determined and later used to gain some insights about students' feelings towards the TEPT in item writing. Secondly, the focus of the scale was finalized and the related items were written. Thirdly, the draft was checked and revised by 3 experts in language and research field. Next, the scale was piloted for exploratory & confirmatory factor analyses. Finally, in March 2021, online questionnaires were launched to target participants across different campuses for those who sat for TEPT. The returned questionnaires were used to screen PEWE' scale & factor reliability to confirm the factor structure of the instrument.

Table 1 Questionnaire Items

| Factors | Questionnaire Items |
|---|--|
| Factor 1 Language learning strategies | |
| 1. Teaching activities (of teachers) | 1. Informing the objectives of each lesson 2. Explaining specific language items, such as words or sentences 3. Going through textbook 4. Organising pair work or group discussions 5. Integrating mixed teaching methods |
| 2. Teaching resources (of teachers) | 6. Textbooks 7. Supplementary materials 8. Multimedia 9. Newspaper/ Magazines 10. Language laboratory |
| 3. Participation on learning activities (of students) | 11. Doing pair work 12. Doing group discussion 13. Expressing ideas |
| 4. In-class - teaching methods | 14. Listening practice 15. Speaking practice 16. Reading practice 17. Writing practice 18. Grammar practice 19. Learning vocabulary 20. Playing English games |
| 5. Out-of-class - English learning | 21. Talking to teachers in English 22. Talk to peers in English 23. Talk to general public in English 24. Watch TV/ movie in English 25. Listen to radio in English program 26. Read newspaper/printed media in English |
| Factor 2 Self-efficacy of students | |
| 1. Benefits of learning English | 27. Understand basic English structure 28. To get a job 29. To further study 30. To communicate with foreigners 31. To understand movie program in English 32. To socialize with more confidence 33. To get more opportunities in the future 34. To pass examinations 35. To satisfy school requirements |



Table 1 Questionnaire Items (Continue)

| Factors | Questionnaire Items |
|---|---|
| 2. Impacts of scores in English in the course study | 37. Effects of English scores 38. Self-image 39. Motivation on learning 40. Anxiety and tension 41. Relationship between teacher and student 42. Job opportunity in the future |
| Factor 3 Students' perceptions of the test (TEPT) | 43. Preference on TEPT 44. Improved learning performance if more practice on mock exam papers 45. TEPT forcing to study harder 46. Test scores used as indicator to evaluate teaching and learning in curriculum 47. Test scores used as indicator to evaluate learner's skills in application of lesson learned 48. Success on doing TEPT resulting from entire English courses in the curriculum 49. TEPT corresponding to English courses of TNSU 50. Current English courses of TNSU enough for future use |

1.2 Semi-structured interview

The interview aimed at exploring the language instructors' perceptions of the impact of the assessment of TEPT. All interviews were audio recorded with the participants' permission and subsequently transcribed for analysis. In an interview, the topics of discussion related to the curriculum and the assessment system such as the course outlines, and course assessment were analyzed to investigate the degree of compatibility between the assessment system and the curriculum in terms of content, objectives and format. In doing so, the aim was to determine the kind of washback the assessment system was meant to produce and how it was expected to be implemented (Watanabe, 2004).

Within an interview session, the teachers of English were therefore asked how they viewed the fact that the TNSU English Proficiency Test (TEPT) was so present in their discussions throughout the academic year. In this research, it mainly focused on the specific washback effects of TEPT that emerged during the course study.

2. Data Analysis

In the quantitative data analysis, the researcher applied the statistical program to analyze the data. In the procedures of data screening, a computer program for statistics was used in order to decode the raw materials within this process. In the first part, descriptive statistics for sample groups explained 1) the general information of the samples, and 2) mean and standard deviation. In the second part, inferential statistics were applied to test the correlation of variables of the hypothesized model and those of empirical data. Data from the questionnaires were input and analysed using descriptive and inferential statistics. The analyses involved frequency distributions and item analysis for descriptive statistics and correlation analyses, as well as regressions for inferential statistics. Data of the study were collected via google forms. All the responses were encoded and prepared for statistical analyses by the researcher. The data was thoroughly screened by different techniques as detailed:

2.1 The correlation analysis was used to describe the strength and direction of the linear relationship between the variables. In this study, a Pearson product-moment correlation coefficient was designed for continuous variables of dependent variables.

