Washback to Classroom Climate: The Case of an IELTS Speaking Preparation Course[1]

Kioumars Razavipour[2], *Fahimeh Hoseini*[3], *& Mahmood Validi*[4]

Received:09/01/2019 Accepted: 03/09/2019

Abstract

In applied linguistics, washback, or backwash, refers to the influence of tests on teaching and learning. Though there is now a large body of research investigating the impact of tests on educational processes and outcomes, the possible effect of language tests on the psychological environment of language classrooms remains underexplored. This study examined the effect of IELTS on the classroom climate of IELTS speaking preparation courses. To isolate the washback of IELTS, the classroom climate of an IELTS speaking preparation course was compared with that of a general conversation course, both taught by the same teacher. Data were collected using classroom observations and a questionnaire. Sixty-seven participants responded to a measure of classroom climate and 4 sessions of each of the noted courses were observed. Data were analyzed, using parametric and nonparametric statistics. It was revealed that IELTS exerted some negative influence on 3 dimensions of classroom climate: student involvement, teacher support, and investigation. In contrast, the washback of IELTS on class cohesiveness, teacher support, investigation, task orientation, as well as on cooperation and equity appeared to be positive. Findings carry implications for language teaching, test preparation, and language classroom management.

***Keywords*:** Test Washback; Classroom Climate; IELTS; Preparation Courses; Speaking

---

[1]Please cite this paper as follows:

Razavipour, K., Hoseini, F., & Validi, M. (2020). Washback to classroom climate: The case of an IELTS speaking preparation course. *Journal of Research in Applied Linguistics*, *11*(1), 21-43.

[2]Corresponding author, Department of English Language and Literature, Faculty of Letters and Humanities, Shahid Chamran University of Ahvaz, Ahvaz, Iran; *razavipur57@gmail.com*

[3]Department of English Language and Literature, Faculty of Letters and Humanities, Shahid Chamran University of Ahvaz, Ahvaz, Iran; *fahimehhosseini69@gmail.com*

[4]Department of English Language and Literature, Faculty of Letters and Humanities, Shahid Chamran University of Ahvaz, Ahvaz, Iran; *validym@yahoo.com*

1. Introduction

High-stakes tests are powerful and important tools, capable of inducing changes in different aspects of educational systems and in societies at large (Hamp-Lyons, 1997). Specifically, language tests are known for their power because of their gate keeping functions for immigration, university admission, and employment (Shohamy, 2014). The power of tests in controlling and directing educational programs has, in particular, been the subject of an extensive body of literature in recent decades (Cheng, Sun, & Ma, 2015).

In applied linguistics, the influence that tests exert on teaching and learning is referred to as washback, or backwash (Hamp-Lyons, 1997). Since the pioneering study of Wall and Alderson (1993) on test washback in Sri Lanka, test washback has become one of the active areas of scholarship in language testing and assessment. As such, various aspects and dimensions of washback including test and language factors (Shohamy, Donitsa-Schmidt, & Ferman, 1996), teacher factors (Alderson & Hamp-Lyons, 1996; Watanabe, 1996, 2004a, 2004b), and test-taker factors (Qin, 2011; Xie & Andrews, 2013) have been studied in the literature on test washback. Nevertheless, an area that seems to have escaped scholars' attention is the effect that language tests exert on the collective classroom environments or classroom climate. Given the centrality of instructional environment in determining learning outcomes (Ghaith, 2003) and its role in affecting student enjoyment and willingness to communicate (Khajavy, MacIntyre, & Barabadi, 2018), this absence of focus on classroom climate in research on test consequences provides a strong imperative to see how tests, especially high-stake language ones, affect classroom climate. Intended to narrow this gap, the present study explored the washback of the Speaking component of IELTS on the classroom climate of the IELTS preparation courses.

2. Literature Review

The existing research suggests that the higher the stakes of a test, the more impacts the test is likely to have on education and society (Hamp-Lyons, 1998; Shohamy et al., 1996). Accordingly, IELTS can possibly have strong washback, given that it is used as a basis for making critical decisions about test-takers (Green, 2006). There is also evidence that washback is a highly complex phenomenon, likely to be confounded by numerous individual and contextual factors (Cheng et al., 2015; Wall & Alderson, 1993; Watanabe, 2004a). With these two premises, it follows that IELTS most likely exerts a strong influence on language teaching and learning, and that contextual and individual factors are at play in determining its washback.

Thus far, several aspects of the washback of IELTS have been studied, including washback to teachers (Naghdipour, 2016), washback to teaching materials

(Saville & Hawkey, 2004), and washback to learners (Green, 2006). In the IELTS preparation courses held in private language schools, Naghdipour (2016) found that, with the exception of the first stage of writing (i.e., composing), teachers ignore other stages of the writing process (e.g., editing, revising, and rewriting), mainly because they think that writing for IELTS is a kind of timed writing where there is no time for editing, revising, and the like. Encouraging students to memorize patterns for writing different parts of the writing tasks for IELTS was another strategy that teachers frequently used in the IELTS preparation courses (Naghdipour, 2016). In the absence of more robust empirical studies controlling for teacher factors, it is difficult to determine whether such observations can be taken as evidence of the washback of IELTS. In fact, one of the main problems of studying test washback is that it is often difficult, if not impossible, to establish causal links (Messick, 1996) between tests and educational processes, as well as outcomes based on descriptive and observational studies lacking a comparative element (Watanabe, 2004a).

To forge such causal links, preparation courses directed at IELTS have been compared with general EAP courses in the literature (Ahmed, 2015; Estaji & Tajeddin, 2012; Green, 2006; Hayes & Read, 2004). Green (2006) discovered considerable differences in the expectations of teachers and students with regard to course content and course outcomes in preparation courses, that is, students' perceptions were found to be affected by teachers and their aims. Using a similar design, Hayes and Read (2004) compared an IELTS preparation course with an EAP one in New Zealand. The findings corroborated those in Green (2006) and many other studies in general education. Hayes and Read (2004) found that the syllabus was remarkably narrowed to maximally reflect the content and format of IELTS. Another study with a comparative element was Ahmed (2016) that found the negative washback of IELTS on the IELTS preparation courses. Ahmed found that the students were mainly seeking to get the required band scores and not to learn the English language. Likewise, Estaji and Tajeddin (2012) reported more pressure and anxiety experienced by the students in the IELTS preparation courses compared to those in non-IELTS courses.

