



Preparing Students to Take Tests

Journal:	<i>TESOL Encyclopedia of English Language Teaching</i>
Manuscript ID	EELT-0321.R3
Wiley - Manuscript type:	Entry
Date Submitted by the Author:	n/a
Complete List of Authors:	Ma, Jia; Queen's University, Faculty of Education Cheng, Liying; Queen's University, Faculty of Education
Keywords:	Language Teaching, assessment, evaluation
Free Text Keywords:	Test preparation, large-scale tests
Abstract:	<p>Test scores derived from large-scale testing are believed by many as an appropriate way to hold educators accountable for students' academic achievement (Nichols & Berliner, 2007). Test preparation has thus been widespread both inside and outside of schools and educational institutions aiming to increase students' test scores. However, the appropriateness of test preparation practices might be sacrificed for emphasis solely on students' score improvement. This article discusses test preparation from theoretical and empirical perspectives across a broad educational context, and suggests pedagogical principles for teachers to integrate appropriate test preparation practices into their classroom instructions.</p>

Preparing Students to Take Tests

Jia Ma & Liying Cheng

Queen's University

jia.ma@queensu.ca and liying.cheng@queensu.ca

Abstract

Test scores derived from large-scale testing are believed by many as an appropriate way to hold educators accountable for students' academic achievement. Test preparation has thus been widespread both inside and outside of schools and educational institutions aiming to increase students' test scores. However, the appropriateness of test preparation practices might be sacrificed for emphasis solely on students' score improvement. This entry discusses test preparation from theoretical and empirical perspectives across a broad educational context, and suggests pedagogical principles for teachers to integrate appropriate test preparation practices into their classroom instruction.

Word Count: 3,000

Large-scale testing has been increasingly used in educational systems across countries for high-stakes purposes of accountability, gatekeeping and policymaking (e.g., Cumming, 2009; Cheng, 2008). These tests, regardless of subject areas being tested and their test constructs, make inferences on students' proficiency based on their performance on a sample of items drawn from a whole content domain of knowledge (Madaus, 1988). The results of such inferences from students' test scores or test performance are used as indicators of students' academic achievement, and are often directly related to a variety of high-stakes decisions, from students' obtaining degree and academic advancement, to teachers' attaining professional certification and promotion, and also to school boards' gaining funds (Cumming, 2009; Madaus, 1988). Due to the snapshot nature of testing as indicators of students' academic achievement and also the relationship between high-stakes decision-making and students' test performance, large-scale testing has affected what and how teachers teach and especially what and how students learn. Consequently, "teaching has been inordinately *skewed toward test preparation*" (Madaus, 1988, p. 36, emphasis added), practices (pedagogy) and principles (appropriateness/ethicality) of preparing students to take tests has thus gained increasing attention in many fields, including curriculum, educational measurement and language assessment (e.g., Green, 2007; Haladyna & Downing, 2004; Popham, 2001; Powers, 1985) since mid-1980s.

Framing the Issue

Three key terms have been used to define preparing students to take tests—coaching, teaching to the test, and more commonly, test preparation. *Coaching* is applied to commercial programs operated as extracurricular activities that students participate in outside of school, and refers more to short-term instructions targeted to improve students' test performance on *a particular* examination and does not necessarily have the aim to improve their academic skills. Therefore, the term coaching usually has a negative connotation that students can be coached to maximize their test performance but may not have a corresponding increase in their academic abilities per se. Coaching is often used in the educational measurement field and examined whether and to what extent coaching might influence students' test scores. *Teaching to the test* is often used in school settings and is usually discussed in the curriculum literature. Similar to coaching, teaching to the test implies that teachers' instructions focus on actual test items with an aim to improve students' test scores, or simply item-teaching. However, *teaching to the test* can also mean that teachers build their instructions around curriculum objectives sampled by tests to enhance students' test performance by improving their knowledge of content domains, or curriculum-teaching (Popham, 2001). *Test preparation* can be applied to a broad context including both inside and outside school activities, and is the term used in the broad research field. It is more neutral and inclusive, and is defined as a variety of activities to review contents sampled by tests and to practice test-taking skills in an attempt to improve students' test scores (e.g., Crocker, 2006). As defined in these terms, the ultimate goal of test preparation is the same—to improve students' test scores. However, these terms differ slightly in focus depending on the contexts where they are used and studied.

Some researchers also use the delivery mode to understand and study test preparation. Test preparation can be school-based, commercial-based, computer-based, and book-based according to the delivery mode (e.g., Montgomery & Lilly, 2012). School-based test preparation is integrated into the curriculum and offered by classroom teachers within school settings. Commercial-based test preparation is fee-charged short-term instructions operated by commercial agencies with the claim of effectively increasing students' test scores. Computer-based test preparation is test preparation whose content is administered through a computer, where students have control of the speed, and the amount of test preparation they choose to engage in. Book-based test preparation is structured on commercial publishers' practice books and test publishers' official guides.