2.2 Besides, an assumption of Normal Distribution was also investigated. In technical terms, the Assumption of Normality claims that the sampling distribution of the mean is normal or that the distribution of means across samples is normal. Otherwise, if the errors are not truly random, then too this assumption might not be valid. If the error distribution is not normal and the assumption of normality is made, then this could lead



to an incorrect statistical analysis and thus erroneous conclusions. Normal distribution is screened so as to confirm that the independent variables are not highly correlated, (if not), where $r=0.9$ or greater, (Tabachnick, B. G. & Fidell, L. S., 2001), otherwise, the effects of such independents on the outcome variable cannot be separated. In other words, one of the predictor variables can be nearly perfectly predicted by one of the other predictor variables. In practice, the researcher decided to include independent variables with an inter-correlation not greater than .70.

2.3 In this study, multiple regression was employed to explore the relationship between the dependent variable (washback effects on TEPT) and a number of independent variables in order to investigate how well the set of those predictors were able to predict or contribute to the particular outcome.

For the interview transcripts, content analysis was carried out based on the themes listed in the framework of this study, namely the teachers' perceptions of test importance, curriculum, administrative support and policy and their language teaching strategies when preparing for the TEPT. The qualitative data from the interview were then used to complement, support and explain the quantitative findings. The semi-structured interview was divided into two parts: 1) General information of the correspondent, and 2) Opinions to TNSU English Proficiency Test (TEPT).

Results

The perceptions of the TNSU English Proficiency Test (TEPT) in this study were identified through the criteria of interpreting scores, which varied according to the mean score for each variable.

Criteria of Interpreting Scores

- 4.21- 5.00 = highest level
- 3.41- 4.20 = high level
- 2.61- 3.40 = moderate level
- 1.81- 2.60 = low level
- 1.00- 1.80 = lowest level

Table 2 Sum of mean squares between groups

| | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|-----|-------------|------|-------------|
| Between Groups | .049 | 2 | .024 | .044 | .957 |
| Within Groups | 138.791 | 253 | .549 | | |
| Total | 138.840 | 255 | | | |

This table gives both between groups and within groups sums of squares, degree of freedom and significance value (Sig.). In the column Sig., the value was at .957, which was greater than .05, then there was no difference in the mean score of perceptions on TEPT among the students from the three faculty members.

Besides, in order to ascertain whether the difference between groups is statistically different and not likely to have occurred by chance, one way to assess the importance of the findings is the calculation of the 'effect size', known as 'strength of association'. Though, there are a number of different effect size statistics, the most common of which is eta squared, (Cohen, J. (1988). Eta squared represents the proportion of variance of the dependent variable that is explained by the independent variables. Values for eta squared can range from 0 to 1. To interpret the strength of eta squared values, the following guidelines can be used, (Cohen, 1988).

- .01 = small effect;
- .06 = moderate effect; and
- .14 = large effect

To calculate the effect size from eta squared, the formula is:

$$\text{Eta squared} = \frac{\text{Sum of squares between-groups}}{\text{Total sum of squares}}$$

therefore,

$$\begin{aligned} \text{Eta squared} &= \frac{.049}{138.840} \\ &= 0.0004 \end{aligned}$$

The resulting eta squared value was 0.0004, which in Cohen's (1988) terms would be considered a small effect size. Hence, it indicated that the group difference among the three faculties were not statistically different in terms of effect size.



Table 3 Perceptions on TNSU English Test (TOPIN)

n=256

| Variables | Mean | S.D. | Level |
|--|-------------|-------------|----------|
| Dependent variable | | | |
| Perceptions on TNSU English Test (TOPIN) | 3.35 | .73 | moderate |
| Independent variables | | | |
| Teaching activities (TACT) | 3.63 | .73 | high |
| Teaching resources (SOURC) | 3.44 | .79 | high |
| Learning activities (LACT) | 3.67 | .80 | high |
| In-class activities (INCL) | 3.52 | .77 | high |
| Out-class activities (OUTCL) | 2.91 | 1.02 | moderate |
| Benefits of English (Use) | 3.44 | .76 | high |
| Effects of scores (EFFECT) | 3.09 | .66 | moderate |

To conclude, it was indicated that, regarding the level of the perception of the variables in this study, it was found that the performance of the dependent variable (TOPIN), and the other independent variables was mostly at moderate to high level. Though, it revealed that the factor of outside-class activities (OUTCL) indicated the lowest mean score. This finding implied that the students had limited opportunities to communicate in English outside a classroom at a low level.

