

*Statistics Corner: Questions and answers about language testing statistics:*

## **Extraneous variables and the washback effect**

James Dean Brown (University of Hawai'i at Manoa)

**QUESTION:** In your 1988 book *Understanding Research in Second Language Learning* you mention different types of extraneous variables such as subject expectancy, the halo effect, and the Hawthorne effect. Can you explain the difference between these terms? Also, what is the relation of these terms with washback? Finally, can you explain why some language researchers prefer to avoid the term "washback"?

**ANSWER:** Let me rephrase these questions and address them in the following order: (a) What are the different extraneous variables that researchers must guard against? (b) What is "washback", and why do some language researchers avoid the term? (c) What is the relationship between extraneous variables and washback?

### ***What Are the Extraneous Variables That Researchers Must Guard Against?***

In Brown (1988), I described a whole set of extraneous variables that might affect the correct interpretation of a statistical study. I categorized these into four main categories: environment issues, grouping issues, people issues, and measurement issues.

*Environment issues.* Environment issues include naturally occurring variables (i.e., those which occur naturally in the research setting, like noise, temperature, adequacy of light, time of day, seating arrangements, etc.) and artificiality (unnatural arrangements within the study, e.g., the effects of students performing in front of a video camera, or under other artificial conditions).

*Grouping issues.* Grouping issues are related to the initial make up of the groups or changes in their composition over the course of the study. These include self-selection (the practice of letting research participants choose to enter a study or decide which group to join, e.g., volunteerism), mortality (the effects of participants dropping out of the study), and maturation (the effects of different experiences on various participants, e.g., other simultaneous learning, puberty, family catastrophes, etc.).

*People issues.* People issues include the Hawthorne effect (the fact of being included in a study may affect the behavior of the participants in a study, and therefore affect the results), halo effect (the human tendency to respond positively to a person, e.g., the researcher, treatment teacher, etc., may affect the results of the study), subject expectancy (the participants may guess what a study is about and then consciously or subconsciously "help" or resist the objectives of the research), and researcher expectancy (instances where the attitudes and motivations of the researchers themselves affect or color the results of a study).

*Measurement issues.* Because the results of a study are only as good as the data upon which they are based, it is crucial to insure that the measures themselves are not introducing extraneous variables such as the *practice effect* (the potential influence over time of the measures in a study on each other, e.g., the effect of a pretest on a subsequent similar posttest), *reactivity effect* (the influence of parts of a measure on subsequent performance on other parts of the measure, e.g., answering questions early in a questionnaire might cause participants to form opinions or attitudes that would affect their answers later in the questionnaire), *instability measures* (the degree to which inconsistent or unreliable measurements affect the study), and *instability of study results* (the degree to which the results of a study are likely to occur again if the study were replicated).



For the most part, extraneous variables are a threat to the internal reliability and validity of a research project. Essentially, such extraneous variables, if not controlled, or otherwise accounted for in a study, are all potential *intervening variables* (i.e., unanticipated variables that could explain the outcomes of a study as well as the conclusions drawn by the authors). As I put it in my 1988 book, "In statistical studies, there are a number of problems that can arise - both within a study and from outside of it that may create major flaws in its *validity*, i.e., the degree to which a study and its results correctly lead to, or support, exactly what is claimed. The problems themselves result from extraneous variables that are relevant to a study but are not noticed or controlled."

### ***What Is "Washback" and Why Do Some Language Researchers Avoid the Term?***

For readers who may not be familiar with the term "washback", let's look briefly at some definitions: For Shohamy, Donitsa-Schmidt, and Ferman (1996), washback is "the connections between testing and learning" (p. 298); to Gates (1995), it is "the influence of testing on teaching and learning" (p. 101); and for Messick (1996) washback is "the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning" (p. 241). Clearly then, the washback is roughly speaking the effect of testing on the teaching and learning processes. An example that often comes up in Japan is the effect of the university entrance examinations in Japan on high school language teaching and learning.

I wasn't aware that some researchers in our field are avoiding the term "washback". However, I can see how that might be the case, for two reasons. First, the very existence of the concept of washback has been questioned (see for instance, Alderson and Wall, 1993). However, since 1993, a considerable literature has emerged on the topic of washback, which seems to indicate that washback does exist (see, for example, Cheng & Watanabe, forthcoming). As shown in Table 1, washback can be analyzed into aspects of a curriculum that negative washback can affect and ways that positive washback can be fostered (see Brown, 1999, for discussion of both positive and negative washback, or Brown, 2000 for more details on fostering positive washback).

