Test takers' perspectives on an English language test in Iranian higher education: A washback study

Kioumars Razavipour, Mahboobeh Mansoori and Zohreh Gooniband Shooshtari
Shahid Chamran University of Ahvaz, Iran

Over the last three decades, the study of test preparation and test washback has emerged as an indispensable area of inquiry in language assessment. Yet, how test takers' motivation and perceptions of test design and content might mediate test preparation has not been given sufficient attention. Taking the general English module of a high stakes university admission test as its focus, this study explored how test takers' motivation and perceptions of test uses and design are related to their test preparation practices. To this end, 110 test takers responded to two questionnaires: one on test takers' motivation and perceptions, and the other on their preparation practices. The collected data were analysed using partial least squares structural *equation modeling* (PLS-SEM) to assess the postulated washback models, informed by *expectancy value* (EV) motivation theory. It was found that positive perceptions of test content were associated with more intensive test preparation. Also, test takers' perceptions of test uses were found to be related to the value they placed on test taking. By contrast, perceptions of test uses and test value did not significantly contribute to test preparation. In general, a test washback model with direct paths from test takers' perceptions to test preparation appeared to account for more variation in test preparation than one without such direct paths. Findings imply that given the complexity of test preparation and washback, theories of motivation must be complemented by broader social considerations to explain test preparation and test washback.

Introduction

Stakeholders' mental representations of educational events and environments are powerful predictors of the success or otherwise of education. This is why the failure of many educational innovations is attributed to teachers' false representations of the innovations (Wall, 1996). Likewise, students' conceptions of instructional environments have been shown to be a more powerful predictor of their learning than the actual instructional events (Brown & Hirschfeld, 2008, p. 3). Assessment as a component of education is bound to be of varying representations to students and test takers. Such representations of assessments in turn influence learners' approaches to learning in general and test preparation in particular (Xie & Andrews, 2013).

From an educational measurement perspective, learners' perceptions of assessments are of paramount importance because they are likely to compromise construct and consequential facets of test validity (Xie, 2011). This is so because a mismatch between students' perceptions of the skills and abilities intended to be measured by the test and the intentions of test designers can potentially undermine the intended positive impact envisioned by test designers and educational policy makers. This is more of a concern where the sole function or one of the functions of high stakes tests is to engineer educational reform. As a case in point, Qi (2005) found that different perceptions of test

use undermined the intended washback of the NMET (National Matriculation English Test) in China.

In the literature on washback, aside from a few studies (Xie, 2011; Xie, 2015; Xie & Andrews, 2013), test takers' perceptions of language tests have not received much attention. The three above-mentioned studies were carried out in relation to the *College English Test* (CET), used for exit purposes in Chinese universities. The current study seeks to extend research on test takers' perceptions of test demands and uses in relation to the washback of an annual national admission test for entering graduate English language programs in Iran, locally known as the *MA Language Test* (MALT). For details on the design and functions of MALT, interested readers are referred to Razavipour, Gooniband Shooshtari and Mansoori (2018).

Another motivation for this study is that most studies on test washback have thus far been qualitative (Xi & Andrews, 2013); hence, there is a need for theory-informed research. In this study, we drew on *expectancy-value* (EV) theory of motivation (more on EV theory below), which in the context of test taking, posits that test takers' perceptions of the test uses and demands affect their motivation and their test preparation strategies. Following Xie and Andrews (2013), two theoretical models were postulated and their adequacy was assessed using *partial least squares structural equation modeling* (PLS-SEM) , which makes less stringent assumptions about the data and the sample size (Ravand & Baghaei, 2016) in comparison with covariance-based SEM, which requires rather larger samples. Given our small sample size, we deemed PLS-SEM more appropriate.

We should note that the terms test washback and test preparation, though closely related, are not exact synonyms, as test washback is wider in domain than test preparation. In addition, test preparation does not necessarily imply test washback, as a test is only one possible cause for test preparation.

Individual factors mediating test washback

Test washback is generally defined as "the extent to which a test influences language teachers and learners to do things they would not necessarily otherwise do" (Alderson & Wall, 1993, cited in Messick, 1996, p. 241). Yet, the relationship between tests and instruction is not linear. Rather, washback is considered to be a complex, multi-faceted phenomenon interacting with and being mediated by social, institutional, and personal factors (Watanabe, 2004). Alderson and Wall (1993) foresaw the relevance of personal motivational factors to washback studies decades ago. However, in research into washback, it is often the teacher factors that are focused upon. As such, many teacher factors such as teachers' attributions (Watanabe, 2004), departmental status (Anagnostopoulos, 2003), and beliefs about the subject matter (Smagorinsky, Lakly & Johnson, 2002) have been investigated in connection with test washback.

Concerning learners, factors such as learners' possible selves (Zhan & Andrews, 2014), self-conceptions and attitudes (Haggerty, 2010), students' conceptions of assessment (Brown & Hirschfeld, 2008), learners' expectations (Green, 2007), assessment literacy

(Smith, Worsfold, Davies, Fisher & McPhail, 2013), as well as family income and level of education (Buchmann, Condron & Roscigno, 2010) have been studies in connection with test preparation and washback. Using a diary study, Zhan and Andrews investigated the test preparation practices of three Chinese students preparing for CET and found that students' test preparation was related to "their imagined possible CET selves" (p. 71). In a quantitative study, Haggerty (2010) administered a questionnaire to 341 Korean students in a private language institute and found that positive attitudes and beliefs towards the second language (L2) community and L2 testing were associated with more intense test preparation. Brown and Hirschfeld (2008) used exploratory and confirmatory factor analyses to build a scale of students' conceptions of assessment, and they found that the conception that assessment makes students accountable was positively correlated with achievement. Moreover, adopting an experimental design, Smith and colleagues (2013) discovered that test takers' higher levels of assessment literacy impact positively on their performance. Moreover, the kind and intensity of test preparation have been found to be related to the socioeconomic status of test takers and their parents' level of education (Buchmann et al., 2010) . Finally, Razavipour et al. (2018) found that test takers' perceptions of test design and demands do not vary significantly across institutional contexts.

Though the studies reviewed above have contributed each in their own way to our understanding of the role of individual factors in mediating test washback, research on relationships between test takers' perceptions of assessment and washback remains rather limited.

Perceptions of assessments and washback

There has been a recent surge of interest in investigating test takers' perceptions of the content and uses of tests and the subsequent influences that such perceptions have on language learning and test preparation. Sato and Ikeda (2015) found discrepancies between test takers' perceptions of what the test items measure and what they are intended to measure. Hsieh in Taiwan found positive perceptions of the uses of TOEIC as a graduation requirement. In Vietnam, Nguyen and Gu (2020) reported perceived negative impact of TOEIC as an exit requirement. Likewise, Im and Cheng (2019) found discrepancies between employees' perceptions of TOEIC content and those of the designers. Similar findings about TOEFL have been reported in Indonesia and Canada (Karjo & Ronaldo, 2018; Fox & Cheng, 2015).

Xie & Andrews (2013) found that instrumentally motivated test takers are more likely to improve on the intended construct of the test. Likewise, Xie and Andrews (2013) concluded that test takers with higher self-efficacy and those advocating higher task values engage in more intense test preparation. Similarly, it has been found that "favorable perceptions of test validity" are associated with "a higher level of engagement in both desirable language learning activities and focused test preparation" (Xie, 2015, p. 57). Xie concluded that test takers' positive perceptions of test validity foster positive washback but do not necessarily reduce negative washback. Using factorial analysis in designing a measure of test impact on learners, Samaie and Mohammadi (2017) found that prior

awareness of the test design and format constitutes a component of a five-factor scale of test impact. Finally, in the context of IELTS preparation courses, Green (2007) found that test takers had various perceptions of test demands, and such perceptions were more important than the course type or the course content in determining outcomes.

Most washback studies done thus far have used a qualitative design and one shortcoming in qualitative washback studies is that they are not informed by a substantive theory (Xie, 2011). Thus, to systematically establish a liaison between test takers' perceptions of test content and uses on the one hand and test washback on the other, we need a comprehensive conceptual model on the basis of which hypotheses can be generated, and paths of influences from test takers' perceptions to test influences can be postulated. Expectancy-value (EV) theory (Wigfield & Eccles, 2000) has proved a useful heuristic in investigating learning for tests (see Xie & Andrews, 2013).