On the other hand, the factors of learning and teaching activities contained rather high mean scores at 3.67 (S.D.=.80) and 3.63 (S.D.=.73) respectively, which implied that the students showed their preferences in the current in-class activities to some extent.

Table 4 Assessment of the suitability of the data for factor analysis

| Variables | Correlation matrix | KMO | Bartlett's Test-Sig. | Component matrix | % of Variance | Deleted |
|--|--------------------|------|----------------------|------------------|---------------|---------|
| Dependent variables | | | | | | |
| Perceptions on TNSU English Test (TOPIN) | .561-.691 | .905 | .000 | .752-.868 | 66.633 | FORC |
| Independent variables | | | | | | |
| Teaching activities (TACT) | .542-.695 | .835 | .000 | .819-.868 | 69.839 | |
| Teaching resources (SOURC) | .540-.691 | .812 | .000 | .794-.880 | 70.259 | LAB |
| Learning activities (LACT) | .540-.736 | .682 | .000 | .788-.893 | 73.652 | |
| In-class activities (INCL) | .603-.748 | .886 | .000 | .839-.862 | 71.771 | GAME |
| Outside-class activities (OUTCL) | .545-.665 | .854 | .000 | .753-.910 | 76.000 | TVE |
| Benefits of English (USE) | .523-.686 | .932 | .000 | .651-.893 | 69.020 | TPASS |
| Effects of scores (EFFECT) | .529-.611 | .863 | .000 | .750-.871 | 70.327 | GRAD |

In the process of data generation, in considering the number of factors to retain for the analysis, a number of criteria were used to clean the data, as follows:

First, factor analysis involved inspecting the correlation matrix of .3 and above, which indicated the sufficiency of strength of the relationship among the variables or items (Pallant, 2005). Cohen (1998) suggests how to interpret the value of the Pearson correlation (r) value as follows:

- r=.10 to .29 or r=-.10 to -.29 = small
- r=.30 to .49 or r=-.30 to -.49 = medium
- r=.50 to 1.0 or r=-.50 to -1.0 = large

Accordingly, in this study the correlation matrix below .5, which might not be appropriate for factor analysis were deleted as detailed: (1) Language lab [LAB], (2) English game [GAME], (3) TV English program [TVE], (4) Test pass [TPASS], (5) Graduate [GRAD], and (7) Force study hard [FORC]. Hence, the correlation matrix with coefficients greater than .5 and above with large value of correlation (r), showing the strength of the relationship among the items were retained for further analysis.

Hence, it was evident that a number of items were excluded from the scale, as they displayed the values opposed to the criteria, as mentioned beforehand. In sum, the remaining variables included in the factorability indicated that Bartlett's Test of Sphericity was statistically significant at .05 or smaller, coupled with the calculating the Kaiser-Meyer-Olkin Measure of sampling adequacy (KMO) of roughly .6, as suggested as the



minimum value for a good factor analysis. Additionally, due to Kaiser’s criterion, the components (factors) to be extracted all had an eigenvalue of 1 or more, indicating that these factors were retained for further investigation in the multiple regression. Hence, it can be concluded that the sample size in this study was adequate for the factorability of the data.

Table 5 Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | R Square Change | Change Statistics | | | Sig. F Change |
|-------|-------------------|----------|-------------------|----------------------------|-----------------|-------------------|-----|-----|---------------|
| | | | | | | F Change | df1 | df2 | |
| 1 | .584 ^a | .341 | .339 | .59999 | .341 | 131.681 | 1 | 254 | .000 |
| 2 | .691 ^b | .478 | .474 | .53520 | .137 | 66.213 | 1 | 253 | .000 |
| 3 | .701 ^c | .491 | .485 | .52952 | .013 | 6.459 | 1 | 252 | .012 |
| 4 | .729 ^d | .531 | .523 | .50940 | .040 | 21.305 | 1 | 251 | .000 |
| 5 | .760 ^e | .578 | .569 | .48418 | .047 | 27.827 | 1 | 250 | .000 |
| 6 | .815 ^f | .664 | .655 | .43316 | .086 | 63.366 | 1 | 249 | .000 |
| 7 | .830 ^g | .689 | .681 | .41694 | .026 | 20.749 | 1 | 248 | .000 |