In addition to research in institutes, a few studies have addressed the washback of IELTS in noninstructional settings. Mickan and Motteram (2009) found that, in their preparation for the IELTS exam, the applicants relied on their peer experiences and commercial IELTS preparation materials, with heavy reliance on test-related items and tasks. In like manner, Allen (2016) investigated the washback of IELTS on learners and their learning in noninstructional settings. Allen found that IELTS exerted positive washback on the productive skills of speaking and writing, but not on the receptive skills of reading and listening.

In terms of the trichotomous model of washback (Hughes, 2003), the review of the extant literature indicates that research into the washback of IELTS has focused mainly on participants and processes. Yet, an important dimension of instructional process that has been given short shrift in the literature is classroom environment, or climate, which is crucial to the success of learning and instruction. Perhaps, this lack of attention to the climate of language classes in washback studies is because, to the best of our knowledge, in none of the theoretical models of washback (Bailey, 1996; Hughes, 2003) has classroom climate been postulated as a potential target of washback. Nor was classroom climate mentioned in the fifteen washback hypotheses proposed by Alderson and Wall (1993). Similarly, in Watanabe's (2004) model, though macro and microcontexts are considered to be important in shaping and mediating washback, no explicit mention is made of classroom climate.

In modern educational systems, it is the class, not the individual, which is "the primary unit of instruction" (Olson, 2009, p. 349). The classroom "is the true center of the educational experience, and . . . it is here, through the teaching-and-learning process, that education happens" (Wright, 2005, p. 1). Classroom climate refers to the quality and characteristics of classroom life (Beld et al., 2018). Sink and Spencer (2005) defined classroom climate "as a smaller learning context (in contrast to school) in which students and learners have got a lot of experiences and feelings" (p. 38). Classroom climate has a number of dimensions, including "social, educational, psychological, and physical learning environment" (Dunn & Harris, 1998, p. 100). In particular, the relationships between the teacher and students, as well as the interactions among students are a big influence on students' experience of learning and education (Beld et al., 2018; Wright, 2005). Research also suggests that classroom climate is a powerful determinant of learners' motivation, self-efficacy, and ultimate educational attainment (Beld et al., 2018; Rivkin, Hanushek, & Kain, 2005; Wright, 2005).

In contrast to general education, the study of classroom does not enjoy a long standing research tradition in language education. Nonetheless, every idea about language teaching and learning implies a certain classroom environment (Wright, 2005). For instance, different language teaching methods like communicative language teaching, community language learning, and desuggestopedia favor particular learning environments (see Larsen-Freeman & Anderson, 2013). Likewise, in studies on language learner anxiety (Horwitz, Horwitz, & Cope, 1986) and language classroom management (Wright, 2005), consideration is inevitably given to classroom climate. What is evident is that a tense, insecure, hostile classroom climate does not contribute to successful language learning (Hamre & Pianta, 2005).

Aside from a few studies, the effect of high-stakes testing on classroom climate is underresearched, even in general education (Plank & Condliffe, 2013). In

a survey study of Ohio teachers, Rapp (2002) found that a high percentage of the teachers reported that the climate of their classes was negatively affected by state testing policies. In specific, the surveyed teachers pointed to "losses of autonomy, insight, creativity, and love of learning" (p. 216). Similarly, the promotion of "teacher-centered instructional strategies" in the wake of No Child Left Behind (NCLB) policy was reported in Plank and Condliffe (2013, p. 1152). Plank and Condliffe assessed the classroom climate in terms of three dimensions: emotional support, classroom organization, and instructional support. Classroom organization refers to teacher's management of time, attention, and behavior. Research shows that both emotional support and classroom organization are associated with better academic achievement (Pianta & Hamre, 2009). Plank and Condliffe (2013) argue that though the desirability of aligning curricula to standards remains controversial, there is far less debate about the desirability of optimal classroom climate. "Regardless of what a teacher is teaching, how he or she interacts with students influences multiple measures of students' social and academic development" (p. 1154). Using observational data from classrooms collected in a span of 2 years, Plank and Condliffe compared second grade classes, which were not subject to accountability, with third grade classes, which were subject to accountability testing, during January, when accountability pressure was high, and during May, when accountability pressure subsided. They found that, in contrast to second-grade classes, third-grade classes registered a significantly low level of instructional support. In other words, high-stakes testing was associated with less concept development, less higher-order thinking, less quality feedback, and less valuing of students' viewpoints and language. Likewise, Anagnostopoulos (2006) found that school accountability policies had negative impacts on emotional and instructional support for the students who most needed such support. In particular, Anagnostopoulos found that such testing policies reinforced a discourse of moral boundary distinctions that would, further, worsen the classroom dynamics through generating social exclusion of those who failed to meet the standardized tests' passing criteria.

In applied linguistics, there is a dearth of studies focusing explicitly on how testing policies and practices affect the everyday dynamics of language classes. In their comparison of English courses aiming at preparing students for TOEFL with General English courses in the United States, Alderson and Hamp-Lyons (1996) found that the lack of laughter and innovation on the part of the teacher characterized the TOEFL preparation courses. Expressing such findings in the terms of classroom climate dimensions, Alderson and Hamp-Lyons' findings mean that the TOEFL preparation courses were poor on instructional and emotional support dimensions of classroom environment. Yet, the authors could not attribute such influences to the TOEFL, *per se*, as teachers' beliefs and preferences are a constant mediator of policy

environment and classroom practice (Diamond, 2007). Surveying a small sample of Iranian English teachers, Riazi and Razavipour (2011) found that the teachers complained of a lack of agency because of the pressure they experienced from national and state testing policies. The lack of classroom observation data, however, limits the validity of the findings of this study.