In essence, the EV theory comes down to two basic questions people ask of themselves when they consider doing a task: do I want to do it? And if so, can I do it? The first question captures the value dimension and the latter the expectancy dimension or self-efficacy. In other words, the degree of willingness to do something and the self-efficacy one feels in doing it are associated with more success. EV theory posits that "individuals' choice, persistence, and performance can be explained by their beliefs about how well they will do on the activity and the extent to which they value the activity" (Wigfield & Eccles, 2000, p. 68). Accordingly, the choices that people make and their performance on the chosen tasks, *achievement–related choices and performance*, (the middle rectangle on the right part in Figure 1) depend on their beliefs about their ability to do the task. These ability beliefs are called *expectation of success* in the EV theory. In fact, the construct is close in meaning to the construct of self-efficacy (Bandura, 1977). The difference is that the former is about the expected outcome while self-efficacy is concerned with the individual's overall confidence towards the chosen task.

Furthermore, performance on a task also depends on the extent the task is valued by the individual (Wigfield & Eccles, 2000), which corresponds to *subjective task value* in Figure 1. According to Wigfield and Eccles (2000), the value attached to a task is determined by *incentive value*, *attainment value*, *utility value*, and *costs*. The theory also posits that *expectation of success* and *subjective task value* are influenced by individuals' *goals* and *perceptions of task demands*, respectively (the left cell in Figure 1). Thus, under *child's goals and general self-schemata*, the three factors of *self-schemata*, *short-term goals*, and *long-term goals* affect individuals' evaluation of their chances of success (*expectation of success*) as well as the value they attach to the task (*subjective task value*), which would in turn affect performance and achievement.

Taking test preparation as a special type of learning, the correspondence between the EV theory and how washback to the learner takes place is as follows. Perceptions of test use are test takers' short-term g*oals*, affecting the *subjective task value* they accord to success on the test, which would in turn determine their extent and type of test preparation. In a sense, goals are in interaction with the "level of the stakes" of a test (Hamp-Lyons, 1998, p. 329). High stakes tests drive test takers to set themselves more important goals

associated with the test. This would increase the value they attach to the test (*subjective task value*), leading to expending more effort and time in test preparation; hence, more intense washback.

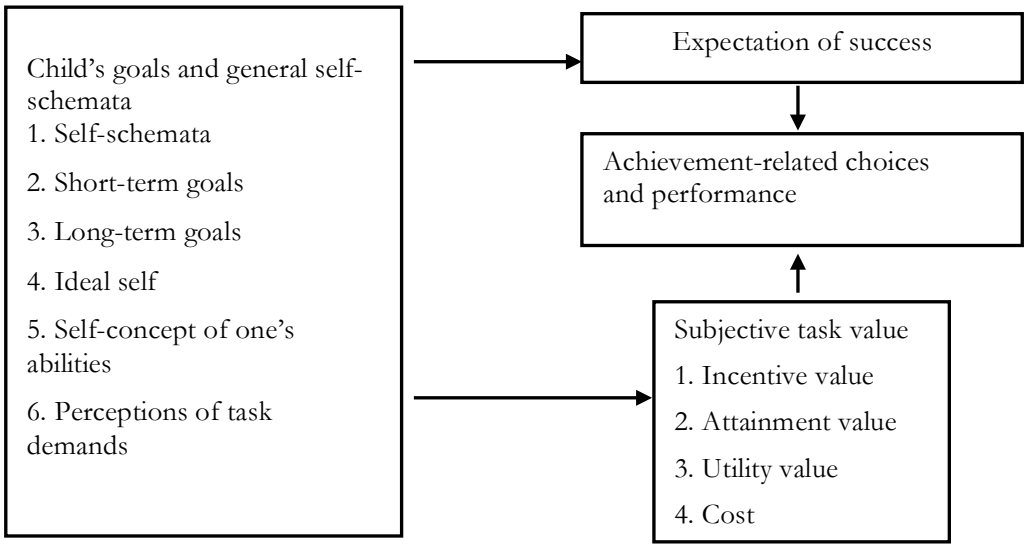

Figure 1: Expectancy-value model of achievement motivation (taken from Wigfield & Eccles, 2000)

Perceptions of task demands are the reality test takers construct for themselves of the test content and of what preparation for the test entails. In other words, test takers' subjective construal of test content and of the knowledge and skills they deem essential for optimal performance on the test constitutes their perceptions of test content and demands. Test takers' judgments of test demands would affect their expectation of success. For instance, if they estimate, based on their beliefs about their ability and about the difficulty of the test, that the test is too easy or too challenging, this would lower their *expectation of success*, and in turn their engagement in test preparation.

Following Xie and Andrews (2013), we postulated two conceptual washback models (Figures 2 and 3). Each model consists of four reflective latent variables and the formative variable, test preparation (TP). In a formative variable, the indicators cause variation in their corresponding construct not the other way around, as is the case with reflective variables (Henseler et al., 2014). Test takers' perceptions of test use (PTU) and their perceptions of test content (PTC) are exogenous latent variables. The endogenous variables include test value, expectation of success (ES), and test preparation (TP).

Figure 2 represents the washback model based entirely on the EV theory (Model 1 henceforth), with no direct paths of influence from PTU and PTC to TP. According to the EV theory, goals and perceptions of task demands affect the value one attaches to the task and the expectation of success on the task; values and expectations would in turn affect one's choices, persistence, and strategies towards the learning task. Accordingly, the

EV theory predicts no paths of influence from goals and task demands (PTU and PTC in our case) to learning (test preparation).

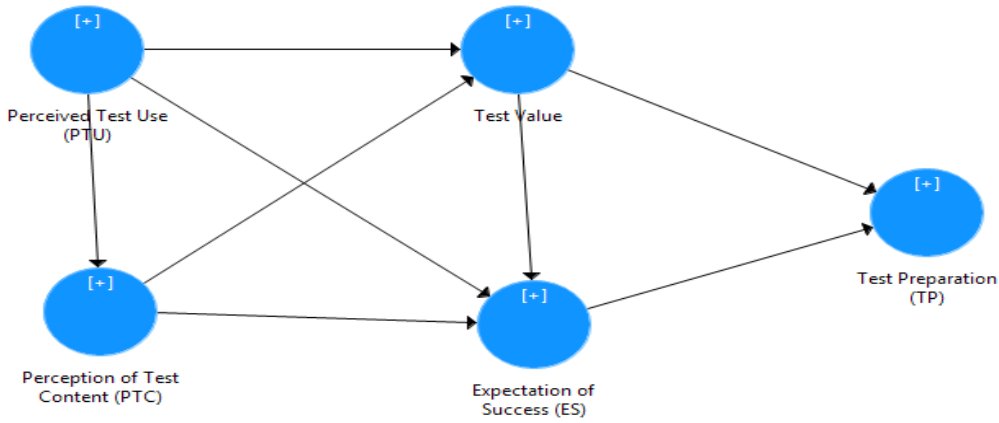

Figure 2: The proposed Model 1 of washback without direct paths from PTC and PTU to TP

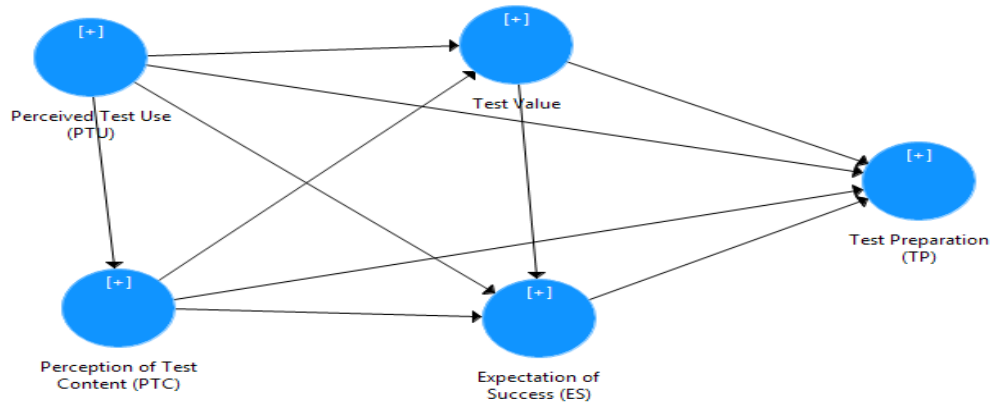

Figure 3: The proposed Model 2 of washback with direct paths from PTU and PTC to TP

Figure 3 demonstrates the model of washback according to which PTC and PTU influence TP both directly and indirectly, through test value and expectancy (Model 2, henceforth).