- a. Predictors: (Constant), TACT
- b. Predictors: (Constant), TACT, SOURC
- c. Predictors: (Constant), TACT, SOURC, LACT
- d. Predictors: (Constant), TACT, SOURC, LACT, INCL
- e. Predictors: (Constant), TACT, SOURC, LACT, INCL, OUTCL
- f. Predictors: (Constant), TACT, SOURC, LACT, INCL, OUTCL, USE
- g. Predictors: (Constant), TACT, SOURC, LACT, INCL, OUTCL, USE, EFFECT
- h. Dependent Variable: TOPIN

In the Table labelled Model Summary, it showed that after the variables (EFFECT, OUTCL, TACT, LACT, USE, SOURC, INCL) entered, the variance in the dependent variable (TOPIN) was explained by the model at 68 per cent ($r^2=.689$). Accordingly, it indicated that the independent variables resulted in the perceptions of students towards the TEPT at this degree. In addition, in the column marked R Square Change in the output presented above, the value of R Square Change was .026, indicated that if the variables of TACT, SOURC, LACT, INCL, OUTCL, USE were removed, the model could explain only at 26 per cent. These findings were confirmed by the Sig. F Change value at $p<.005$.

Discussion

This research was intended: 1) To investigate the phenomena of the washback effect in the light of using English proficiency tests as a gatekeeping device, 2) To listen to voice how the main participants within the TNSU context reacted to the implementation of English proficiency tests as a graduation requirement, and 3) To explore the nature and scope of the washback effect on aspects of institutional policies, teachers’ and students’ perceptions, and teachers’ behaviors, within the context of the English graduation regulation. More specifically, the study aims at offering teachers and students an opportunity to voice their concerns and opinions about the graduation regulation and to inform the university authorities of teachers’ and students’ opinions.

Firstly, in line with the intentions of the curriculum and assessment developers, the TEPT was found to provide the students with an authentic and meaningful context for learning and using English. The students considered that English was useful to their future career and in their daily life. However, the test itself did not help thereby increase their intrinsic motivation or force to study English harder (where the variable FORC ($r=.561-.691$) was removed from the model) among students of all proficiency levels. This finding was in contrast with the results of the study by Pan (2014), which reported positive washback on intrinsic motivation of higher-proficiency students. On the other hand, the study by Buyukkeles (2016), which indicated “no significant washback” on students’ intrinsic motivation “regardless of their language proficiency”. This phenomenon can be perhaps explained by the difference in assessment design of the test itself. The positive feedback investigated the impact of one-off standardized exit tests. Whereas the TEPT did not enforce the prerequisite condition of passing in relation to the graduation requirement. As prescribed in the TNSU’s announcement of language assessment exit examination, it required that the students needed to pass at least B1 level of the CEFR standard. However, as a matter of fact, it was revealed that some of the students (approx. 10 per cent) did not pass the set standard (Registration office of TNSU Bangkok Campus, 2022), still they were permitted to graduate. Possibly, due to the unstable policy implementation, it was more likely not to increase students’ intrinsic motivation for learning



English. This situation responded to the results of the study, in terms of the low coefficient value of the independent variables of TEST PASS (TPASS, $r=.523-.686$) and GRADUATE (GRAD, $r=.529-.611$) were removed from the model. Therefore, it was assumed that the TEPT was unable to be used as a gatekeeping device for graduation due to the failure to enforce the policy implementation to use the English proficiency test as a graduation requirement.

