

Test-takers' Perceptions of Test Fairness: A Washback Study on Test Information Given Prior to a High-stakes Writing Test

In-hye Choi
(Seoul National University)

Choi, In-hye. (2016). Test-takers' Perceptions of Test Fairness: A washback study on test information given prior to a high-stakes writing test. *Foreign Language Education Research, 19*, 1-18.

The development of conceptual frameworks for fairness in language testing has significantly expanded the scope of discussions on fairness. Empirical research motivated by these theories, however, has been minimal. There is no research, for example, that thoroughly examines the perceptions of test takers concerning fairness as outlined in the frameworks. Taking the context of the admissions officer system of universities in Korea, this paper looks at how the information released by universities before their exams affect students' perceptions of test fairness. Furthermore, the paper attempts to examine the washback of such perceptions as test-takers prepare for the test. Two Korean universities that conducted written tasks for its early admissions processes were selected. Participants were six actual test-takers, who were extensively surveyed and interviewed. The results show that students placed much more faith in the assessment when the information regarding the test construct was relayed in a clear and consistent manner. Such results suggest that test developers need to pay acute attention to the information they provide to test-takers prior to the test, and that test-takers should be aware of the rights that they have regarding the disclosure of test information.

Key words: washback, fairness, test accountability, high-stakes exam, test-taker perception

I. Introduction

1. Fairness: The Ethics of Language Testing

How important should test developers regard the issue of fairness in their language tests? Some researchers have treated ethics in languages tests as an extension of validity, denoting that achieving validity achieves fairness. Messick (1989), for example, regarded testing ethics as consequential validity.

The trend is shifting. As Hamp-Lyons (2001, p.1) observes, language testing currently "seems to be moving to [...] an ethical phase". Initially driven by broader

social justice theories (Jensen, 1980) and crystallized by conceptual frameworks of fairness by Kunnan (2000, 2010), researchers are beginning to recognize that fairness is a critical component in language testing that comes before validity. Indeed, as Kunnan (2000, p.10) argued, tests are not “valid and reliable or even authentic and interactive” if they are not fair.

The central idea in Kunnan’s framework of fairness for language testing is that fairness is a complex construct whose multifaceted components must be considered throughout the entire process of test development. One important component of fairness that Kunnan includes in his framework of test evaluation is ‘access’, which refers to how accessible a test is to test takers financially, geographically, personally, educationally, and finally, in terms of the familiarity of test conditions and equipment. The underlying assumption for educational access and the familiarity of test conditions is particularly interesting. Specifically, educational access refers to the opportunity for test-takers to learn the content and the types of tasks that are demanded in the test. Familiarity with testing conditions and equipment refer to whether test takers are “familiar with the materials (such as computers), the procedures (such as reading a map), and conditions (such as using planning time)” (Kunnan, 2010, p.41). In other words, the two types of access denote that test takers must know as much as possible about the test *prior* to taking it. Providing such information makes tests accountable—stakeholders must believe that the intended uses of an assessment is justified. Consequently, test takers need to ensure that test takers perceive the test as accountable not only during the test and once they receive the results, but before the test as well, as they prepare for the assessment.

2. The Dearth of Fairness and Washback Studies on Language Tests

How does the application of the fairness framework affect students in the real world? Sadly, although social justice theories and Kunnan’s conceptual frameworks for fairness in language testing has significantly expanded the scope of fairness, empirical research by the frameworks have only barely begun (Xi, 2010, p.147):

“For one thing, current empirical research in language testing has been piecemeal. The studies have typically focused on only one of a number of different aspects of fairness at any one time. These aspects may include differential item functioning (DIF) investigations across sub-groups (see Kunnan, 2000 and Ferne & Rupp, 2007 for comprehensive reviews of DIF research in language testing), the influence of construct-irrelevant test taker characteristics on test performance (Alderson & Urquhart, 1985a, b; Zeidner, 1986; Hale, 1988; Kunnan, 1995; Clapham, 1998; Taylor et al., 1998), the

influence of interviewer behavior on examinees' speaking scores across studied groups (Brown, 2003), the influence of gender bias in oral interviews (O'Loughlin, 2002), the influence of gender bias in oral interviews (O'Loughlin, 2002), the invariance of factor structures of test scores across groups (Swinton & Powers, 1980; Hale et al., 1989; Oltman et al., 1990; Ginther & Stevens, 1998; Stricker et al., 2005), and the reliability of multiple-choice test scores across L1 groups (Brown, 1999.)"