The significance of this study lies in its potential to bring together a number of factors contributing to test washback, which have thus far been researched in isolation from one another. This allows for the possibility of examining the interaction and path directions between and among different factors. The causality interpretation which is made possible using SEM has the potential to shed further light on the complex nature of test washback

by allowing us to know whether perceptions of test content and uses directly affect test preparation or whether the influence of perceived test uses and test content perceptions are moderated by intermediate factors of values and expectancy in the models (Figures 2 and 3). Within this spirit, the current study addresses the following questions:

1. Do test takers' perceptions of MALT content and uses have a direct significant impact on their test preparation?
2. Are the paths of influence from test-takers' perceptions of test content and uses to test preparation mediated by test takers' test value and their expectations of success?

Method

Participants

The data for this study was collected in June 2016. A total of 170 English major students contributed data to the current study. They were conveniently sampled from three national universities based in Ahvaz, a metropolis in the Southwestern province of Khouzestan, Iran. Participation was voluntary and the participants were assured of the confidentiality of the data. Sixty participants were recruited for the pilot phase of the study. Of the remaining 110 participants who completed the finalised questionnaires, 65 were females and 45 were males, a sample commensurate with the general nature of student population in the country's often female-dominated language departments. The participants of the main study were senior undergraduates, as well as graduate students who had already been admitted to graduate programs through MALT. The latter group was included for the retrospective account they could provide about their conceptions of and preparation for MALT.

Instrumentation and analysis

This study utilised two questionnaires to capture test-takers' perceptions of test content and functions as well as their preparation strategies. To construct the questionnaires, we drew on three sources. We first interviewed 11 MALT applicants for their views on the test's content and uses, as well as their preparation practices. Second, we benefited from the two instruments developed by Xie and Andrews (2013) for a similar purpose. Third, our own experience, as two former MALT test takers, was also helpful in deciding on appropriate items for the preparation questionnaire. For data collection, we used two Likert scale questionnaires for quantifying test perceptions and preparation.

The perception questionnaire was intended to elicit data about test takers' ideas about the content and uses of the MALT (see Appendix 2). Items on MALT content were intended to tap into test takers' perceptions of the knowledge and skill areas that they thought MALT seeks to measure. More specifically, the participants were asked about what knowledge and thinking processes were required to do well on different sections of the MALT (e.g., grammar, vocabulary, and reading comprehension).

With regard to the MALT uses, questionnaire items were intended to capture the two dimensions of EV theory, expectancy and value. As such, to gather information on their self-efficacy and expectation of success on MALT, they were asked how confident they were about success on the test, and about their perceived difficulty of MALT. Regarding the value construct, we asked the participants about their motivation for taking MALT, what benefits they would get from taking MALT, and how important success on MALT was for them.

The MALT preparation questionnaire (see Appendix 1), consisting of 61 Likert type items, captured test takers' test preparation practices in four subscales, namely, test analysis and evaluation (14 items), test-taking skills (12 items), drilling target skills (26 items), and socio-affective strategies (9 items). For PLS-SEM analysis, composite scores were computed by averaging item scores constituting each subscale. Thus, in the PLS-SEM model, test preparation was measured with four indicators, which were the computed composite scores for each of the four subscales. This was done because unlike reflective measurement models, in PLS-SEM there are limits on the number of indicators allowed for formatively measured models. According to Henseler et al. (2014), "formative measurement has an inherent limit to the number of indicators that can retain a statistically significant weight. The maximum possible outer weight is $1/\sqrt{n}$, where n is the number of indicators" (p. 128). As such, including more than five items in formative models would render insignificant outer loadings (for further information, see the Appendix). For data analysis, we used partial least squares structural equation modeling (PLS-SEM). The choice was made for two reasons. First, due to the relatively small sample, the data did not satisfy the assumptions of mainstream SEM analysis and secondly, PLS-SEM is better suited to exploratory purposes (Hair, Hult, Ringle & Sarstedt, 2017). The software *SmartPLS 3* was used for data analysis. In this study, test preparation was measured formatively, as it is difficult to conceptualise it as a trait with a causal relationship to its indicators. Rather, it is more plausible to consider test preparation as a trait being shaped by a set of manifest variables.

Results

This study was designed to investigate how test takers' construal of MALT content and utility shapes the kind and degree of washback it exerts on their test preparation practices. To that end, two models, based on expectancy-value theory, were proposed (Figures 2 and 3), assessed, and compared via PLS-SEM.

Model evaluation in PLS-SEM is carried out in two stages: measurement model evaluation and structural model evaluation. Table 1 summarises the evaluative criteria in each stage.

In the following section, results for measurement model evaluations are presented first. Afterwards, we report the results for structural model assessment.

Table 1: Evaluation of PLS-SEM results (adapted from Henseler et al., 2014)

Evaluation of measurement models	Internal consistency (composite reliability) Indicator reliability Convergent validity (average variance extracted) Discriminant validity
Evaluation of the structural model	Coefficient of determination (R^2) Predictive relevance (Q^2) Size and significance of path coefficients f^2 effect sizes q^2 effect sizes (not appropriate with models including formatively measured constructs, according to Garson, 2016)

Assessing measurement models

To assess internal consistency, composite reliability is preferred to Cronbach's alpha, which has been criticised for its unrealistic assumptions, resulting in underestimation of reliability (Henseler et. al., 2014; Henseler & Sarstedt, 2013). Instead, composite reliability is considered a more robust index of reliability, thanks to its more realistic assumptions. Table 2 indicates that all composite reliability coefficients are above the acceptable .7 value. To evaluate the convergent validity of the outer models, average variance extracted (AVE) and outer loadings must be checked. AVEs of .5 suggest that at least half of the variance in the manifest variables is being explained (Ravand & Baghayi, 2016). Accordingly, the AVEs for ES (.446), PTU (.535), and value (.664) are above or close to the required level. It is relatively low for PTD (.337) and quite low for TP (.197). The lower index of TP is expected because preparation for language tests cannot be operationally defined as a unidimensional construct, since it entails a diverse number of sub-constructs such as cognitive strategies, test wiseness, and socio-effective test preparation strategies (Hawkey, 2006, pp. 54-58). It is partly for this same reason that test preparation was defined as a formative measure in this study.

Table 2: Reliability and convergent validity of the measurement models

	Cronbach's alpha	rho_A	Composite reliability	Average variance extracted (AVE)
Expectation of success (ES)	0.731	0.809	0.810	0.446
Perception of test content (PTC)	0.671	0.697	0.773	0.337
Perception of test use (PTU)	0.580	0.620	0.774	0.535
Test preparation (TP)	0.895	0.913	0.907	0.197
Test value (TV)	0.522	0.633	0.795	0.664

Discriminant validity is another criterion in evaluating measurement models in PLS-SEM. Consistent with the logic of the multi-trait multi-method approach (Campbell & Fiske, 1959), the AVE of each construct must be higher than its correlation with any other construct in the model. Comparing the AVEs in Table 2 with cross correlations in Table 3 (see Appendix 3) shows that the constructs discriminate well.

Table 3 (see Appendix 3) shows that for the ES construct, all indicator loadings are obviously high. For PTC, however, the loadings are below the suggested cut-off point. Yet, in social sciences, it is suggested that indicators with outer loadings above .4 be retained unless their elimination results in significant increases in composite reliability, which was not the case for PTC indicators. For PTU, except for PTU3 (.607), outer loadings for other items were quite high. Indicators for the value construct also loaded high on their corresponding construct. Finally, those for test preparation were moderate to high, though as a formative measure such criteria of evaluation do not apply to TP. Overall, it seems that there is sufficient support for the discriminant validity of the reflective constructs of ES, PTU, PTC, and TV as no traitor indicator is in evidence and all items have their higher loadings on their corresponding constructs.

A further step in measurement model evaluation is examining the multicollinearity of the indicators. The variance inflation factor (VIF) values on the far right column in Table 3 show that all VIF values are below the threshold value of 5; hence, no serious multicollinearity issue for path model estimation.

Another consideration in assessing discriminant validity is the Fornell-Larcher criterion, according to which, the AVE of every construct must be larger than its correlation with any other construct in the model (Garson, 2016). Meeting this requirement indicates that the construct shares more variance with its associated indicators than with any other construct.

Table 4: Discriminant validity: Fornell-Larcker criterion

	Expectation of success	Perceived test use	Perception of test content	Test preparation	Test value
Expectation of success	**0.809**				
Perceived test use	0.169	**0.730**			
Perception of test content	-0.174	0.027	**0.615**		
Test preparation	0.262	0.188	0.399	Formative measure	
Test value	0.249	0.340	0.192	0.254	**0.817**

In Table 4, the AVEs are the values on the diagonal, which, according to Fornell-Larcker criterion, must be larger than all the values below them in the same column. The four AVEs (in bold font) are larger than the correlations of constructs with each other, attesting to the distinctness of the constructs, hence discriminant validity.