Secondly, with regards to the washback of the TEPT on teaching, findings of the study suggested that the teachers had a good understanding of the rationale and requirements of the TEPT and were committed to make the most of the assessment system to improve their students' language learning. The dependent variable of Perceptions on TNSU English Test (TOPIN), comprised 1) preference on TEPT, 2) improved learning performance if more practice on mock exam papers, 3) test scores used as indicator to evaluate teaching and learning in curriculum, 4) test scores used as indicator to evaluate learner's skills in application of lesson learned, 5) success on doing TEPT resulting from entire English courses in the curriculum, 6) TEPT corresponding to English courses of TNSU, 7) current English courses of TNSU enough for future use. Whereas the item of TEPT forcing to study harder (FORC) was initially removed due to the low coefficient value, since it was decided to keep only the correlation matrix with coefficients greater than .5 and above with large value of correlation (r), so as to retain the strength of the relationship among the items for a step-further in the analysis. Explicitly, the study showed that the perception of the performance of the dependent variable [TOPIN] ($\bar{X}=3.35$, $S.D.=.73$), and the other independent variables were mostly at moderate to high level ($\bar{X}=2.91-3.67$, $S.D.=.66-1.02$). This implied that the senior students who sat for the TEPT perceived on their moderate preference in term of their attitudes towards the backwash effect of TEPT. The factors of Learning activities (LACT, $\bar{X}=3.67$, $S.D.=.80$), indicated the highest mean scores, followed by Teaching activities (TACT), $\bar{X}=3.63$, $S.D.=.73$ respectively, which implied that the students showed their preference in the current teaching and learning activities of TNSU. The result implied that the students showed their preferences in the current in-class activities to some extent. Though, it revealed that the factor of outside-class activities (OUTCL) indicated the lowest mean score ($\bar{X}=2.91$, $S.D.=1.02$). This finding implied that the students had limited opportunities to communicate in English outside a classroom at a low level, though it differed in term of the regions located in different regions.

Thirdly, when considering the top policy makers of TNSU, the TEPT was found to help the teachers monitor their students' progress and make necessary adjustments in their instruction. The teachers' provision of constructive feedback on the students' performance with guidance on how to enhance their language skills as well as involvement of their students in self-assessment can be identified as the main positive impact of the TEPT on teaching. This finding is in line with the results of the study by Yu (2010), which reported washback on teachers' assessment practices. However, the teachers reported during the interviews that other than having to focus on academic content, the test result of TEPT did not cause any major changes in their teaching methodology which they generally described as being communicative in nature. This, they reasoned, was because the current methods they were using already promote student-centred learning and communicative language teaching, and therefore were congruent with the TEPT. This corroborates the findings of the study by Alderson and Hamp-Lyons (2016) to the effect that "the amount and type of washback will vary according to [...] the extent to which the test is counter to current practice". The English teachers all reported that they just taught as indicated in the course outline, and did not make much changes due to the TEPT. On the other hand, the findings of the study indicated that the TEPT increased the students' confidence in using English and encouraged them to take more responsibility for their own learning, as it was intended by the curriculum and assessment developers. The students reported engaging more in independent out-of-class learning activities (OUTCL), such as talking with friends in English, watching English videos, listening to English songs, reading English books and playing games in English. This finding is in agreement with some earlier studies, which reported the impact of internal assessment on students' out-of-class learning activities (Buyukkeles, 2016; Pan, 2014; Yu, 2010).

To sum up, the study showed that after the group of independent variables (EFFECT, OUTCL, TACT, LACT, USE, SOURC, INCL) entered in the model of regression, it indicated that the students who sat for the TEPT recognized the importance of test at 68 per cent, and if those variables were removed from the model, the value of R Square Change was at .026, indicating that without the mentioned independent variables, the students just perceived their preference on TEPT at only 26 per cent. On the contrary, it could be assumed that the TEPT was not judged as a graduation requirement. Though it was not officially announced, it was understandable among those undergraduates since the item of 1) 'TEPT forcing to study harder (FORC)', 2) to pass examinations (TEST PASS -TPASS), and 3) to satisfy school requirements (GRADUATE-GRAD) were previously removed from the model due to the low coefficient value of less than .5, which was not appropriate for factor analysis of the dependent variable (TOPIN).



Conclusion

The adoption of standardized English Proficiency Test as a tool for assessing students' English competence for graduation is becoming more and more common in higher education. Generally, testing is commonly believed to have an effect on teaching and learning. A case in point would be a teacher's adaptation of teaching materials and methodology to assist students preparing for an exam. When a test has an impact on the educational process involving the teacher and the students, it is generally termed as "backwash" (Hughes, 2003). Depending on whether that impact proves harmful or beneficial, a distinction is drawn between negative and positive washback. A test is said to have negative washback when a restricted definition of linguistic competence is taken as a basis for the development of its content or design, which results in teaching and learning contexts becoming narrow in scope. Alternatively, positive washback occurs when a test stimulates good teaching and learning practices.