None of the noted studies had as its explicit focus the study of the washback of high-stakes tests on classroom climate. To the best of our knowledge, there remains a gap in the literature regarding how classroom climate changes in response to high-stakes tests. The noted gap is of particular significance not only to language testing and education, but also to general education because of the polemic views on the relation between institutional policy and L2 classrooms. On the one hand are those who endorse a buffering, decoupling view of classrooms (Bidwell, 2001), according to which classes may change merely symbolically to external pressures. On the other hand, other scholars maintain that institutional policies do penetrate classrooms in substantial ways (Coburn, 2004). There is provisional evidence supporting each of the noted views. The current study is a modest contribution to filling this void in the literature in language testing and general education. Specifically, the study sought to investigate whether, and the extent to which, classroom climate was affected by IELTS. In specific, the following questions guided this study:

1. To what extent does IELTS induce washback on the classroom climate of IELTS preparation speaking courses?
2. Which dimensions of classroom climate are affected by the IELTS washback?

3. Methodology

3.1. Setting and Participants

In the design of washback studies, the participants must ideally be selected, rather than sampled (Watanabe, 2004a). Likewise, to be better positioned to trace the presence of washback, including a comparative element in the design of the study, improves internal validity (Watanabe, 2004a). Moreover, washback is a phenomenon that can only be investigated in naturalistic settings, as it is impossible to simulate the effects of high-stakes tests when the stakes are low or nonexistent (Alderson & Wall, 1993; Bailey, 1996). Mindful of these requirements, to study the washback of IELTS on the climate of test preparation courses, we selected two intact classes: A general conversation course and an IELTS speaking preparation course, both of which were taught by the same teacher during the fall semester in 2016. The two courses had to be taught by the same teacher because teacher characteristics, beliefs, and attributions have been found to be important mediating factors in shaping test washback (Wall,

1996; Wall & Alderson, 1993; Watanabe, 1996; 2004). In other words, by choosing two courses, we controlled the teacher variable. This requirement of choosing classes that are taught by the same teachers imposes constraints on the sample size of washback studies.

Of the two courses observed in this study, the non-IELTS class was a credit-bearing conversation course with 15 freshman English major students. The class was held twice a week in a language laboratory. The IELTS course was part of a comprehensive IELTS preparation program, covering all the four language skills in separate classes at the English Language and Literature Department of Shahid Chamran University, based in the Southwest of Iran. There were 21 students in the IELTS speaking course. The teacher was a seasoned assistant professor of TEFL with 20+ years of teaching language skills, both at tertiary and secondary education levels. For the purpose of this study, four sessions (240 min) of each course were observed, at which point data saturation was reached.

To make quantitative analysis possible, in addition to participants in the two observed courses, another sample of 39 participants were invited to fill out the questionnaires. Twenty-two participants were sophomore English major students in the same English department and 17 were students in the IELTS speaking preparation courses offered in two private language institutes. In sum, the study had 75 participants, of which 67 responded to the classroom climate scale. Table 1 provides details with regard to the design of the study:

Table 1. *Design and Participants of the Study*

	Course	*N*	Total	*N* of Absentee Participants at the Day of Administering Questionnaire	*N* of Observed Sessions	Length of Observed Time (min)
Participants Observed	Conversation Course	15	36	0	4 Sessions	240 min
	IELTS Course	21		8	4 Sessions	240 min
Participants Not Observed	Conversation Course	22	39	0		
	IELTS Course	17		0		
Total		75		8	8 Sessions	480 min

3.2. Instruments and Data Analysis

The data were collected using classroom observation and a questionnaire. To capture the pertinent classroom climate dimensions, an observation scheme called a Revised Classroom Climate Assessment Tool (C-CAT), taken from Leff et al. (2011), was used. C-CAT consists of two domains of compliance and classroom responsiveness. Each domain, further, consists of several subdomains of target

behaviors and the subdomains, in turn, contain a number of behavioral indications. The compliance domain consists of (a) noncompliance/disruptive behavior (arguing and defiance), (b) teacher reprimand, (c) transition in classroom teaching, and (d) interruption in teaching. Classroom responsiveness domain includes (a) teacher praise (verbal praise or nonverbal praise), (b) teacher assistance, (c) teacher encouragement, (d) classroom level of interest or enthusiasm, and (e) classroom level of focus and being on-task.

In total, eight class sessions were observed. The observations started from the third week of the semester when the courses settled into their weekly routines. Each session of the conversation course lasted for 75 min, the last 60 min of which were observed in each observation session, as the first 15 min during each session were usually about class businesslike calling the role, greeting the students, and so on. Each session of the IELTS speaking course lasted for 90 min. We observed only two-thirds (1 hour) of each session because the other one-third was spent on marginal issues, which were not the concern of the study. For the purpose of this study, C-CAT was filled out during observation because it is preferable to video-recording for its being embedded in the real contextual events of the classroom. "Live observations may offer greater validity in the sense that they can capture everything that is happening in the classroom, and not just what is visible on a video" (Curby, Johnson, Mashburn, & Carlis, 2016, p. 767). To counter the observer effect on the observed classes, the second researcher would, unobtrusively, sit at the back of the classroom, where her presence was minimally visible.

The second instrument was the What is Happening in this Class (WIHIC) questionnaire, taken from Pickett and Fraser (2002). Using a measure of assessing classroom environment primarily completed by class members "has the advantage of characterizing the setting through the eyes of the actual participants and capturing the data that the observer could miss or consider unimportant" (Pickett & Fraser, 2002, p. 6). WIHIC comprises of 56 Likert-scale items (*Almost Never*, *Seldom*, *Sometimes*, *Often*, and *Almost Always*), tapping seven dimensions of classroom climate: student cohesiveness, teacher support, involvement, investigation, task orientation, cooperation, and equity. The questionnaire was administered to the participants by the second researcher towards the end of the fourth observation session.

To offset the possible effects of differential English proficiency of the respondents, the questionnaire was translated into Persian. To ensure translation equivalence, back translation was done. An expert in the field of applied linguistics, with a good mastery of Persian and experience with quantitative research, edited the translated version and suggested substantial revisions because the translated version was deemed to be too literal a rendering of the English version, which would fail to capture the nuances of classroom environment in the context of the present study.