Up to this point, the reflectively measured models of ES, PTC, PTU, and Test Value were assessed. However, the test preparation measurement model was not discussed. The criteria used in PLS-SEM to evaluate formative measures are different from those used for reflective measurement models because the core logic of internal consistency among indicators of the same traits does not hold with formative models (Garson, 2016). According to Henseler et al. (2014), to assess formative models, four main characteristics must be investigated: content validity, convergent validity with another similar measure,

multicollinearity, and the significance as well as the relevance of the indicators. To maximise content representation, in addition to reviewing the literature, we conducted in-depth interviews with eleven participants to capture a comprehensive picture of all the possible test preparation practices. Concerning the second criterion, it was literally impossible to think of an item that could represent the entirety of the construct of test preparation practices. Furthermore, expert opinion regarding the relevance and adequacy of the content was sought. Finally, multicollinearity indexes were examined for each indicator of the construct of test preparation practices (see Appendix 3) and the VIF values were below the threshold level.

Assessing the structural models

In PLS-SEM, structural model evaluation is done using the criteria mentioned in Table 1. Unlike CBS-SEM, bootstrapping and blindfolding procedures are used to generate heuristic criteria for assessing structural models in PLS-SEM (see Henseler et. al., 2014; Ravand & Baghayi, 2016). In the remaining of this section, results from each of the noted criteria are presented for Model 1 and Model 2, respectively.

To make sure the estimated path coefficients are not biased, multicollinearity must first be checked prior to PLS-SEM model evaluation. All the VIF values of the four endogenous variables (PTC, PTU, ES, and Test Value) were found to be far below the threshold value of 5 (see Table 3 in Appendix 3).

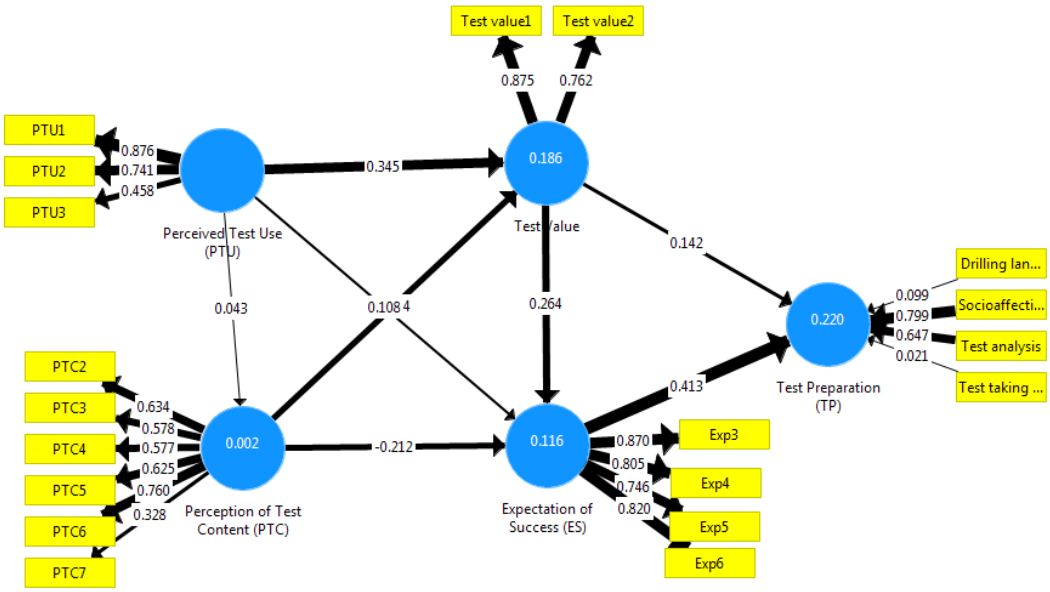

Figure 4: Washback Model 1 with path coefficients and R square values
(use PDF reader 'zoom in' function to read text)

Assessing Model 1

Figure 4 illustrates the outcome of PLS-SEM algorithm for Model 1 with path coefficients and R^2 values. The thickness of the arrows is commensurate with the strength of the relationship between constructs. Consistent with the EV theory, the strongest coefficient is between ES and TP (.41), followed by the one between PTU and Test Value (.345). In contrast, the weakest paths are PTU to PTC (.043), and PTU to ES (.108), which are rather at odds with the EV theory of learning.

The values in blue circles are R^2 values. The R^2 value from PTU to PTC is conspicuously low (.002), indicating that test takers' goals in taking MALT may have little to do with their knowledge of test demands, perhaps due to the 'diploma disease' that is very common in the country. People pursue academic degrees regardless of what it takes.

Since we seek to explain variation in MALT washback (test preparation), the R^2 value that is of most interest to us is that for test preparation (TP), which is .22 (see Figure 4). According to Henseler et al. (2014), "R^2 values of 0.25, 0.50, and 0.75 for target constructs are considered as weak, medium, and substantial" (p. 198). Therefore, Model 1 fails to explain a considerable amount of variation in MALT washback.

To decide whether individual parameter estimates are of statistical significance, we used a nonparametric bootstrapping procedure to generate p-values or the f^2 effect sizes.

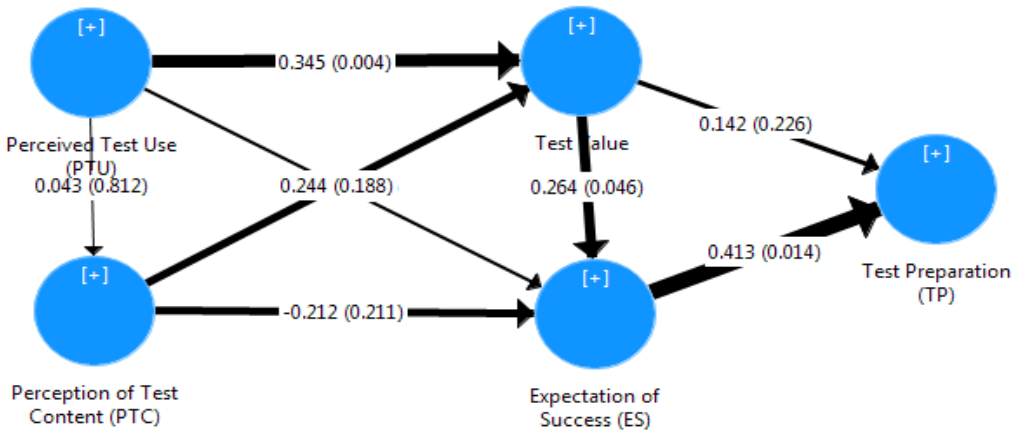

Figure 5: f^2 effect sizes for Model 1

Figure 5 gives the f-square effect size indices, values in parentheses on the paths. According to Garson (2016), f^2 is a measure of the change in R^2 if a certain variable is dropped from the model. Accordingly, PTC (p=.004), ES (p=.014) and test value (p=.046) are the only constructs in the model that, if eliminated, would lead to significant changes in the magnitude of the variance explained. The dropping of the other constructs in the model does not cause significant reduction in the variance explained by the model. These values indicate that for MALT candidates, high stakes uses of the test significantly

contribute to the value they attach to the test. Test value significantly affects their test self-efficacy, which in turn affects their test preparation practices.

Assessing Model 2

Figure 6 demonstrates the path coefficients and the R^2 values for Model 2, which unlike Model 1, entails direct paths from PTC and PTU to test preparation. The values in the circles are R^2 values and the strength of relationships is indicated by the thickness of the paths. As such, the strongest path is from PTC to Test Preparation, followed by the path from PTU to test value. In other words, test takers' perceptions of test content predicts the largest amount of variation in test preparation practices geared at MALT, and the extent they value the MALT is more determined by their perceptions of test uses rather than by their PTC. Paths from PTU to PTC and from test value to test preparation are the weakest in the model, respectively.