This paper focuses on university undergraduate students and uses data from a questionnaire survey to investigate their views of the application of standardized TEPT tests as the graduation benchmark. The results of this study show that standardized TEPT tests are viewed as insufficient to reflect what is learnt and taught in a foreign language classroom and likely to make English instruction become test-driven. The researcher concludes that the adoption of standardized TEPT tests as an assessment tool should be determined on the basis of students' English competence and learning situations, and implemented in the context of overall curriculum planning and needs analysis. Accordingly, in this study it indicated that the teachers of English did not change their teaching methods, on the other hand they just put more emphasis on listening and speaking skills relevant to the TEPT. Basically, the teachers just followed the course outlines, intended to enable the students to a more proficiency in language. Though, partially, the results of the TEPT helped fix the weak points in teaching and learning.

On the other hand, the present study provides some insights into the washback of the internal assessment on the teaching and learning of English in Thailand National Sports University, but it has certain limitations. First, this study investigated the washback of an English language assessment system or TEPT, and its findings were specific to this particular research context and assessment system. Theoretically, there should have been some relationship between learners' perceptions of the test, and how these perceptions influenced their course of action, for instance in terms of language learning strategies when preparing for the test. However, in practice, due to the fact that the policy enforcement to use the results of the test as a graduation requirement was ignored, accordingly the students did not show their self-efficacy in recognizing that the TEPT itself truly affected their academic success, (where 'to pass examination [TPASS]' and 'to graduate [GRAD]' were removed). Though, it is commonly agreed that the higher the test consequence is, the more serious it will influence the students as well as the teachers in terms of their attitude, motivation, and teaching-learning process. On the contrary, in this study it was found that the students did not show their intrinsic motivation to study harder to pass the test, since the item of 'TEPT forcing to study harder (FORC) was initially removed and excluded from the model of an analysis. Therefore, among this uncertain circumstance of the policy implementation on the adoption of standardized English Language Proficiency test or TEPT as a gatekeeping device for graduation, the connections between language policy and practices were corrupted. The TEPT was, accordingly, unable to offer a lens to examine the proficiency level of those undergraduates before their graduation. As a result, TNSU then failed to use the TEPT as a mobility change to uplift the standardized benchmarking in language standard.

Hence, whenever the policy implementation is fully monitored to ensure that the language graduation requirement would be implemented as intended, a more research design involving a larger number of participants is necessary to be conducted for investigation in TNSU in order to have a more insightful understanding of the washback of internal language assessment in the context of the university. Despite its limitations, the present study is one of few studies that have provided empirical evidence concerning the impact of internal language assessment on teaching and learning. As the findings of the study suggested, the washback constituted an area of research deserving further exploration. It would be more interesting to explore the relationship between learners' perceptions of the test and how these perceptions influence their course of action, including their language teaching and learning strategies, curriculum or any other mediators or variables that may or may not contribute to the influence. The future study should find more evidences of washback on learners, and on the association between the learners and their reactions towards a test, for example by amplifying or reducing their effort to learn because of their perceptions of the test. Then the study thus should seek to explore how learners' perceptions determine the different extent of washback they experience by looking at the language learning strategies they adopt when preparing for a high-stake English language test. Concerns about the lack of empirical studies aimed at determining the relationship between how the two most important stakeholders of testing, namely the teachers and learners, are affected have been raised. Thus, this study was prompted by the scarcity of washback studies focusing predominantly on the learners, but at the same time considering the perspective of the teachers as interlinked.



Lastly, the findings of this study could help the TNSU work toward an improvement of the language assessment system, curricular, teaching and learning and, alongside, enhancing the language learning achievement at the very earliest stage of the students in the early stage of their education. At the same time, the policy on language assessment as a prerequisite requirement for graduation should be explicit and not ambiguous in the implementation phrase to ensure higher standards of English proficiency among graduating students, otherwise it would fail on its own merits within its messy system.

