

<p>Liying Cheng, Youyi Sun and Jia Ma (2015). <i>Review of washback research literature within Kane's argument-based validation framework</i>. Language Teaching, 48, pp 436-470 doi:10.1017/S0261444815000233</p>	<p>articles copied to or from</p>
<p>Awareness of the importance of testing consequences is not new; it has existed since the birth of modern testing (Latham 1877)</p>	<p>Awareness of the importance of testing consequences—the power of testing—is not new; it has existed since the birth of modern testing. As far back as 1877, Latham (1877, p. 1)</p> <p>Cheng: <i>Geopolitics of Assessment</i>, 2017</p>
<p>The prevalence of large-scale, high-stakes testing worldwide (Ungerleider 2003; Cheng 2008; Klinger, DeLuca &amp; Miller 2008) has given rise to growing attention and concerns over its impact on stakeholders, the educational system, and society at large. The result has been a fast growing body of research in education investigating intended and unintended relationships between testing, teaching, and learning. Given the range and extent of testing consequences reported worldwide, it is critical that testing practices yield valid data about student achievement and performance. p1</p>	<p>The prevalence of large-scale and high-stakes testing worldwide has given rise to growing attention and concern over its impact on stakeholders, the educational system, and society at large. The result has been a fast-growing body of research in education investigating relationships among testing, teaching, and learning (Cheng, 2014). Given the range and extent of testing consequences reported worldwide, it is critical that testing practices yield valid data about student achievement and performance.</p> <p>Cheng: <i>Geopolitics of Assessment</i>, 2017</p>
<p>Messick (1996) also points out that the consequences of tests are likely to be a function of factors both within the test itself and within the setting where the test is implemented. He recommends, from the test design perspective, minimizing the two threats to construct validity – construct under-representation and construct irrelevant variance – in order to enhance the quality of the test and thus promote positive washback p437</p>	<p>Messick also points out that the consequences of tests are likely to be a function of factors both within the test itself and within the setting of the test. He recommends the examination of the two threats to construct validity—construct under-representation and construct-irrelevant variance—in order to enhance the quality of the test and thus promote positive washback.</p> <p><b>Consequences, Impact, and Washback</b> The Companion to Language Assessment Volume III. Evaluation, Methodology, and Interdisciplinary Themes Part 9. Designing Evaluations Liying Cheng First published: 11 November 2013 <a href="https://doi.org/10.1002/9781118411360.wbcla071">https://doi.org/10.1002/9781118411360.wbcla071</a></p>
<p>Bailey (1996: 268), however, argues that any test, whether good or bad, can have either positive or negative washback depending on whether ‘it impedes or promotes the accomplishment of educational goals held by learners and/or program personnel’. She focuses on the specificity of this phenomenon, which could induce differential impact on test stakeholders within a range of teaching and learning contexts – a view increasingly shared by many researchers in language testing (e.g., Cheng 2008). p437</p>	<p>Bailey (1996, p. 268), however, argues that any test, whether good or bad (in terms of validity), can have either negative or positive washback, depending on whether “it impedes or promotes the accomplishment of educational goals held by learners and/or program personnel.” She focuses on the specificity of this phenomenon, which could induce differential impact on test stakeholders within a range of teaching and learning contexts—a view increasingly shared by many language testers (e.g., Cheng, 2008).</p> <p><b>Consequences, Impact, and Washback</b> ,The Companion to Language Assessment Volume III. Evaluation, Methodology, and Interdisciplinary Themes Part 9. Designing Evaluations Liying Cheng First published: 11 November 2013 <a href="https://doi.org/10.1002/9781118411360.wbcla071">https://doi.org/10.1002/9781118411360.wbcla071</a>November 2013 <a href="https://doi.org/10.1002/9781118411360.wbcla071">https://doi.org/10.1002/9781118411360.wbcla071</a></p>

<p>More recently, Bachman (2005) proposes a validity framework with a set of principles and procedures for linking test scores and score-based inferences to test use and the consequences of test use, an area in which, he argues, more research needs to be conducted. p437</p> <p>Bachman (2005), when he proposed a framework with a set of principles and procedures for linking test scores and score-based inferences to test uses and their consequences. p448</p>	<p>Bachman (2005) proposes a validity framework with a set of principles and procedures for linking test scores and score-based inferences to test use and the consequences of test use—an area in which he argues for more research to be conducted.</p> <p><b>Consequences, Impact, and Washback</b> ,The Companion to Language Assessment Volume III. Evaluation, Methodology, and Interdisciplinary Themes Part 9. Designing Evaluations Liying Cheng First published: 11 November 2013 <a href="https://doi.org/10.1002/9781118411360.wbcla071">https://doi.org/10.1002/9781118411360.wbcla071</a></p>
	<p>Bachman (2005) proposes a framework with a set of principles and procedures for linking test scores and score-based inferences to test use and the consequences of test use.</p> <p><b>Washback, Impact and Consequences</b> Language education in Europe: The Common European Framework of Reference (pp.2479-2494) Edition: 2nd DOI: 10.1007/978-0-387-30424-3_186 Publisher: Springer Science + Business Media LLC., Editors: Elana Shohamy, Nancy H. Hornberger</p>
	<p>...a set of principles and procedures for linking test scores and score-based inferences to test use and the consequences of test use.</p> <p><b>Building and Supporting a Case for Test Use</b> Lyle F. Bachman 2005</p>
<p>McNamara (2000) uses two terms to distinguish between two levels of this phenomenon: ‘impact’, the effects of tests on the macro-levels of education and society; and ‘washback’, the effects of language tests on micro-levels of language teaching and learning inside the classroom. In this sense, the difference between impact and washback resides in the scope of the effects of testing – a more narrow one of washback and the all-encompassing one of impact (Hamp-Lyons 1997). These two specific terms are used in this review as individual researchers use them in their studies. p.438</p>	<p>McNamara (2000), however, uses two terms to distinguish between two levels of this phenomenon: “impact”—the effects of tests on macro-levels of education and society, and “washback”—the effects of language tests on micro-levels of language teaching and learning inside the classroom. In this sense, the difference between impact and washback resides in the scope of the effects of testing—which gives us a view of test consequences falling between the more narrow one of washback and the all-encompassing one of impact (Hamp-Lyons, 1997). These two specific terms are discussed in this chapter as individual researchers use them.</p> <p><b>Consequences, Impact, and Washback</b> ,The Companion to Language Assessment Volume III. Evaluation, Methodology, and Interdisciplinary Themes Part 9. Designing Evaluations Liying Cheng First published: 11 November 2013 <a href="https://doi.org/10.1002/9781118411360.wbcla071">https://doi.org/10.1002/9781118411360.wbcla071</a></p>
<p>the IUA can be specified as a network of inferences and assumptions leading from the test performances to the score interpretations, and to any decisions of test uses. The decision inference links interpretations of the test score to decisions of score uses and is evaluated in terms of consequences, i.e., expected success in achieving goals at a reasonable cost and with acceptable consequences (Kane 2013)</p>	<p>The IUA can be specified as a network of inferences and assumptions leading from the test performances to the conclusions to be drawn and to any decisions based on these conclusions (Crooks, Kane, &amp; Cohen, 1996; Kane, 1992, 2006; Shepard, 1993). <b>Validating the Interpretations and Uses of Test Scores</b> (Kane, 2013)</p>