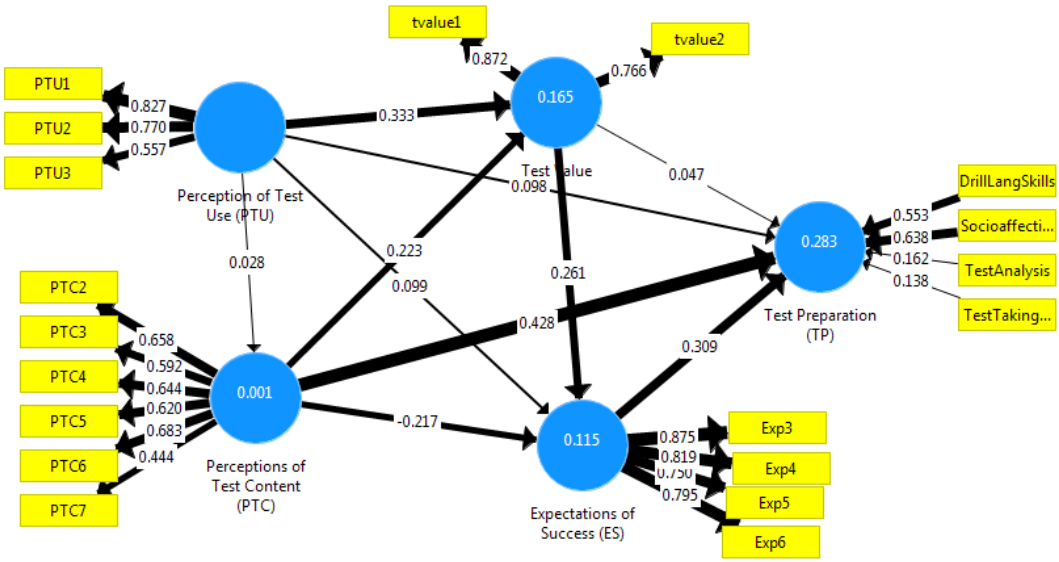

Figure 6: Washback Model 2 with path coefficients and R square values
(use PDF reader 'zoom in' function to read text)

As to the overall capacity of Model 2 in explaining the test takers' preparation for MALT, the corresponding R^2 value shows that .283 of variation in test preparation is predicted. By PLS-SEM criteria, this is a medium R^2 value, meaning that a considerable portion of variation in MALT preparation goes unexplained.

To check the statistical significance of parameter estimates, the nonparametric bootstrapping procedure was used to generate p-values or the f^2 effect sizes; a measure of the change in R squared value when a certain causal path is removed from the model (Garson, 2016; Henseler et al., 2014).

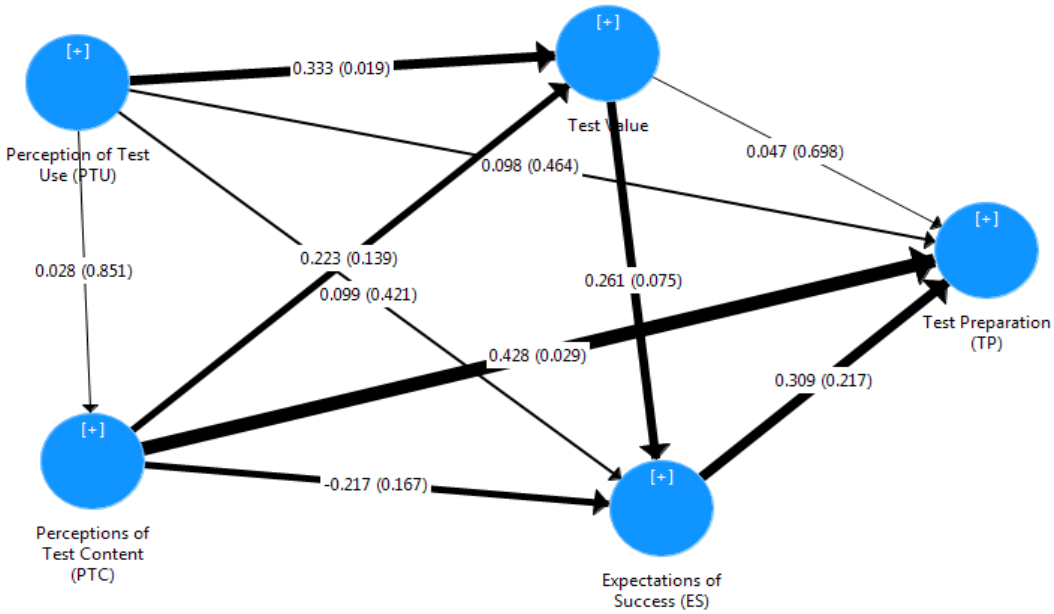

Figure 7: f^2 effect sizes for Model 2

In Figure 7, values on the paths are the p-values for the f^2 effect sizes. As it can be seen, there are two significant parameters in the model: PTC to test preparation (p=.029) and PTU to test value (p=.019). This observation suggests that for MALT test takers, endorsing high stakes uses of MALT is associated with valuing the test. It also indicates that perceptions of the necessary skills or knowledge for successful test performance are associated with more test preparation.

Discussion

This study examined the paths of effects from MALT test takers' perceptions of test content and use to their test preparation practices. The mediating role of motivational factors of value and self-efficacy was also studied. In so doing, two washback models, informed by the EV-theory, were postulated and assessed via PLS-SEM.

In broad terms, Model 1, being entirely based on the EV theory, was found to enjoy less predictive adequacy than Model 2. In Model 2, perceptions of test content appeared to be the best predictor of how test takers go about preparing for the test. Indeed, the direct path from perceptions of test content contributed the largest explanation for the outcome variable of test preparation (see Figure 7). Furthermore, test takers' perceptions of the stakes of the test were significantly associated with the value test takers accorded to MALT. Yet, test value did not seem to necessarily translate into variation in test preparation practices. This is consistent with the findings in Xie and Andrews' (2013) study. They also found that test takers' perceptions of test use did not explain a significant amount of variation in test preparation. This may sound counter intuitive given the

consensus in the literature that the higher the stakes of a test, the more likely test takers are to engage in test-directed learning (Hamp-Lyons, 1998; Alderson & Hamp-Lyons, 1996). Perhaps, there are other social, cultural and personal forces at play that mitigate the expected relationship between perceived stakes of a test and its washback. As such, the perceived direct linear relationship between test stakes and test preparation should be taken with more caution.

In addition, the prominent role found in this study for test takers' perceptions of test content partly paralleled Xie and Andrews**'** (2013) findings, in which the direct path from perceptions of test content to test preparation was significant. Similarly, as was the case with Xie and Andrews' study, test takers endorsing high stakes uses of the test appeared to value test taking. Test takers' expectations of the assessment outcomes also seemed to be an important contributing factor to test preparation, which echoed findings from prior research (Xie & Andrews, 2013).

In terms of mediation path analysis, findings were at odds with those in the literature. In Xie and Andrews' (2013) study, except for the perceptions of test use to self-efficacy, all the other three mediating paths were significant. However, none of the mediating four paths from the endogenous perceptions variables to test preparation appeared to be significantly larger than zero in this study, casting doubt on the explanatory potential of the expectancy-value motivation theory to explain test washback in various contexts. One possible reason might be that being primarily psychological, the theory is limited to the motivational characteristics of the individual learner or test taker. Hence, it fails to capture the contextual contingencies of the settings wherein a test operates.

In sum, in both models, a high proportion of variance in test takers' test preparation behaviour remained unexplained. The postulated models (1 and 2) explained only 22% and 28% of variation in the endogenous variable of test preparation, respectively (see Figures 4 and 6). This finding runs counter to Xie and Andrews' (2013) findings. This might have to do with the different nature of the two tests. For one thing, MALT is an entrance language test whereas CET, investigated in Xie and Andrews, was an exit one. The two tests were different also with regard to their test taker populations. In our case, test takers were English major students, who are intrinsically motivated to learn English and have a future career in the field. Therefore, it might be justified to claim that not all their test preparation is solely directed by the test's demands or its high stakes uses. In terms of the value dimension of washback (Hawkey, 2006; Watanabe, 2004), the lack of a perfect fit between the models and the data might be counted as evidence of positive washback in that test takers do not regulate all their language learning behaviour based on test demands. In an exit test, however, test takers' motivation for test taking might be rather different; that is, test preparation is likely to be more test directed and the washback more intense. In such cases, the exit test is likely perceived as an obstacle that should be overcome and once this is accomplished, the test's content may not be necessarily of relevance to what test takers will do in their future career. In other words, in an exit test like CET, there might be little, if any, target language use domain envisioned by test takers. Therefore, the test is considered a terminal end and the end is more likely to justify the means, hence more negative washback.