Recommendation

Hopefully, the TEPT should emphasize a more continuous approach to assessment that includes greater provision for feedback from not only teachers but students throughout the study years. At the end of each year, however, students are required to sit for the test prior to their upgrade to the higher grade. Therefore, the TEPT should be used to determine placement into the different levels during their undergraduate programmes at the university. In case the students did not pass the required yearly standard of English score, additional courses of English should be included in their study plan, designed to prepare the students for the final TEPT in their final year of study. In this regard, the TEPT would be of more concern about raising student's motivation on language in preparation for studies in the degree programmes. Seen as a whole, the TEPT therefore prepares students for the standardized examination as well as for real world and authentic language related tasks using a combination of formative and summative assessment. All these findings have relevant implications and should therefore lead to recommendations for further training or studies. It is recommended a larger sample size with a longer time frame should be extended to increase the degree of generalization of the study and in order to make the findings more valid and reliable.

References

- Alderson, J. C., & Wall, D. (2013). Does washback exist? *Applied Linguistics*, 14(2), 115-129. Retrieved September 4, 2021, from doi.org/10.1093/applin/14.2.115
- Alderson, J. C., & Hamp-Lyons, L. (2016). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280-297. Retrieved December 2, 2021, from https://doi.org/10.1177/026553229601300304
- Buyukkeles, G. (2016). *The washback effect of a high-stakes exit test on students' motivation in a Turkish pre-university EFL preparatory school*. Master's dissertation, University of Reading.
- Chapman, D. & Snyder, C. W. (2000). Can high stakes national testing improve instruction: Reexamining conventional wisdom. *International Journal of Educational Development*, 20, 457-474.
- Chen, I. W.-L., & Hsieh, J. J.C. (2019). English language in Taiwan: An examination of its use in society and education in schools. *English language education across Greater China*, 70-94.
- Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, 48(4), 436-470. Retrieved November 12, 2021, from https://doi.org/10.1017/S0261444815000233
- Cohen, J. (1988). *Statistical Power analysis for behavioral science*. Hillsdale, NJ: Erlbaum.
- Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (2nd edition)*. Thousand Oaks, CA: Sage.
- Chu, H. Y. (2009). *Stakes, needs and washback: An investigation of the English benchmark policy for graduation and EFL education at two technological universities in Taiwan*. National Taiwan Normal University.
- Hughes, A. (2003). *Testing for language teachers. (2nd edition)*. Cambridge: Cambridge University.
- Pallant, Julie. 2005. *SPSS Survival Manual: A step by step guide to data analyzing using SPSS version 12 (2nd edition)*. Bunkingham: Open University Press.
- Pan, Y.-C. (2014). Learner Washback Variability in Standardized Exit Tests. *TESL-EJ: Teaching English as a Second or Foreign Language*, 18(2), 1-30. Retrieved January 2, 2022, from http://anu.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwnV1LSwMxEB6UXgTxbaQg3hbmya7jXsSkRYPvb14LZvHwiKs2t2KuudSba19qDgLRAlA0lmJl_mmw9AaiiserfgEmcYYh3jiuDVA9YWwzhhOYPtHEJmnK9owgdRPSEHY7bmT9J7bPhsCzXt9JZNBzPGC3Ly8RiQjRdtrabGOnQw1U5Jy0F16eIFJmVgSHJFVCE
- Registration office of TNSU Bangkok Campus. (2021, 20 April). *Announcement of TNSU English Proficiency Test Academic year 2020/2021*.
- Sukyadi, D., Mardiani, R. (2011). The Washback Effect of the English National Examination (ENE) on English Teachers' Classroom Teaching and Students' Learning, *K@ta*, 13(1), 98-113.
- Tabachnick, B. G. & Fidell, L. S. (2001). *Using Multivariate Statistics (4th edition)*. New York: HarperCollins.



- Tsai, Y., & Tsou, C. (2009). A standardized English Language Proficiency test as the graduation benchmark: student perspectives on its application in higher education. *Assessment in Education: Principles, Policy & Practice, 16(3), 319-330*. Retrieved August 25, 2021, from <https://doi.org/10.1080/09695940903319711>
- Watanabe, Y. (2004). "Methodology in washback studies". cited in L. Cheng, Y. Watanabe, & A. Curtis. (Eds). *Washback in language testing: Research contexts and methods*. Lawrence Erlbaum, 19-36.
- Yu, Y. (2010). *The washback effects of school-based assessment on teaching and learning: A case study*. Doctoral Dissertation, University of Hong Kong.