The internal consistency of WIHIC inspected via Cronbach's alpha ranged from .772 to .896 for the seven dimensions of the classroom climate inventory.

Prior to the data analysis, the requirements of parametric statistics were checked. In particular, the skewness and kurtosis values for each subscale of the questionnaire were inspected. In addition, the internal factor structure of the questionnaire was examined, using Partial Least Squares Structural Equation Modeling (PLS-SEM). Further, boxplots were examined to locate possible extreme scores in each subscale. The results of each of the noted procedures are reported in the next section. To find (possible) differences between the two courses, descriptive statistics and independent samples *t* tests were used. The observation data collected via C-CAT were analyzed, using vhi-square test because such data are of the nominal type dealing with frequencies of different tokens of behavior.

4. Results

Data cleaning is a crucial step in doing quantitative analysis. According to Tabachnick, Fidell, and Ullman (2007), missing values higher than 15% should be properly dealt with, as they are likely to distort the data. In keeping with the above guidelines, the responses of the one participant who had skipped 20 items of the questionnaire were removed from the final pool of data. To check whether the data lend themselves properly to parametric statistical analysis, the skewness and kurtosis values of each questionnaire item were inspected and were all found to be within the allowed range limit of +2 to -2 (Bachman, 2004). Likewise, the skewness and kurtosis values of each of the original subscales of the questionnaire pointed to lack of violation of the normality assumption (see Table 2):

Table 2. *Skewness and Kurtosis Values of Questionnaire Subscales*

	Student Cohesiveness	Teacher Suppose	Involvement	Investigation	Task Orientation	Cooperation	Equity
N Valid	66	66	66	66	66	66	66
Skewness	-.12	.24	.05	-.06	.33	.26	.99
Std. Error	.29	.29	.29	.29	.29	.29	.29
Kurtosis	-.67	-.41	-.62	-.56	.23	-.41	.35
Std. Error	.58	.58	.58	.58	.58	.58	.58

The structural validity of WIHIC was investigated using PLS-SEM. Our choice of PLS-SEM was informed by the fact that the sample size prohibited the use of covariance-based SEM. When it comes to sample size and the distributional requirements of the data, PLS-SEM makes less stringent assumptions (Hair Jr, Hult, Ringle, & Sarstedt, 2016). Figure 1 illustrates the loadings of items on their corresponding factors:

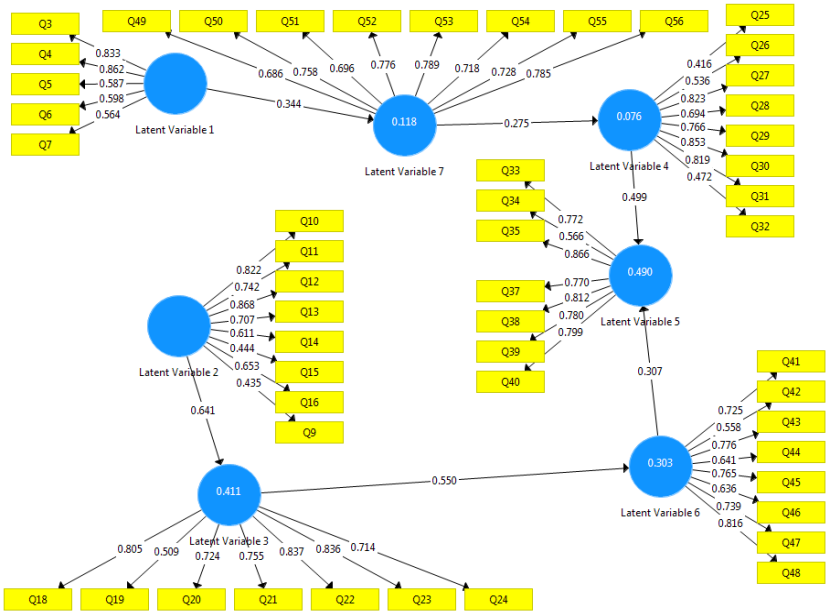

Figure 1. Outer Loadings of Indicators on the Factors in PLS-SEM

As Figure 1 demonstrates, except for items # 2, 8, 17, and 36, which were dropped because of their poor loadings on their corresponding factors, the remaining items all satisfactorily loaded on the intended factors. It is of note that the paths of relationships among the factors were not the concern of this study, and the paths observed in Figure 1 are there only because in the context of PLS-SEM, for the software to produce output, variables cannot be isolated from one another.

Prior to conducting group comparison boxplots, an informative graphic representation of data is highly recommended (Field, 2009; Larson-Hall & Herrington, 2009). The following seven boxplots represent the distributional characteristics of the data on each of the classroom climate dimensions of the questionnaire:

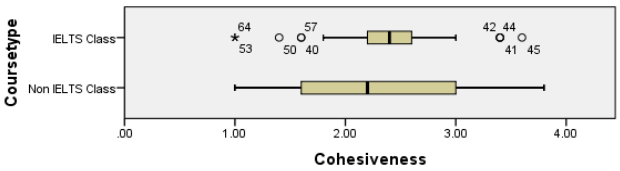

Figure 2. Boxplots of Participants' Scores on Cohesiveness Dimension of Classroom Climate

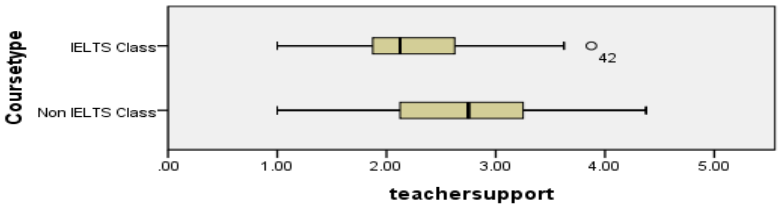

Figure 3. Boxplots of Participants' Scores on Teacher Support Dimension of Classroom Climate