Since Alderson and Wall's (1993) pioneering study, the complexity of test washback has been emphasised by numerous studies (Cheng et al., 2015; Tsagari, 2011; Zhan & Andrews, 2014, *inter alia*). Considering the complex nature of test washback and the role that theory plays in non-experimental inquires (Byrne, 2010), it is not surprising for a single theory to fail at explaining all or most variation in a complex, social phenomenon like washback, for all models and theories are simplifications of highly complex phenomena in the world. "A critical principle in model specification and evaluation is the fact that all of the models that we would be interested in specifying and evaluating are wrong to some degree." (MacCallum, 1995; p. 17).

Concluding remarks

This study investigated the role of motivational factors as posited in the expectancy-value theory in mediating test washback. Of the two perceptions constructs, perceptions of test uses did not seem to mediate test washback to a considerable extent. In other words, test takers' mental representations of the test's high stakes do not seem to be of significant explanatory power, which comes across as counter intuitive. This may have to do with the broader socio-educational factors surrounding higher education in Iran, where millions of people are pursuing graduate and postgraduate degrees with no employment opportunities in prospect for them. The younger generation, facing the dilemma of whether to stay at home and feel unemployed or to pursue a graduate degree and feel busy, commonly choose the latter with no particular short or long term goals in mind. This may partly explain the observation that the value component of the expectancy-value theory does not bear on test preparation practices.

On the other hand, test takers' knowledge of test content and design explained the largest variation in test directed learning. Such perceptions also influence test takers' expectations of success on the test, which in turn affect the extent of engagement in test preparation. In other words, it is the test takers' understanding of the test demands that determines the strategies and practices they adopt for a test, regardless of what drives them to achieve their understanding of test content.

Limitations and future directions

This study took test takers' perceptions of MALT's content and uses as exogenous variables. Yet, it is plausible to speculate that test takers' perceptions are determined by antecedent individual and contextual factors. The question of how various personal, cultural and exam-related factors interact to determine test takers' perceptions is still open for further inquiry. This would necessitate adopting a broader theory wherein learners' social and individual backgrounds are included. Thus further inquiry utilising more complex theoretical models accommodating a wider range of socio-cultural factors hold the promise to further our understanding of the washback mechanism.

In explaining our findings we speculated that entrance and exit exams might motivate test takers in different ways. However, this has to be substantiated through empirical research. Thus, how differential uses of tests scores shape washback to test takers' motivation and

test preparation merits further inquiry. Finally, this study drew on the expectancy value theory for the formulation of washback models. Other theories of motivation (e.g., achievement goal theory) can cast further light on the mechanism of washback. In addition, both the magnitude of our data and our analytic procedures should be acknowledged as two major limitations of the study. The small sample size required using PLS-SEM, which is rather more for exploratory, rather than confirmatory, purposes (Ravand & Baghaei, 2016). Finally, in any quantitative study, the defensibility of the findings rests, to a large extent, on the quality of data collection instruments. The fact that we did not undertake a strong validation program for our instruments is one additional reason to interpret the findings with caution.

References

Alderson, J. C. & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing,* 13(3), 280-297. https://doi.org/10.1177/026553229601300304

Alderson, J. C. & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129. https://doi.org/10.1093/applin/14.2.115

Anagnostopoulos, D. (2003). Testing and student engagement with literature in urban classrooms: A multi-layered perspective. *Research in the Teaching of English*, 38(2), 177-212. https://www.jstor.org/stable/40171636

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191. https://doi.org/10.1037/0033-295X.84.2.191

Brown, G. T. L. & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice,* 15(1), 3-17. https://doi.org/10.1080/09695940701876003

Buchmann, C., Condron, D. J. & Roscigno, V. J. (2010). Shadow education: Theory, analysis and future directions: A rejoinder. *Social Forces*, 89(2), 483-490. https://www.jstor.org/stable/40984543

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications and programming* (2nd ed.). New York: Routledge. [3rd ed.] https://www.routledge.com/Structural-Equation-Modeling-With-AMOS-Basic-Concepts-Applications-and/Byrne/p/book/9781138797031

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105. https://psycnet.apa.org/doi/10.1037/h0046016

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. https://doi.org/10.1017/S0261444815000233

Fox, J. & Cheng, L. (2015). Walk a mile in my shoes: Stakeholder accounts of testing experience with a computer-administered test. *TESL Canada Journal*, 32(9), 65-86. https://doi.org/10.18806/tesl.v32i0.1218

Garson, G. D. (2016). *Partial least squares: Regression and structural equation modeling.* Asheboro, NC, USA: Statistical Publishing Associates. https://www.smartpls.com/resources/ebook_on_pls-sem.pdf

Green, A. (2017). Learning-oriented language test preparation materials: A contradiction in terms? *Papers in Language Testing and Assessment,* 6(1), 112-132. http://hdl.handle.net/10547/622430

Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, 13(2), 39-51. https://doi.org/10.6018/ijes.13.2.185891

Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education: Principles, Policy & Practice*, 14(1), 75-97. https://doi.org/10.1080/09695940701272880

Haggerty, J. F. (2010). *High-stakes language testing and young English language learners: Exploring L2 motivation and L2 test validity in South Korea.* Unpublished PhD dissertation, Carleton University, Ottawa. https://curve.carleton.ca/9c7e4af8-7963-4436-aeb0-8784410720a7

Hair, J. F., Hult, G. T. M., Ringle, C. M. & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM).* 2nd ed. Thousand Oaks: SAGE. https://au.sagepub.com/en-gb/oce/a-primer-on-partial-least-squares-structural-equation-modeling-pls-sem/book244583

Hamp-Lyons, L. (1998). Ethical test preparation practices: The case of the TOEFL. *TESOL Quarterly,* 32(2), 329-337. https://doi.org/10.2307/3587587

Hawkey, R. (2006). *Impact theory and practice. Studies in language testing*, 24. Cambridge: Cambridge University Press. https://www.cambridge.org/us/cambridgeenglish/catalog/teacher-training-development-and-research/impact-theory-and-practice/impact-theory-and-practice-paperback

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., Ketchen, D. J. Jr., Hair, J. F., Hult, G. T. M. & Calantone, R. J. (2014). Common beliefs and reality about PLS: Comments on Rönkkö and Evermann (2013). *Organizational Research Methods*, 17(2) 182-209. https://doi.org/10.1177/1094428114526928

Hsieh, C. N. (2017). The case of Taiwan: Perceptions of college students about the use of the *TOEIC*® tests as a condition of graduation. *ETS Research Report Series*, 2017(1), 1-12. https://doi.org/10.1002/ets2.12179

Im, G.-H. & Cheng, L. (2019). The test of English for international communication (TOEIC®). *Language Testing*, 36(2), 315-324. https://doi.org/10.1177/0265532219828252

Karjo, C. H. & Ronaldo, D. (2019, June). The validity of TOEFL as entry and exit college requirements: Students' perception. In *Proceedings of the Eleventh Conference on Applied Linguistics (CONAPLIN 2018).* Atlantis Press. https://download.atlantis-press.com/article/125911480.pdf

MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 16-36). Thousand Oaks, CA: SAGE. https://au.sagepub.com/en-gb/oce/structural-equation-modeling/book4796

Messick, S. (1996). Validity and washback in language testing. *Language Testing,* 13(3), 241-250. https://doi.org/10.1177/026553229601300302

Nguyen, H. & Gu, Y. (2020). Impact of TOEIC listening and reading as a university exit test in Vietnam. *Language Assessment Quarterly*, 17(2), 147-167. https://doi.org/10.1080/15434303.2020.1722672

Ravand, H. & Baghaei, P. (2016). Partial least squares structural equation modeling with R. *Practical Assessment, Research & Evaluation,* 21, Article 11. https://doi.org/10.7275/d2fa-qv48

Razavipour, K., Gooniband Shoushtari, Z. & Mansoori, M. (2018). The structural invariance of a model of washback to test takers' perceptions and preparation: The moderating role of institutions. *Journal of Teaching Language Skills*, 36(4), 171-201. https://doi.org/10.22099/JTLS.2018.28165.2440

Samaie, M. & Mohammadi, S. (2017). On the development and validation of a scale of test impact on test takers (TITT). *International Journal of Language Testing*, 7(2), 155-177. https://cdn.ov2.com/content/ijlte_1_ov2_com/wp-content_138/uploads/2019/07/456-2017-7-2.pdf

Sato, T. & Ikeda, N. (2015). Test-taker perception of what test items measure: A potential impact of face validity on student learning. *Language Testing in Asia*, 5(article 10). https://doi.org/10.1186/s40468-015-0019-z