Table 1. *Negative and Positive Effects of Washback.*

	
<b>NEGATIVE WASHBACK CAN AFFECT:</b>	<b>WAYS POSITIVE WASHBACK CAN BE FOSTERED:</b>
<ul style="list-style-type: none"><li>* <b>Teaching</b></li><li>* <b>Course content</b></li><li>* <b>Course characteristics</b></li><li>* <b>Class time</b></li></ul>	<ul style="list-style-type: none"><li>* <b>Alter test design factors</b></li><li>* <b>Change test content factors</b></li><li>* <b>Adjust test logistics factors</b></li><li>* <b>Modify test interpretation factors</b></li></ul>

Second, many authors simply use other terms for the same basic concept as washback and thereby avoid the term. For example, in the general education literature, this concept is sometimes referred to as *backwash*, while elsewhere it is referred to variously as *test impact*, *test feedback*, *curriculum alignment*, and *measurement-driven instruction*. So, in direct answer to your question, I'm not sure language researchers are avoiding the concept of washback, but, as with so many concepts in the language teaching literature, various authors may be using different terminology to discuss it.

Washback, whether it is positive or negative, can be a potential boon or threat to language teaching curriculum (broadly defined) because, through washback, a test can steer a curriculum in one direction or another (in terms of teaching, course content, course characteristics, and/or class time) either with or against the better judgment of the administrators, teachers, students, parents, etc.

From the point of view of testing, thinking about washback can help us to think about test validity. Washback becomes negative washback when there is a mismatch between the construct definition and the test, or between the content (e.g., the material/abilities being taught) and the test. Given that the definition of validity is the degree to which a test is measuring what it claims to measure, any such mismatch between the construct or content that a test is designed to measure and the test, would be a threat to the test's validity.

For example, as long as the official English language teaching curriculum in Japan was *yakudoku* (roughly translated as the grammar-translation reading method), those university entrance examinations that tested in ways consistent with *yakudoku* could be viewed as valid to the degree they matched the curriculum. However, once the government ministry issued the 1993 guidelines for communicative language teaching, a mismatch was created between the *yakudoku* entrance examinations and any curricula that had actually responded to the ministry's guidelines. Thus, the *yakudoku* entrance examinations are seen by some to be creating negative washback on the communicative curriculum.

Thinking about washback can also lead us to think about the consequential basis for test validity in terms of the social consequences of test use and the values implications of test interpretations, but that is a story for another day (for more on these topics, see Messick, 1988, 1989; Brown, 1999).

### ***What Is the Relationship Between Extraneous Variables and Washback?***

Before I received your questions, I had never previously considered the relationship between the extraneous variables listed in the first section of this article and the concept of washback discussed in the second section. To be perfectly honest, when I first considered the question, I didn't see any connection whatsoever. After all, the effect of extraneous variables is a *research* issue, and the washback effect of test results is a *curriculum* issue.

However, as often happens, in writing about the relationship between extraneous variables and washback, I began to see that there was a connection: extraneous variables can have unintended, but nonetheless important, consequences on research very much in the way that test washback can have unintended, but nonetheless important, consequences on curriculum. Looking at the connection from another angle, extraneous variables can be seen as having a sort of washback effect on research either in positive or negative ways depending on whether such variables were accounted for in the research. Similarly, washback can be viewed as an extraneous variable affecting curriculum either in positive or negative ways depending on whether washback was accounted for in the curriculum and anticipated in the test design and use.

While I may be stretching things a bit here, there is no question that extraneous variables are an important aspect of the research endeavor and that washback is an important aspect of the testing/curriculum endeavor. So thank you for raising these questions.

### **References**

- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Brown, J. D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design*. London: Cambridge University Press.
- Brown, J. D. (1999). The roles and responsibilities of assessment in foreign language education. *JLTA Journal*, 2, 1-21.
- Brown, J. D. (2000). University Entrance Examinations: Strategies for creating positive washback on English language teaching in Japan. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 3 (2), 4-8. Retrieved March 1, 2001 from the World Wide Web: [http://jalt.org/test/bro\\_5.htm](http://jalt.org/test/bro_5.htm)

Cheng, L. & Watanabe, Y. (2004). *Washback in Language Testing: Research Contexts and Methods*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gates, S. (1995). Exploiting washback from standardized tests. In J. D. Brown & S. O. Yamashita (Eds.), *Language Testing in Japan* (pp. 101-106). Tokyo: Japanese Association for Language Teaching.

Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 33-45). Hillsdale, NJ: Lawrence Erlbaum Associates.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp. 13-103). New York: Macmillan.

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

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

**HTML:** [http://jalt.org/test/bro\\_14.htm](http://jalt.org/test/bro_14.htm) / **PDF:** <http://jalt.org/test/PDF/Brown14.pdf>