

Exploring Washback in Japanese EFL Classrooms: Processes and instruments for researching washback

LTRC Poster Session

July 30th, 1999 at Tsukuba International Congress Center

Presenter: WATANABE, J. Yoshinori

International Christian University

A very brief summary of the research into washback of the Japanese university entrance examinations (slightly revised version of Watanabe, 1997)

I. Conceptualising washback as a phenomenon

1. Dimensions: Washback seems to operate on several dimensions

- Intensity: strong (test will influence everything/everyone involved in the test) / weak (test will influence only some aspects of instruction, learning, programme, society, etc.)
- Specificity: specific (the influence that only some specific test types or specific aspects of the test will have; e.g., a listening test will induce learners to focus on this skill in preparation) / general (the influence that any test will have; e.g., tests will give incentive to learners)
- Nature: positive (test will help achieve the educational goal/purpose) / negative (test will become an obstruct to achieving the educational purpose/goal)
- Length: short (the influence of the test will last only for a short period of time; the influence vanishes soon after the test administration) / long-term (the influence of the test will last for a considerable length of time; e.g., the influence of the entrance exam on students lasts even after they enter college)

2. Factors to be considered: Washback will be function of the following factors.

- Test factors : method, format, content, skills tested, etc.
- Test use factors: purpose of the test (placement, achievement, proficiency, etc.), decisions that will be made on the basis of the test result, method of score reporting, frequency of test administration, etc.
- Prestige factors (designated by Brown, 1997: stakes of the test, status of the test in the whole educational system.
- Personal factors: past educational backgrounds, teaching/learning

- experience, beliefs about the best method of teaching/learning, etc.
 - Micro-context factors: the school setting, where the preparation for the test is being done.
 - Macro-context factors: the society where the test is used.
- 3. Variables to be investigated:** washback will be observed on all or a combination of some of the following variables.
- Personal variables (the ‘intervening’ variables which may be influenced by the exam, and also that may influence a subsequent behaviours: anxiety (debilitating and facilitating; trait and state); motivation; perception about the test (what characteristics test-takers, instructions perceive the test to have, what they really know about the test), etc.
 - Activity/ interaction variables: how/what teachers teach, how/what learners learn, the interaction that is going on between the two in the classroom, the teaching/learning materials being used.

II. The research

1. Methodology

- Ethnography (interviews, classroom observations): (LeCompte & Preissle, 1993)
 - ethnography elicits phenomenological data that represent the world view of the participants being investigated and participants’ constructs are used to structure the research;
 - ethnography employs participant and non-participant observation to acquire firsthand, sensory accounts of phenomena as they occur in real world setting;
 - in ethnographic research, the researchers seek to construct descriptions of total phenomena within their various contexts and to generate from these descriptions the complex interrelationship of causes and consequences that affect human behaviour toward and belief about phenomena;
 - the ethnographic researchers use a variety of research techniques to amass their data.
- Cross-comparative: compared several teaching situations which had a variety of backgrounds. (N.B. two types of longitudinal research into washback; longitudinal 1: before and after the administration of the test and longitudinal 2: before and after the new exam is implemented.)

2. Description of Macro-context factors: the Japanese university entrance examinations.

- Three types of universities/colleges: national (N=98; 18.4%)/ local public (N= 46; 8.6%) / private (N=502; 89.8%).
- Two major types of entering universities/ colleges: recommendation based on the school record/ taking the test (each university/college uses its own produced test, except national/local public, where NCUEE, produced by the Ministry of Education is required as the first stage exam, followed by the second stage exam).
- The exam student normally sits more than one exam, ranks them according to the (in many cases) prestige of the universities/colleges.

3. Description of micro-context factors: An attempt was made to gather data from various contexts, where exam preparation would be done.

- Yobiko (a special exam preparatory school): preparing students with a particular emphasis on the English section of the exams (of national, local public, and private).
- High school A: a traditional rural high school, teachers seeking for the ideal method of teaching English for entrance exam and communication. Very few yobiko nearby, so students (100% of whom wished to enter higher education) relied on the school for preparing for the exam.
- High school T: teaching English for communicative purposes, implementing innovative approaches (speech contests, inviting students from overseas, etc.). Located in rural area, no yobiko nearby, so students (100% of whom wished to enter higher education) placed total reliance on the school for preparation.
- High school I: located in metropolitan area, there were many returnee students (100% of whom wished to enter higher education), graduating from high schools where English was the first language. There were many yobiko nearby, so students (teachers said) were going to the special school for preparing for the exam.

4. Description of participants (personal factors)

- Yobiko: Teacher G (late 20's), believed in the effectiveness of traditional teaching (e.g., reading aloud, translation, grammatical explanations, etc.), graduated from a national university, was interested in cognitive linguistics/// Teacher H (mid-30's), believed in the possibility of teaching English both for communication and examination, was graduated from a private university, was interested in communicative approaching of teaching English.