Smagorinsky, P., Lakly, A. & Johnson, T. S. (2002). Acquiescence, accommodation, and resistance in learning to teach within a prescribed curriculum. *English Education,* 34(3), 187-213. https://www.jstor.org/stable/40173127

Smith, C. D., Worsfold, K., Davies, L., Fisher, R. & McPhail, R. (2013). Assessment literacy and student learning: The case for explicitly developing students 'assessment literacy'. *Assessment & Evaluation in Higher Education,* 38(1), 44-60. https://doi.org/10.1080/02602938.2011.598636

Tsagari, D. (2011). Washback of a high-stakes English exam on teachers' perceptions and practices. *Selected Papers on Theoretical and Applied Linguistics*, 19, 431-445. http://ejournals.lib.auth.gr/thal/article/view/5521

Wall, D. (1996). Introducing new tests into traditional systems: Insights from general education and innovation theory. *Language Testing*, 13(3), 334-354. https://doi.org/10.1177/026553229601300307

Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318-333. https://doi.org/10.1177/026553229601300306

Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe & A. Curtis (Eds), *Washback in language testing: Research contexts and methods* (pp. 19-36). New Jersey: Lawrence Erlbaum Associates. https://www.routledge.com/Washback-in-Language-Testing-Research-Contexts-and-Methods/Cheng-Watanabe-Curtis/p/book/9780805839876

Wigfield, A. & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68-81. https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., Tonks, S. M. & Klauda, S. L. (2009). Expectancy-value theory. In K. R. Wentzel & A. Wigfield (Eds), *Handbook of motivation at school* (pp. 55-75. New York: Routledge. [2nd ed] https://www.routledgehandbooks.com/doi/10.4324/9781315773384

Xie. Q. (2015). Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism. *System,* 50(June), 56-68. https://doi.org/10.1016/j.system.2015.03.002

Xie, Q. (2011). Is test takers' perception of assessment demand related to construct validity? *International Journal of Testing*, 11(4), 324-348. https://doi.org/10.1080/15305058.2011.589018

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. https://doi.org/10.1177/0265532212442634

Zhan, Y. & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: Insights from possible self-theories. *Assessment in Education: Principles, Policy & Practice*, 21(1), 71-89. https://doi.org/10.1080/0969594X.2012.757546

Zhan, Y. & Wan, Z. H. (2014). Dynamic nature of washback on individual learners: The role of possible selves. *Assessment & Evaluation in Higher Education*, 39(7), 821-839. https://doi.org/10.1080/02602938.2013.872769 [retracted 3 Dec 2014, https://doi.org/10.1080/02602938.2014.971544]

Appendix 1: Test preparation questionnaire

The test preparation questionnaire used a five-point Likert type scale ranging from "apply to me entirely" (5), "apply to me most of the time" (4), "apply to me" (3), "somewhat apply to me" (2), "does not apply to me at all" (1).

The mean and SD of items otherwise scaled (e.g., the first three items in Drilling language skills subscale) were open-ended and the mean and SD were computed using the arithmetic average of the number of hours the participants reportedly spent on each language skill or subskill.

Subscale	No.	Item	Mean	SD
Test analysis and evaluation	1	I spent more time improving my areas of weakness in English.	3.58	1.061
	2	I analysed MALT question types to identify frequently assessed points and tricky questions.	3.38	1.249
	3	I did not read English materials irrelevant to MALT.	3.38	1.381
	4	I adjusted my test-taking strategies according to my performance on practice tests.	3.15	1.107
	5	I spent my time mostly preparing for MALT.	3.45	1.246
	6	I analysed MALT test papers to identify the level of difficulty in each section.	3.44	1.238
	7	All my study materials were related to MALT.	3.22	1.222
	8	I read MALT coaching books to know more about frequently assessed points and test preparation strategies.	3.35	1.208
	9	I summarised the mistakes I make in practice tests to avoid making similar mistakes.	2.88	1.210
	10	I spent more time on those areas that can be easily improved during test preparation.	3.85	1.077
	11	I analysed MALT score distribution to judge the relative importance of sections.	3.81	1.351
	12	I tried different test-taking strategies and find one that best suits me.	2.73	1.100
	13	I analysed my own performance so that my test preparation can be more purposeful.	3.17	1.099
	14	How many MALT mock tests have you taken?	4.47	0.965
Drilling language skills	1	How much time did you spend practising Reading Comprehension every day?	3.25	1.237
	2	How much time did you spend practising Grammar?	2.76	1.116
	3	How much time did you spend practising Vocabulary and Idioms?	2.21	1.126

	4	During test preparation, I kept on practising my spoken English.	2.65	1.384
	5	I kept on reading English newspapers/websites.	2.61	1.174
	6	I kept on listening to English radio broadcasts.	3.36	1.276
	7	I kept on reading English texts aloud.	2.75	1.177
	8	I kept on writing diaries/blogs in English.	2.07	1.029
	9	I kept on communicating with English native speakers whenever possible.	2.42	1.266
	10	I kept on using English whenever possible, e.g. writing emails.	3.05	1.244
	11	I tried to improve my reading comprehension skills.	3.37	1.132
	12	I timed myself to improve my reading speed.	3.57	1.215
	13	I read any and each text I come across intensively.	3.34	1.025
	14	I focused on understanding difficult and complex sentences in the passages.	3.81	1.070
	15	I spent more time learning new vocabularies available in the past MALT papers.	3.75	1.230
	16	I prepared long lists of English vocabularies and their meanings and memorised them.	3.12	1.200
	17	I said the words aloud to activate my auditory memory and relate them to the words I already know.	3.01	1.250
	18	I kept looking up the difficult words in the dictionary.	4.16	1.060
	19	I prepared flash cards containing new words and their meanings.	3.29	1.370
	20	I studied 1100 words, 504 words, and GRE books.	4.08	1.200
	21	I improved my word skills by looking up and memorising new words in materials on Linguistics, Teaching methodology, and Language testing.	3.28	1.228
	22	I studied IELTS preparation books.	2.43	1.490
	23	I studied TOEFL preparation books.	2.39	1.410
	24	For the reading part, I studied textbooks I had used during my undergraduate studies.	3.13	1.260
	25	I studied English newspapers and journals.	2.31	1.220
	26	I made use of applications like Flashcard Maker and Any memo.	1.96	1.290
Socio-affective strategies	1	I encouraged myself to keep on working hard on English.	3.78	1.04
	2	I tried to learn from others.	3.75	1.07
	3	When did you start preparing for MALT?	3.45	1.59
	4	I tried to build up my confidence in MALT.	3.89	0.97
	5	I sought teachers' advice on how to improve my MALT performance.	3.42	1.237
	6	I shared my test preparation experiences with friends and classmates	3.47	1.06
	7	I rewarded myself when satisfied with my performance.	2.70	1.27
	8	I consulted senior students about MALT.	3.60	1.25
	9	I set my sleep and wake time in order to be fresh on the MALT exam day.	3.09	1.26

Rehearsing test taking skills	1	I memorised long lists of words and their meanings.	3.65	1.161
	2	I memorised all frequently assessed words.	3.76	1.15
	3	I practised arriving at the correct answer through logical elimination of alternatives.	3.48	1.10
	4	When I met difficulties, I made the best guess based on my life experience and background knowledge.	3.47	1.09
	5	I scanned key words in the text after reading and understanding questions carefully.	3.91	1.08
	6	I searched for the answers in the text according to the sequence of questions.	3.12	1.29
	7	I scanned key words in the text before reading and understanding individual sentences carefully	3.32	1.08
	8	I sought frequently assessed questions.	3.41	1.18
	9	I trained the skill for making inference from the context.	3.63	1.02
	10	I trained the skill of grasping the gist intensively.	3.79	1.07
	11	I tried to improve my understanding of difficult and complex sentences.	3.48	1.09
	12	I practised choosing the correct answers by eliminating wrong alternatives.	3.58	1.11

Appendix 2: Test perception questionnaire

The test perception questionnaire used a six-point Likert type scale ranging from "apply to me entirely" (6), "very much apply to me" (5), "apply to me" (4), "somewhat apply to me" (3), "does not apply to me" (2), "does not apply to me at all" (1)