Making the Case

Starting in the 1950s, a number of educational researchers have investigated the effects of commercial test preparation courses (coaching) on test scores. Inspired by these early investigations, more researchers have attempted to unpack and reach an understanding of test preparation from their particular research perspectives. Test preparation has been studied in a range of three specific areas: (a) “teaching to the test” and “measurement-driven instruction” in the field of curriculum because such practice may narrow curricula (e.g., Madaus, 1988; Popham, 2001); (b) “test impact”, “construct-irrelevant variance”, and “consequential validity” in the field of educational measurement because of its effects on test scores and test score uses (e.g., Haladyna & Downing, 2004); and (c) “washback” in the field of language education and

testing because of its influence on language teaching and learning (e.g., Alderson & Hamp-Lyons, 1996; Cheng & DeLuca, 2011; Green, 2007). Studies on test preparation across fields were derived from a common concern—the influences of test preparation on the accuracy of students' test scores as an indicator of their academic competency, or theoretically speaking, the influences of test preparation on the validity of test scores.

Validity refers to the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores. Simply put, if a plausible interpretation of a student's mastery level can be derived from the test score, the validity of the test score is achieved. However, if a test score is not the actual representation of a student's mastery level of a content domain, the interpretation and inference based on this test score cannot be accurate; therefore, the validity of this test score is threatened. With differences on its focus of instructional contents, test preparation thus might threaten the validity of test scores. Messick (1996) has stated that if test preparation emphasizes the instruction of test-wiseness strategies, students might answer some test items correctly using test-wiseness strategies rather than their actual knowledge. In this case, students' increased scores cannot represent the equivalent improvement of students' knowledge, and this type of test preparation threatens the interpretation and use of the increased test scores. Meanwhile, if students are prepared with test familiarization and solutions of dealing with psychological influences (e.g., anxiety reduction), they might better perform their mastery of knowledge with reduced levels of influences on their performance. This type of test preparation thus improves the validity of test scores as such a practice minimizes the construct irrelevant variance.

Although studies on test preparation cover many dimensions, including practices, effects, and principles, how test preparation influences students' test scores draws immediate interest. Since the 1950s, a large body of educational measurement studies has examined high-school students' test score gains resulting from coaching programs (e.g., Montgomery & Lilly, 2012; Powers, 1985). The results of these studies have shown that coaching programs increase student test scores by 20 to 30 points on vocabulary and math subtests of the Scholastic Aptitude Test (SAT) (Montgomery & Lilly, 2012). Studies in language education and testing investigate this issue in a slightly different way; instead of measuring effect sizes in the unit of score gains, these studies looked at whether students who take test preparation courses perform significantly better in comparison to students who do not. These studies have shown conflicting results: (a) significantly better performance of students taking test preparation courses (e.g., Hayes & Read, 2004); and (b) no significant advantage over students taking test preparation courses (e.g., Green, 2007). Therefore, whether test preparation can significantly influence test scores is still under debate and needs further empirical explorations (Montgomery & Lilly, 2012).

Pedagogical Implications

Although the degree to which test preparation influences test scores is still under exploration, it is a shared understanding among researchers across educational fields that a criteria or a code of practice should be established to ensure the appropriateness or ethicality of test preparation practices, practically speaking, to enable teachers to perform preparation activities appropriate for students' improvement in knowledge and the ability of a content domain. Since the 1980s, educational researchers have proposed principles to examine test preparation appropriateness in

two strands: theoretical-based and practice-based (Table 1).

Theoretical-based principles can fall into two categories: (a) to examine a continuum of test preparation activities ranging from ethical to unethical behavior (Mehrens & Kaminski, 1989), and (b) to create sets of specific standards for evaluating the appropriateness of these activities, such as professional ethics (Popham, 1991) and educational value (Crocker, 2006). It can be seen from this table below that the theoretical-based principles have been developed to include more concrete dimensions to evaluate test preparation activities since the 1980s. Crocker's (2006) specifies the following criteria: *validity* requires that test preparation improves the validity of test score interpretation; *academic ethics* requires test preparation activities to be consistent with ethical standards of the educational profession; *fairness* means all test takers should have equal access to preparation opportunities; *educational value* means test preparation should improve both test takers' scores and their content knowledge; *transferability* requires test preparation to teach test takers skills that can be used in different examination situations.

One concern of these theoretical-based principles is practicality. Classroom teachers still found these standards too general to follow in judging their own preparation activity (Popham, 1991). Therefore, practical-based principles of evaluating test preparation practices have been proposed to help teachers focus their test preparation practices on curriculum instruction rather than test items. Turner (2009) identified five types of test preparation practices that can benefit learning (Table 1): (a) teaching to the content domain covered by the curriculum, (b) using a variety of assessments, (c) reviewing/teaching test-taking strategies, (d) promoting students' motivation, and (e) managing time and frequency of test preparation. They further suggested lists of practical teaching activities for each type of test preparation to ensure appropriateness. For example, Turner (2009) suggested teachers could create opportunities for students to present their understanding of the same content knowledge in different forms and contexts such as independent work, oral presentations, and written essays. When teaching students test-taking skills, teachers might review previous years' test papers to analyze task requirements, and help students be familiar with what aspect of knowledge they are required to know and understand, and in what task format (e.g., write an argument essay) and how they are required to complete test tasks (e.g., write an essay in several steps, including forming an argument, brainstorming supportive evidence, outlining structure). When planning a timeline for test preparation activities, teachers are suggested to consider regular intervals (e.g., bi-weekly or monthly) throughout the whole school year to schedule test review activities and also in the weeks approaching the test.