- High school A: Teacher A (late 20s), was trying to teach English for both communication and exam, employing a variety of methods (using video, pair work, etc.)/// Teacher B (late 30s), was happy with a traditional approach (e.g., grammar-translation, coral reading, etc.).
- High school B: Teacher C (mid 50s), was trying to innovate education through exam preparation, implementing innovative approaches to teaching English for examination (games, pair-work, group-work, etc.)
- High school I: Teacher E (mid 50s), believed returnee students would need improvement in grammar and translation as language skills, thus emphasising these skills. ///Teacher F (mid-20s), was trying to use exam preparation as a chance of developing communication ability.

6. Description of test factors:

- target exam: the English section of the Japanese entrance examination (as of 1994)
- purpose: screening
- status: varies according to teachers/students (see 2 1))
- stakes: varies for individual teachers/students (see 2 1))
- contents: to describe the content of the target exam of each of the courses observed, the following steps were taken.
 - interviewed high school teachers about their target exams;
 - analysed the target exams in terms of the categories that were derived from teacher interviews (e.g., listening, paper test of pronunciation, gap-filling, etc.)

7. **Predictions:** derived by analysing documents (i.e., newspapers, magazines, TV broadcast, and interviewing people who were actually involved in preparing for the exams) + test analyses

Washback on specific dimension

Test content

- translation → (literal) translation
- grammar → use of metalanguage + reference to formal aspects of English + teacher's error correction of students' response (use or form)
- summary → reference to coherence + picking up main ideas
- listening test → aural/oral use of English

Test methods

- constructive response → use of extended length of English
- selective response → reference to test-taking techniques

- indirect test methods (e.g., written test of pronunciation, gap-filling of written dialogue, unscrambling sentences into a coherent text) → aural/oral practice, writing practice

Language used in the test

- English (e.g., for instructions, mcq items) → use of English
- Japanese → use of Japanese

Washback on general dimension

- tense classroom atmosphere → incidents of laughter
- limited types of classroom organisations → length of time spent on each organisation type (e.g., lockstep, pair work, group work, etc.)
- stifled innovation → incidents of innovative teaching employed (e.g., games, teacher produced materials, etc.)
- students' attitudes → asking questions spontaneously

8. Research assumptions: In order to completely prove that washback exists, it is necessary to exclude all the possibilities other than exams which may affect teaching/learning.

Washback on specific dimension

- 1) Washback on specific dimension addresses the question, “would teaching/learning become different if the exams were to change?”
- 2) Washback is considered to exist on specific dimension if at least the following conditions were met.

	Teacher C		Teacher D
Exam A	↑	←(2) [same]→	↑
	(1) [different]		(1) [different]
Exam B	↓	←(2) [same]→	↓

Notes:

- Exam A is different from Exam B in their contents and methods.
- Exam A may be of use at the same period of time when Exam B is of use. (Cross-sectional study)
- Exam A may be a revised version of Exam B. (Longitudinal study)
- Teacher C is different from Teacher D.
- Each shaded cell represents classroom events and materials being used.

- (1) Action/ interaction variables (including teaching materials) are different in the courses taught by the same teacher.
- (2) Action/ interaction variables (including teaching materials) are similar in the courses taught by two different teachers.

Washback on general dimension

- 1) Washback on general dimension addresses the question, “would teaching/learning become different if there were no exams?”
- 2) Washback is considered to exist on general dimension if at least the following conditions were met.

	Teacher C		Teacher D
Exam prep. lessons	↑	←(2) [same]→	↑
	(1) [different]		(1) [different]
Non-exam lessons	↓		↓

Notes:

- In exam prep(eration) lessons, teachers aim at a variety of target exams.
- Teacher C is different from Teacher D.
- Each shaded cell represents classroom events and materials being used.

- (1) Action/ interaction variables (including teaching materials) are different in exam-prep and non-exam lessons taught by the same teacher.
- (2) Action/ interaction variables (including teaching materials) are similar in exam-prep lessons taught by two different teachers, and the action/ interaction/ materials are those which can be predicted from each teacher’s target exams.

10. Data gathering and analyses

Lessons observed

- Yobiko (a total of 1,797 minutes): Teacher G teaching for a private exam + a group of national exams /// Teacher H teaching for the same pair of the exams
- High school (a total of 652 minutes): all the teachers teaching for regular main-stream and special exam preparation courses.

Instruments: To be presented at an individual session.

Process: To be specified at an individual session.

- Before observations: interviewed teachers (about the purpose, goal, materials to be covered, etc.); received/borrowed textbooks.
- During observations: wrote field-notes, audio-record lessons.
- After observations: summarised the impressions, analyse the materials used, completed field-notes listening to the tape/ looking into memory.
- Developed coding sheet (on the basis of the field-notes) for frequency

count.

- Coded data by listening to the audio-tape.
- Computed frequencies.
- Examined the results in the light of the research assumptions.

11. Results

(Note: TE = Target exam.)