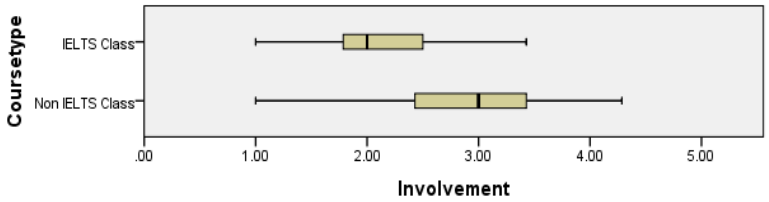

Figure 4. Boxplots of Participants' Scores on Involvement Dimension of Classroom Climate

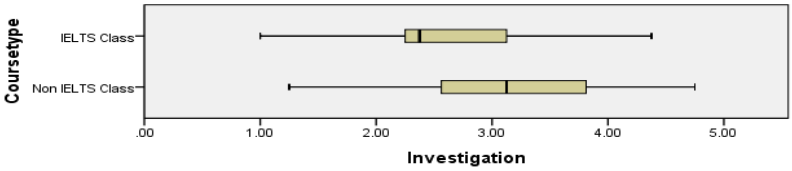

Figure 5. Boxplots of Participants' Scores on Investigation Dimension of Classroom Climate

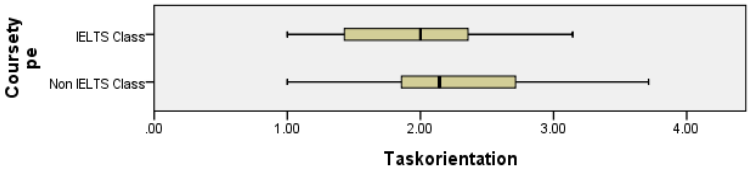

Figure 6. Boxplots of Participants' Scores on Task Orientation Dimension of Classroom Climate

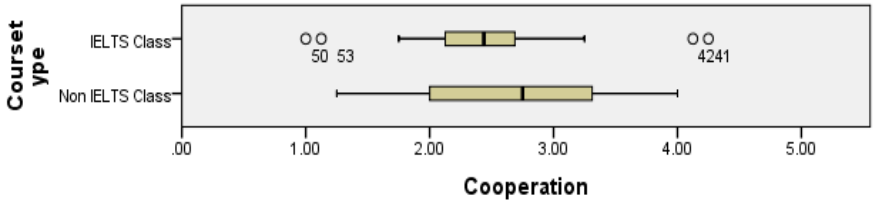

Figure 7. Boxplots of Participants' Scores on Task Orientation Dimension of Classroom Climate

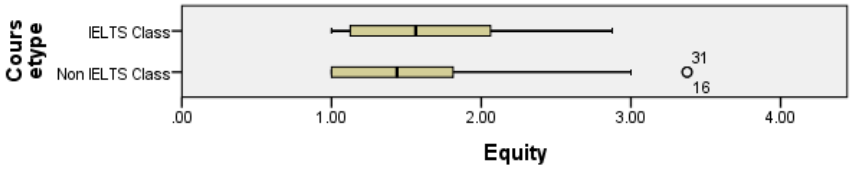

Figure 8. Boxplots of Participants' Scores on Equity Dimension of Classroom Climate

As the boxplots in Figures 2-8 indicate, there are a few extreme scores in some of the measured dimensions. Still, we did not deem it necessary to remove those items on the grounds that first, given the limited range of Likert-scale scores, extreme scores are not likely to remarkably alter the shape of the distribution and, secondly, doing so is in keeping with meeting the independence of the data (Field, 2009). The boxplots also show that except for cohesiveness and equity, the median score of participants in the General English course is higher than the median score of participants in the IELTS preparation course on all the other five dimensions of classroom climate. This may be taken provisional evidence that the classroom climate of the IELTS preparation courses is not as favorable as that of the General English course. Moreover, the lowest median score for both groups of participants is on the equity dimension, which is strikingly low, around 1.5. Table 3 represents the mean and standard deviations of the WIHIC dimensions in the IELTS and General English courses:

Table 3. *Descriptive Statistics of Participants' Scores on Dimensions of WIHIC in the Two Courses*

Group Statistics					
	Class	*N*	Mean	Std. Deviation	Std. Error Mean
Cohesiveness	Non-IELTS Class	35	2.29	.76	.129
	IELTS Class	29	2.40	.66	.122
Teacher Support	Non-IELTS Class	37	2.66	.80	.132

Involvement	IELTS Class	26	2.23	.74	.145
	Non-IELTS Class	37	2.82	.90	.148
Investigation	IELTS Class	27	2.16	.63	.121
	Non-IELTS Class	36	3.11	.85	.141
Task Orientation	IELTS Class	27	2.63	.80	.155
	Non-IELTS Class	37	2.27	.70	.116
Cooperation	IELTS Class	28	1.96	.60	.114
	Non-IELTS Class	36	2.72	.80	.134
Equity	IELTS Class	28	2.47	.71	.135
	Non-IELTS Class	36	1.57	.67	.111
	IELTS Class	28	1.62	.55	.104

Consistent with information in the boxplots, Table 3 indicates that the mean scores of the two dimensions of student cohesiveness and equity in the IELTS course are higher than those of the General English course. On the other five dimensions, the General English course appears to have a more favorable climate. To see whether the apparent differences are larger than chance differences, independent samples *t* tests were conducted:

Table *4. Independent Samples t Test*

	Levene's Test for Equality of Variances		*t* Test for Equality of Means						
	F	*Sig.*	*t*	*df*	*Sig.* (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Cohesiveness	3.215	.078	-.607	62	.546	-.109	.18	-.47	.25
Teacher Support	.470	.496	2.150	61	.036	.429	.20	.03	.82
Involvement	3.139	.081	3.274	62	.002	.66223	.20	.25	1.06
Investigation	.434	.512	2.282	61	.026	.483	.21	.05	.90
Task Orientation	.986	.325	1.853	63	.069	.309	.16	-.02	.64
Cooperation	2.992	.089	1.284	62	.204	.249	.19	-.13	.63
Equity	.110	.741	-.259	62	.796	-.040	.15	-.35	.27

Table 4 shows that the two courses were significantly different on three dimensions: teacher support, involvement, and investigation. Teacher support in the IELTS preparation course was significantly lower than that of the General English course, $t(61) = 2.15$, $p = .036$. Likewise, the degree of involvement was significantly lower in the IELTS preparation course, $t(62)= 3.27$, $p = 002$. The IELTS course obtained a significantly lower score on the investigation dimension, too, $t(61) = 2.28$, $p = .026$. The obtained effect sizes computed using eta-squared (Hatch & Lazaraton, 1991; Pallant, 2013) were .07, .08, and .14 for teacher support, investigation, and

involvement dimensions, respectively. The first two effect sizes are considered medium and the third is regarded as a large effect size (Pallant, 2010).