Table 1.

Principles of evaluating test preparation practices synthesized in the literature

Category	Principles
	<ul style="list-style-type: none"> • A continuum from ethical to unethical behaviours (Mehrens & Kaminski, 1989)
	<ul style="list-style-type: none"> • Professional ethics • Educational defensibility (Popham, 1991)
	<ul style="list-style-type: none"> • Validity • Academic ethics

Theoretical-based	<ul style="list-style-type: none"> • Fairness • Educational value • Transferability (Crocker, 2006)
Practice-based	<ul style="list-style-type: none"> • Including curriculum objectives; integrating test content into curriculum • Familiarizing students with various assessment approaches; • Instructing and reviewing test-taking strategies; • Managing time and frequency of test preparation throughout the year • Motivating students to do their best on tests (Turner, 2009)

Teachers could also align their test preparation practices with both the theoretical-based and the practice-based principles to judge the appropriateness of their own test preparation practices. For example, teachers often instruct students to manage their time for test tasks as one common test-taking strategy to ensure test tasks completion within a designated time limit. This test preparation activity can reduce the possibility of having insufficient time to complete test tasks due to not managing time well. This particular activity can help reduce the construct irrelevant variance (in this case, insufficient time) that is probably unrelated to students' academic competence but can negatively influence students' test scores. This test preparation activity conforms to Crocker's (2006) criteria of validity because it increases the plausibility of interpreting students' test scores and validity of students' test scores. Teaching content domains (e.g., specific language skills like reading, listening etc.) that are sampled by tests (e.g., language proficiency tests) is aligned with the criterion of educational value or educational defensibility (Crocker 2006; Popham, 1991), because it reduces the concern that test preparation practices might be limited to the contents (e.g., test items) that appear on tests. In addition, teachers can use empirical evidence as legitimate resources of performing appropriate test preparation activities. It has been found that students preferred teachers' diagnosis of their weakness in specific language skills (e.g., pronunciation in speaking, vocabulary in reading and writing etc.) and needed opportunities for participation, questioning and practicing language skills in test preparation classes as they do in their regular class (e.g., Alderson & Hamp-Lyons, 1996). Students taking test preparation courses believed teachers' instructions aiming for improving their English competence could contribute to their better test performance (Wang, Ma, & Cheng, 2013). These test preparation activities emphasizing general English competence improvement rather than focusing on coaching test items alone were shown to be associated with students' higher test scores on English language proficiency test. To conclude, more ongoing empirical investigations will shed more light on pedagogical implications that enable teachers to perform appropriate test preparation activities.

SEE ALSO: Test impact & washback; Ethics; Test anxiety; Accommodations

References

- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing, 13*(3), 280-297.
- Bangert-Drowns, R. L., Kulik, J. A., Kulik, C. L. (1983). Effects of coaching programs on achievement test performance. *Review of Educational Research, 53*, 571-585.
- Cheng, L. (2008). The key to success: English language testing in China. *Language Testing, 25*, 15-37.
- Crocker, L. (2006). Preparing examinees for test taking: Guidelines for test developers and test users. In S. M. Downing, & T. M. Haladyna, (Eds.), *Handbook of Test Development* (pp. 115–128). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cumming, A. (2009). Language assessment in education: Tests, curricula, and teaching. *Annual Review of Applied Linguistics, 29*, 90-100.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practices, 23*, 17-27.
- 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: Research contexts and methods* (pp. 97–112). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Madaus, G. F. (1988). The Distortion of Teaching and Testing: High-Stakes Testing and Instruction. *Peabody Journal of Education, 65*(3), 29-46.
- Mehrens, W. A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practices, 8*(1), 14-22.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*, 241- 256.
- Montgomery, P., & Lilly, J. (2012). Systematic reviews of the effects of preparatory courses on university entrance examinations in high school-age students. *International Journal of Social Welfare, 21*, 3-12.
- Popham, W. J. (1991). Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice, 10*(4), 12-15.
- Popham, W. J. (2001). Teaching to the test? *Educational Leadership, 58*(6), 16-20.
- Powers, D. (1985). Effects of test preparation on the validity of a graduate admission test. *Applied Psychological Measurement, 9*, 179-190.
- Turner, S. L. (2009). Ethical and appropriate high-stakes test preparation in middle school: Five methods that matter. *Middle School Journal, 41*(1), 36-45.
- Wang, H., Ma, J., & Cheng, L. (2013). *English language learning through TOEFL iBT test preparation: A case of a test preparation centre in China*. Paper presentation at TESOL 2013, Dallas, Texas, United States.

Further Readings

- Cheng, L. & DeLuca, C. (2011). Voices from test-takers: Further evidence for language assessment validation and use. *Educational Assessment, 16*(2), 104-122.
- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education, 14*, 75-97.