Washback on specific dimension

Test content

- translation: even where TE did not include translation, the teachers who believed in its effectiveness was using this method.
- grammar: irrespective of TE, teachers (e.g., Teacher G and F), who felt it necessary to teach grammar to students, used metalanguage, and referred to formal aspects of English. Teachers rarely corrected students errors.
- summary: teachers tended to refer to coherence of the text when TE includes summary, but not necessarily placed importance on getting main ideas.
- listening test: even where TE included listening section, teachers did not necessarily teaching listening. In the cases where this skill was taught (Teacher F of school I, and Teacher C of school T), there were a school atmosphere aural/oral use of English was taken for granted.

Test methods

- constructive response: even where TE included writing, this skill was not necessarily taught, because 'score method/ criteria' was not known.
- selective response: Teachers deliberately avoided referring to test-taking techniques. They believed that real English skills would lead to students' passing of the exam.
- indirect test methods (e.g., written test of pronunciation, gap-filling of written dialogue, unscrambling sentences into a coherent text): did not lead to practice in the target skill, because teachers did not know an appropriate method of teaching these skills.

Language used in the test

- There was a tendency that teachers used English where the instruction and multiple-choice questions were written in English than in Japanese. But the use of English was done simply by reading aloud rather than authentic use.

Washback on general dimension

- tense classroom atmosphere: atmosphere of exam classes were not

necessarily tense. Seemed to depend upon teacher's personality (e.g., anxiety, guilt feelings about teaching for the exams).

- limited types of classroom organisations: teachers who were trying to innovate exam classes were using a variety of organisation patterns.
- stifled innovation: there were several cases where teachers were trying to innovate exam classes (e.g., revise listening task taken from past exam papers, preparing hand-outs explaining cultural background to English words, etc.)
- students' attitudes: students rarely asked questions in classes other than high school I, where there were a number of returnee students.

Analyses of the results on dimension

- intensity: weak washback present; unitary notion of washback does not hold (varies according to teachers' educational background, experience, beliefs about the best teaching method)
- specificity: specific test content/method does not necessarily generate specific washback.
- nature: negative (e.g., indirect testing method did not lead to practice) // positive (summary, listening, mcq in English rather than in Japanese, etc.), but depending upon teachers, classes, and schools.
- length: yet to be investigated (see Watanabe, 1992).
- range: yet to be investigated.

Factors that seemed to have been involved in inducing washback

- Test specifications (including, scoring method, key, etc.) are not made public; information about the future exams was not shared among test users. For example, teachers do not know how and on which criteria the writing section of the test is marked, so they are not willing to teach writing for the exams.
- Teachers are likely to have guilt feelings about test coaching. Thus, however 'good' the test might be, washback is not easily coming in.
- Teacher's beliefs about the best teaching method overrides test effects. And those beliefs seem to come from teacher's educational background, teaching training, and university majors. Thus, even where the test is written all in English, the teacher, if he/she strongly believes in the effectiveness of translation method, is likely to translate.
- School atmosphere (micro-context factor) may override the test effects. Thus, even if the target exam includes very little listening, teachers are likely to use listening materials (taken from the past exam paper) for the purpose of exam preparation.

- Cultural tradition (macro-context factor) may override the test effects. For example, students tend to be passive listeners in exam classes, but this may not be due to the effect of the exams, but a cultural tradition may be telling them (implicitly) that students need to listen carefully what teacher says.

III. Conclusion

The test factors alone do not cause washback to the classroom. Washback is induced via various factors, and those include people, micro-context, and macro-context factors.

VI. For future research

Future research needs to investigate;

- longitudinal nature of the washback. A new curriculum has since been implemented, and the exams based on the curriculum began in 1997. Also, the macro-context has been changing, especially in a way in which, the total number of students enrolling in higher education becomes smaller as the birth rate declines. The new data set, including the description of new exams and their washback, need to be plotted against the present data.
- test specifications. Examine if test centres of each university has its test specifications. If not, why not? If yes, what they contain? If yes, can they be made public? Why or why not?
- washback of smaller-scale lower-stake tests (i.e., in-class test, practice test, placement test, etc.). Detailed description about the development and administration of the test and its relationship to action/interaction variables.
- comparison (regarding washback) of the lower-stakes tests with high-stakes tests (e.g., entrance exams, school leaving exams).
- washback to the learner (motivation, attitudes, etc.)

Hypotheses generated from students interviews in the present research

1. An exam which is more important to a student will have a greater effect than an exam which is less important
2. The sections of the exam which are perceived to be more difficult, thus, more discriminatively powerful will induce greater washback than those which are not.
3. The sections which are perceived to be less difficult, thus, less discriminatively powerful, will have a smaller effect.
4. If the test section is perceived to be too difficult, then it will discourage students; thus, washback will not be engineered.

5. Test methods (selective response/ constructive response) will not induce differential washback effects on how students study.
6. The exams may influence students through practice test.
7. Exam preparatory teaching will influence students if it matches their target university exams, but if it does not match, then it will not have an effect.