The second source of data was observation data, collected using C-CAT. Table 4 contains the frequency of each behavior in the two IELTS and General English courses. It is evident that, in most cases, the frequencies are more or less similar. Thus, the frequencies of teacher praise, interruptions in the class, and transitions are rather similar:

Table 5. *Frequency of Occurrence of Each Behavior in IELTS and General English Courses*

Arguing	1	7
Defiance	2	1
Teacher Reprimand	7	4
Transition in Classroom	7	10
Interruption in Classroom	11	13
Teacher Praise	22	17
Teacher Assistance	20	16
Teacher Encouragement	21	19
Interest		
Majority of students were interested.	12	10
Half of students were interested.	8	9
Minority of students were interested.	4	5
Focus		
Majority of students were focused.	18	11
Half of students were focused.	3	7
Minority of students were focused.	3	6

That said, it may be discerned that the General English course seems to be less teacher-centered, as there are seven cases of arguing in the General English course, compared to only one in the IELTS course. Also, attesting to the more dominant role of the teacher in the IELTS course is the frequency of teacher reprimand: seven in the IELTS course compared to only four in the General English course.

To see whether and how far the two courses differed in terms of the frequency of occurrence of each behavior documented in the observation scheme, chi-square test for goodness-of-fit (Gravetter & Wallnau, 2016; Pallant, 2013) was run:

Table 6. *Chi-Square Tests for Observation Scheme Dimensions*

Behavior		Value	*df*	Exact *Sig.* (2-sided)
Arguing	Pearson Chi-Square	5.86	1	.035
Teacher Reprimand	Pearson Chi-Square	.818	1	.36
Transition in Classroom	Pearson Chi-Square	.529	1	..46
Interruption in Classroom	Pearson Chi-Square	.168	1	.68
Teacher Praise	Pearson Chi-Square	.641	1	.423
Teacher Assistance	Pearson Chi-Square	.44	1	.50
Teacher Encouragement	Pearson Chi-Square	1	1	.75
Classroom Level of Interest	Pearson Chi-Square	-	1	Zero Residual
Classroom Level of Focus	Pearson Chi-Square	-	1	Zero Residual

The chi-square test for goodness-of-fit in Table 6 indicates that, except for the arguing behavior, none of the values are significant, showing that the two courses were more or less similar in terms of the behavioral tokens captured by the observational scheme. The two courses, however, differed significantly in the frequency of arguing that took place in the classroom, X^2 (1, N = 8) = 5.86, p = .035).

5. Discussion

The present study investigated the washback of the IELTS speaking module on the classroom climate of the IELTS preparation courses. To do so, the classroom climate of an IELTS speaking course was compared with that of a general conversation course. Back to the research questions, the answer to the first research question is both yes and no: IELTS has washback on some dimensions of classroom climate, but not on others. The second research question was about which dimensions of classroom climate are affected by IELTS. It was found that washback to the student involvement dimension of WIHIC was the most evident. The other two dimensions seemingly affected by IELTS were investigation and teacher support. The effect size for the latter two dimensions was medium. On the other four classroom climate dimensions, including cohesiveness, task orientation, cooperation, and equity, IELTS seemed to have a positive influence, as no significant differences were found between the climate of the IELTS course and that of the General English course on the four noted dimensions.

Generally speaking, regarding the value dimension of washback (Watanabe, 2004a), it appeared that the washback of IELTS on the classroom climate of the preparation course was both positive and negative. In other words, observational and questionnaire data suggest that the climate of the IELTS speaking course resembled that of the General English course in some of the classroom climate dimensions. This could be taken as positive washback. On other dimensions of class climate, however, the classroom climate of the IELTS course seemed to have been negatively affected, as realized through less learner involvement in class activities, less teacher support, and less emphasis on inquiry and problem-solving.

The findings of this study concur with some findings from similar research on the washback of language tests, in general, and the washback of IELTS, in specific. For one thing, the findings attest to the highly complex nature of washback well documented in the literature (Alderson & Hamp-Lyons, 1996; Cheng et al., 2015; Wall & Alderson, 1993). The complexity problem is compounded by numerous macro and microcontextual elements that might moderate or mediate test washback (Watanabe, 2004a). Though the focus in this study was on how classroom climate was affected by IELTS, it is naïve to assume a unidirectional process of influence. That is, classroom climate can plausibly be taken to mediate the washback process. Past research on the washback of IELTS points to both negative and positive effects (Ahmed, 2015; Moore & Morton, 2005). For instance, focusing on the writing task of IELTS, Moore and Morton (2005) found evidence of negative washback of IELTS, due to the fact that the construct of academic writing is underrepresented in the IELTS writing tasks. In specific, they found that IELTS writing tasks differed from academic writing tasks in several dimensions, such as genre, information source, rhetoric function, and object of inquiry.

Similarly, in the present study, we found partial evidence of negative washback effect of the IELTS on a few dimensions of the climate of the preparation course. In particular, the questionnaire data showed that the IELTS course was somehow more teacher-centered, a finding consonant with the existing scant literature (Alderson & Hamp-Lyons, 1996; Hayes & Read, 2004; Plank & Condliffe, 2013). The largest difference between the two courses on the dimensions of classroom climate was about the students' involvement in classroom events, which was significantly lower in the IELTS course. Similarly, observational data indicated that the occurrence of student-centered behaviors like arguing was significantly lower in the IELTS course. Seen in the light of the literature noted above, the observed effect of IELTS on classroom climate was quite low on the specifity dimension of washback (Watanabe, 2004a). In other words, the observed washback seems to characterize test preparation courses, in general, as found in Plank and Condliffe (2013), too.