Subscale	No.	Item		
Perception of test use	1	I take the MALT to get an M.A. degree for job seeking.	5.23	0.97
	2	I take the MALT to boost my English proficiency.	4.89	1.14
	3	I take the MALT to graduate with an M.A degree.	4.50	1.51
	4	I take the MALT to win a PhD scholarship.	4.03	1.68
	5	I take the MALT to meet my parents' expectations.	3.94	1.65
	6	I take the MALT to compete with my classmates.	2.67	1.40
	7	I take the MALT to have research opportunities.	3.47	1.50
	8	I take the MALT because I have nothing important to do.	4.00	1.58
	9	I take the MALT to challenge myself.	3.92	1.49
	10	I take the MALT to enhance my prestige and social status.	5.23	1.25
Expectation of success	1	I have confidence I will have a good performance in General English Section of the MALT.	4.47	1.15
	2	I think, this time, I can have a good performance in General English Section of the MALT.	4.29	1.12
	3	If I study well, I will get the Pass Mark in General English Section of the MALT.	5.25	0.84
	4	If I try enough, I will have a good performance in General English Section of the MALT.	5.28	0.82
	5	If I fail in General English Section, it would be absolutely due to my less than enough effort.	4.77	1.16

	6	Considering the difficulty of the General English Section of the MALT and my own proficiency level, I'm sure I will be successful in it.	4.36	1.17
	7	Considering the question types, the difficulty of the MALT, and my own proficiency level, I have confidence I can enhance my performance in Grammar.	4.86	1.14
	8	Considering the question types, the difficulty of the MALT, and my own proficiency level, I'm confident I can improve my performance in vocabulary and idioms.	4.94	0.97
	9	Considering the question types, the difficulty of the MALT, and my own proficiency level, I'm confident I can enhance my cloze performance.	4.65	0.96
	10	Considering the question types, the difficulty of the MALT, and my own proficiency level, I'm confident I can improve my reading comprehension performance.	4.90	0.96
	11	How do you predict your overall performance in the MALT administration.	4.06	1.14
	12	How do you predict your performance in Grammar section?	3.97	1.11
	13	How do you predict your performance in Vocabulary section.	4.14	1.07
	14	How do you predict your performance in Reading Comprehension section?	4.34	1.08
	15	How do you predict your performance in Cloze section?	4.00	1.12
Test value	1	Having a successful performance on the MALT is of major importance to me.	4.97	1.23
	2	Performing well on the General English Section of the MALT will be useful for my future.	4.52	1.23
	3	Taking the MALT will be helpful for my English learning.	4.09	1.30
Perception of linguistic knowledge	1	The majority of students can answer most of the Grammar questions.	3.21	1.25
	2	It is very difficult to get high marks in the Grammar section.	3.38	1.43
	3	Doing well on Grammar section of the MALT doesn't have much impact on the total score.	2.35	1.01
	4	It is difficult to get a high score on Vocabulary and idioms section.	4.05	1.42
	5	If I fail in the Grammar, it will be difficult for me to get a high score on the MALT General English section.	3.39	1.126
	6	If I fail in Reading comprehension, It will be difficult to be successful in GE section of the MALT.	3.92	1.11
	7	Performing well on Cloze part doesn't have much impact on total score.	2.56	0.92
	8	If I fail Vocabulary and Idioms part, it will be hard to succeed in GE section of the MALT.	4.52	1.15
	9	To perform well on Cloze Part, I must have the knowledge of terminology and collocations.	5.00	1.02
	10	I must have knowledge of syntax and grammar.	4.85	0.10

	11	To have a successful performance on Cloze section, I must know the subtle nuances of synonymous words.	4.65	1.08
	12	During preparation period, I focus on Grammar.	4.32	1.39
	13	While preparing for the MALT, I focus on my word skills.	5.36	0.83
Perception of reading knowledge	1	In order to succeed in Reading comprehension section of the MALT, I must know the meaning of each and every sentence of the passage.	2.67	1.60
	2	In order to succeed in Reading comprehension section of the MALT, I must skim the whole passage to get the gist.	4.75	1.23
	3	In order to succeed in Reading comprehension section of the MALT, I must select and read the main parts of the text.	4.22	1.28
	4	I must be persistent in finishing reading long passages.	3.11	1.61
	5	I must scan the passage to find the key words and locate the important details.	4.77	1.02
	6	It would not be necessary to understand the meaning of each and every sentence in the passage.	4.93	1.24
	7	It would not be necessary to grasp the important details in the passage.	4.19	1.29
	8	I would have enough time to read the whole passage intensively.	2.30	1.30
	9	I must get the gist of passage.	4.61	1.33
	10	I must make a good inference about the concepts and results of the passage.	4.61	1.28
	11	My background knowledge doesn't influence my comprehension of the text.	2.17	1.30
	12	I must be able to understand the author's attitudes and ideas.	3.95	1.30
	13	I must understand the relationships among the sentences.	4.55	1.10
	14	I must use test-taking skills to eliminate wrong options.	4.36	1.28
	15	Background knowledge has not much influence on my performance in Cloze part.	2.46	1.27
	16	To have a successful performance on Cloze section, I must draw a correct inference from the topic of the text.	4.53	1.22
	17	To have a successful performance on Cloze section, I need not understand the whole text completely.	3.43	1.42
	18	To have a successful performance on Cloze section, I must make use of test-taking skills to eliminate the similar options.	4.20	1.34
	19	While preparing for the MALT, I focus on Cloze part.	4.04	1.32
	20	While preparing for the MALT, I focus on my reading comprehension skill.	4.45	1.36

Appendix 3

Table 3: Discriminant validity cross-loadings and VIF of items

	Item	ES	PTU	PTC	TP	TV	VIF
ES3	Overall, I will perform well on MALT	**0.86**	0.20	-0.12	0.25	0.27	2.072
ES4	I will perform well on the Grammar section	**0.83**	0.04	-0.22	0.20	0.21	1.740
ES5	I will perform well on the Reading section	**0.77**	0.18	-0.13	0.20	0.08	1.624
ES6	I will perform well on the Cloze section	**0.76**	0.12	-0.04	0.16	0.22	1.813
PTC2	It is very difficult to get a high score on Grammar section	-0.22	-0.06	**0.68**	0.25	0.09	1.366
PTC3	If I perform poorly in Grammar, it will be difficult to get a high score on MALT	0.00	0.06	**0.58**	0.19	0.14	1.306
PTC4	I must be persistent in finishing long reading passages	-0.08	-0.07	**0.66**	0.22	0.05	1.350
PTC5	I must understand the relationships among sentences	-0.04	0.03	**0.59**	0.29	0.19	1.167
PTC6	I must learn to arrive at the right answer through eliminate wrong options	-0.11	0.09	**0.64**	0.29	0.22	1.227
PTC7	I must improve my knowledge of syntax and grammar	-0.12	0.05	**0.50**	0.18	-0.04	1.166
PTU1	I take the MALT to get an M.A. degree to get a better job	0.14	**0.82**	0.06	0.09	0.36	1.156
PTU2	I take the MALT to earn an M.A. degree	0.19	**0.74**	-0.06	0.15	0.21	1.249
PTU3	I take the MALT to meet my parents' expectations	-0.00	**0.60**	0.05	0.21	0.09	1.170
TP1	Socio-affective test preparation strategies	0.33	0.17	0.19	**0.76**	0.27	1.327
TP2	Test analysis	0.27	0.09	0.04	**0.43**	0.19	1.406
TP3	Test taking skills	-0.02	0.12	0.43	**0.71**	0.15	1.190
TP4	Drilling language skills	0.03	0.07	0.35	**0.61**	0.05	1.604
	Test value1	0.31	0.34	0.04	0.22	**0.90**	1.142
	Test value2	0.03	0.18	0.34	0.18	**0.72**	1.142

ES = Expectation of success; PTC = Perception of test content;
PTU = Perception of test use; TP = Test preparation; TV = Test value;
VIF = variance inflation factor.

Kioumars Razavipour is an assistant professor in the Department of English Language and Literature, Shahid Chamran University of Ahvaz, Iran. His areas of research interest include language assessment and epistemology.
Email: razavipur57@gmail.com

Mahboube Mansouri holds an MA in TEFL from Shahid Chamran University of Ahvaz, Iran, where she works as a General English instructor. She also teaches English in Iran Language Institute. Her research interests include language testing and teaching ESP.
Email: mansourimahboube6488@gmail.com

Zohreh Gooniband Shooshtari is associate professor of TEFL in the Department of English Language and Literature, Shahid Chamran University of Ahvaz, Iran. Her main areas of interest include SLA, psycholinguistics, and TEFL.
Email: zshooshtari@yahoo.com

Please cite as: Razavipour, K., Mansoori, M. & Gooniband Shooshtari, Z. (2020). Test takers' perspectives on an English language test in Iranian higher education: A washback study. *Issues in Educational Research*, 30(3), 1058-1083.
http://www.iier.org.au/iier30/razavipour.pdf