It is, however, difficult to attribute the observed findings to the IELTS with any certainty. For one thing, as past research suggests, learner expectations mediate and moderate washback from tests (Green, 2006; Xie & Andrews, 2013). Similarly, it might be that students' expectations in the IELTS preparation course affect the classroom climate of the IELTS preparation course. Thus, the students in the IELTS preparation course might think that they would make more out of the course if the class is dominated by the teacher because learners are not traditionally seen as legitimate sources of knowledge. In addition to student expectations, their language proficiency might also moderate the intensity, duration, and direction of test washback. The students in the General English course were majoring in English. As such, they were supposedly more proficient than those in the IELTS course, which might explain why the General English course enjoyed more learner participation and involvement.

In addition, teacher factors have been shown to play a major role in shaping test washback (Alderson & Hamp-Lyons, 1996; Anagnostopoulos, 2003; Watanabe, 1996, 2004b). The curricular status of the teacher, in particular, is known to bear on the extent to which teachers are influenced by tests (Anagnostopoulos, 2003). The fact that the course observed in this study was run by a tenured faculty member with a higher curricular status in the department may partly explain why the effect of the IELTS on the preparation course was not remarkably obvious. Teachers with high curricular status are less likely to feel compelled to get out of their comfort zone to tailor their instructional styles to the demands of high-stakes tests (Anagnostopoulos, 2003). Still a further possibility for this lack of obvious IELTS washback on several dimensions of classroom climate might, in fact, be due to the buffering, decoupled nature of classes, in general, as Coburn (2004) maintains.

6. Conclusion and Implications

This study intended to investigate the washback of IELTS to classroom climate. Overall, it was found that, in the IELTS course, the amount of teacher support and the degree of student involvement in class activities were considerably less than in the General English course. Assuming that factors other than test washback did not bear on classroom climate, which is a very difficult assumption to make, we may surmise that, in test preparation courses, the instrumental motivation of students encourages the teacher to forgo humanistic aspects of an ideal language class. Once the end becomes helping learners passing a test, the means to such an end might be perceived to be other than adhering to best practice in language teaching and learning. Yet, for test preparation courses to be educationally defensible (Hamp-Lyons, 1998; Popham, 1991), teachers should resist succumbing to the urge of engaging in practices that are in conflict with current theories and approaches to language

learning. Because if they do, there are two possibilities: Either teachers would help students pass the exam or they would not. If the former, that is, if learners are aided in passing an exam without going through genuine communicative activities, which is what consequentially valid tests seek to foster, the validity of interpretations and decisions made of test scores would be compromised. If the latter, that is, if students fail to pass the exam after test preparation courses, then they have been denied both the chance of experiencing good language learning practices and the chance of passing the test.

Language testing has moved from seeking to measure and develop communicative competence (Bachman & Palmer, 1996) to testing and development of interactional competence (Ahmadi & Montasseri, 2019; Galaczi & Taylor, 2018). Developing interactional competence, which posits that interaction is the coconstruction of discourse across participants in a speech event, is only possible in a classroom environment that puts interaction among learners above teacher-centered pedagogy. For this to happen, the classroom climate should also be rich enough in emotional support and friendliness. Nevertheless, the findings from this study and similar studies suggest that teachers seem to believe that good teaching in test preparation is different from good language teaching, in general (Huang, 2018). Alderson and Hamp-Lyons (1996) maintain that some of what teachers did in TOEFL test preparation courses had zero to do with what it takes to pass TOEFL. It seems reasonable to suggest that there is a collective culture of test preparation among teachers and learners that does not necessarily work in their best interest. Perhaps, this could be remedied during preservice and in-service language teaching programs. The reality is that preparation for different language tests constitutes a big portion of language teaching that goes on around the world (Huang, 2018). Iran biggest language institute (i.e., ILI), with hundreds of branches, has identified its mission as helping learners pass university admission tests. Now that test preparation is so prevalent, it is both justified and necessary to train language teachers in the best practice of running language test preparation practices (Huang, 2018). A key element of such training would be how to foster classroom climate conducive to fostering learners' interactional competence.

The present study had its share of limitations: First, classroom climate is jointly constructed by a classroom teacher and students in the classroom. Ideally, to isolate the effect of a test on classroom climate, one would need to control or neutralize the effect of both the teacher and the students. In the present study, we could not neutralize the effect of learners because it would necessitate choosing one group of learners taking both a General English course and a test preparation course simultaneously with the same teacher. In a similar vein, the fact that only one teacher was observed in the two courses allows for the possibility that the findings are still,

at least partly, attributable to teacher characteristics. Observing multiple pairs of classrooms taught by several teachers would improve the internal validity of the study.

In this study, we used a general measure of classroom climate, which might have not captured the nuances of a test preparation course. Future inquiry might benefit from a measure specifically developed for assessing the climate of exam preparation courses. Yet, another limitation of the study, pertaining to the observation data, was that the observation data were coded by only one coder. Multiple observers or the use of video-recording, with more than one coder, can provide more solid evidence, regarding how high-stakes tests influence classroom climate.

References

Ahmadi, A., & Montasseri, Z. (2019). Interactional competence in paired vs. group oral tests. *Teaching English Language, 13*(1), 1-26.

Ahmed, N. (2015). *The washback effects of teaching academic IELTS on Emirati students' language proficiency at a technical high school in Abu Dhabi*. The British University in Dubai (BUiD).

Alderson, C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing, 13*(3), 280-297.

Alderson, C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115-129.

Allen, D. (2016). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia, 6*(1), 1-20.

Anagnostopoulos, D. (2003). Testing and student engagement with literature in urban classrooms: A multilayered perspective. *Research in the Teaching of English*, 177-212.

Anagnostopoulos, D. (2006). "Real students" and "true demotes": Ending social promotion and the moral ordering of urban high schools. *American Educational Research Journal, 43*(1), 5-42.

Bachman, L. F. (2004). *Statistical analyses for language assessment.* UK: Cambridge University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). UK: Oxford University Press.

Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing, 13*(3), 257-279.

Beld, M., van der Voort, D., van der Helm, G., Kuiper, C., De Swart, J., & Stams, G. (2018). Assessing classroom climate in special education: A validation study of the special education classroom climate inventory. *Journal of Psychoeducational Assessment, 36*(7), 736-749.

Bidwell, C. E. (2001). Analyzing schools as organizations: Long-term permanence and short-term change. *Sociology of Education, 5*, 100-114.

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.

Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education, 77*(3), 211-244.

Curby, T. W., Johnson, P., Mashburn, A. J., & Carlis, L. (2016). Live versus video observations: Comparing the reliability and validity of two methods of assessing classroom quality. *Journal of Psychoeducational Assessment, 34*(8), 765-781.

Diamond, J. B. (2007). Where the rubber meets the road: Rethinking the connection between high-stakes testing policy and classroom instruction. *Sociology of Education, 80*(4), 285-313.

Dunn, R. J., & Harris, L. G. (1998). Organizational dimensions of climate and the impact on school achievement. *Journal of Instructional Psychology, 25*(2), 100-114.

Estaji, M., & Tajeddin, Z. (2012). The learner factor in washback context: An empirical study investigating the washback of the IELTS academic writing test. *Language Testing in Asia, 2*(1), 5-25.

Field, A. (2009). *Discovering statistics using SPSS*. London: SAGE publications.

Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualizations, operationalizations, and outstanding questions. *Language Assessment Quarterly, 15*(3), 219-236.

Ghaith, G. (2003). The relationship between forms of instruction, achievement and perceptions of classroom climate. *Educational research, 45*(1), 83-93.

Gravetter, F. J., & Wallnau, L. B. (2016). *Statistics for the behavioral sciences*. Boston: Cengage Learning.

Green, A. (2006). Washback to the learner: Learner and teacher perspectives on IELTS preparation course expectations and outcomes. *Assessing Writing, 11*(2), 113-134.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*. London: SAGE Publications.

Hamp-Lyons, L. (1997). Washback, impact, and validity: Ethical concerns. *Language Testing, 14*(3), 295-303.

Hamp-Lyons, L. (1998). Ethical test preparation practice: The case of the TOEFL. *TESOL Quarterly, 32*(2), 329-337.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76*(5), 949-967.

Hatch, E. M., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House Publishers.

Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing* (pp. 97-112). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125-132.

Huang, B. (2018). PD in testing preparation for university ESL. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1-7). New York: John Wiley & Sons, Inc.

Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.

Khajavy, G. H., MacIntyre, P. D., & Barabadi, E. (2018). Role of the emotions and classroom environment in willingness to communicate: Applying doubly latent multilevel analysis in second language acquisition research. *Studies in Second Language Acquisition, 40*(3), 605-624.

Larsen-Freeman, D., & Anderson, M. (2013). *Techniques and principles in language teaching*. Oxford: Oxford University Press.

Larson-Hall, J., & Herrington, R. (2009). Improving data analysis in second language acquisition by utilizing modern developments in applied statistics. *Applied Linguistics, 31*(3), 368-390.

Leff, S. S., Thomas, D. E., Shapiro, E. S., Paskewich, B., Wilson, K., Necowitz-Hoffman, B., & Jawad, A. F. (2011). Developing and validating a new

classroom climate observation assessment tool. *Journal of School Violence, 10*(2), 165-184.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241-256.

Mickan, P., & Motteram, J. (2009). The preparation practices of IELTS candidates: Case studies. *International English Language Testing System (IELTS) Research Reports, 10*, 223-243.

Moore, T., & Morton, J. (2005). Dimensions of difference: A comparison of university writing and IELTS writing. *Journal of English for Academic Purposes, 4*(1), 43-66.

Naghdipour, B. (2016). English writing instruction in Iran: Implications for second language writing curriculum and pedagogy. *Journal of Second Language Writing, 32*, 81-87.

Olson, D. R. (2009). The history of writing. In B. Roger, M. Debra, N. Martin, & R. Jeni (Eds.), *The SAGE handbook of writing development* (pp. 6-16). London: Sage.

Pallant, J. (2013). *SPSS survival manual: A step-by-step guide to data analysis using SPSS*. Berkshire: McGraw-Hill Education

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109-119.

Pickett, L. H., & Fraser, B. J. (2002). *The role of learning environment, achievement, and student and teacher attitudes in a science mentoring program for beginning elementary school teachers*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Plank, S. B., & Condliffe, B. F. (2013). Pressures of the season: An examination of classroom quality and high-stakes accountability. *American Educational Research Journal, 50*(5), 1152-1182.

Popham, W. J. (1991). Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice, 10*(4), 12-15.

Qin, X. (2011). Is test-taker perception of assessment related to construct validity. *International Journal of Testing, 11*(4), 324-348.

Rapp, D. (2002). National board certified teachers in Ohio give state education policy, classroom climate, and high-stakes testing a grade of F. *Phi Delta Kappan, 84*(3), 215-218.

Riazi, A. M., & Razavipour, K. (2011). (In) Agency of EFL teachers under the negative backwash effect of centralized tests. *International Journal of Language Studies, 5*(2), 262-283.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458.

Saville, N., & Hawkey, R. (2004). The IELTS impact study: Investigating washback on teaching materials. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing: Research contexts and methods* (pp. 73-96). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Shohamy, E. (2014). *The power of tests: A critical perspective on the uses of language tests*. New York: Routledge.

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing, 13*, 298-317.

Sink, C. A., & Spencer, L. R. (2005). My class inventory-short form as an accountability tool for elementary school counselors to measure classroom climate. *Professional School Counseling, 9*(1), 37-48.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics*. Boston, MA: Pearson

Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing, 10*(1), 41-69.

Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing, 13*(3), 318-333.

Watanabe, Y. (2004a). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing* (pp. 41-58). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Watanabe, Y. (2004b). Teacher factors mediating washback. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 129-146). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Wright, T. (2005). *Classroom management in language education*. New York: Palgrave Macmillan.

Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with structural equation modeling. *Language Testing, 30*(1), 49-70.