Until now, empirical research on fairness has been conducted in only certain aspects of fairness, and none have dealt with the impact of fairness onto the test taker before they take the test. In terms of the test taking process, the concentration has been on during the test and the after the test, when the results are translated into scores. However, tests shape the teaching and learning processes, processes which happen before the test takers take the test. This influence is called washback (Alderson & Wall, 1993). Researchers have underlined the need to enlarge the scope of washback research. In their preface to *Washback in Language Testing*, Cheng, Watanabe, and Curtis (2003, p.xiv) write that

“[i]n the field of language testing, researchers' major interest has been to address issues and problems inherent in a test in order to increase its reliability and validity. However, washback goes well beyond the test itself. Researchers now need to take account of a plethora of variables, including the school curriculum, behaviors of teachers and *learners inside and outside the classroom, their perceptions of the test*, how test scores are used, and so forth”(Italics mine).

According to Alderson and Wall, washback divides largely into two types, washback to the program and washback to the learner. Much research has been made on washback to the program (Qi, 2007), while relatively less notice has been given to washback to the learners, “perhaps because of the difficulty of getting access to the participants” (Cheng et al., p.xv).

However, when considering the fact that learners are the most important stakeholders of language assessments, it is now time that researchers also look into the washback that generates in one of the earliest moments of a test-taker's testing experience: when he or she prepares for the test by examining the information provided by test developers. What affect does such information have on test-takers? How are their perceptions of the fairness of the tests shaped?

II. Context of the Study

1. The Stakes of University Entrance Exams in South Korea

South Korea's university entrance exams hold very high stakes. This stems from Korea's social and cultural belief that assessments act as "gatekeeper[s] of one's success" in life (Kwon, Lee & Shin, 2015, p.4). Getting into a prestigious university is "highly valued" (Finch, 2009, p. 95) because education is seen as a tool for social advancement. As a result, middle and high school students persevere to rank well in their midterms and finals. They study zealously to garner better scores in simulations of the nationwide college entrance exam.

There are largely two ways to enter a university in Korea. The first is the aforementioned exam, the 20 year-old CSAT (College Scholastic Ability Test), which is the most critical standardized tool for college entrance (Kwon et al., 2015). The subject of this paper's study is the second method, the admissions officer system. Similar to the U.S. system, in which each university administers its own criteria for selecting freshmen, the admissions officer system in Korea consists of many different entrance programs that students can choose to apply for.

Although the system opens up doors for students with diverse abilities, researchers have noted that the sleuth of information about myriads of entrance programs are confusing test-takers, parents, and teachers—there were a total of 3,298 admissions officer programs in 2011, with 208 universities juggling an average of 16 different admissions programs. Although washback studies concerning prior-testing information have been conducted in Korea, most have only been dedicated to the overt result (Hwang & Kim, 2012; Kim, 2010; Lee, 2009). The confusing flurry of information, researchers have noted, have played an active role in increasing students' reliance on private institutions. The result is that admissions processes seem to foster a socioeconomic divide: the more one can pay to obtain relevant information, the better admissions result that person will obtain.

This paper takes a different approach from previous studies, in that it takes a closer look at the internal thought processes of the test taker as he or she prepares for the test. Shedding light onto the essential causes of the test takers' behavior prior to the test will provide a better direction for both test developers and test takers in the development of a high-stakes test. Consequently, this paper looks into the relationship between the actual information provided by the university and the test takers' perception of the test.

2. The Test

This paper looks at two admissions officer programs, each administered by a different university, that select high English proficiency students through essay writing tasks. The two prestigious universities will be referred to as University A and University B. In 2014, at least five universities selected English-proficient students through essay tasks (Park, 2013), but in 2015, this number decreased to two. There may be many reasons for the decrease. For one thing, the admissions officer programs are inherently very volatile, and are subject to change every year. The government's interest in simplifying and merging the various programs may be another reason (Park, 2013).

Whatever the true reason may be, this paper chooses to look into English writing exams because 1) they are still very high-stakes tests, influencing many stakeholders, and 2) universities rarely provide information regarding oral interviews to test takers before the interview, making an in-depth comparison analysis difficult for research purposes. The following table outlines brief descriptions of the admission processes studied in this paper, based on 2016 results:

TABLE 1
Description of the Studied Tests

	University A		University B	
Admission process (weighted % of each process)	STEP 1 Essay(100%)	STEP 2 Interview(100%)	STEP 1 Resume(40%) - Student record - Official English Test Scores ¹	STEP 2 Essay(60%)
Time	50 mins	6 mins	N/A	100 mins
Competition	Number of students admitted: 101 Number of applicants: 1,793 Competition: 17.75:1		Number of students admitted: 25 Number of applicants: 313 Competition: 12.51:1	
Constructs tested	Passage Comprehension Logical Reasoning Writing Skills		Expressiveness Reasoning Grammar	

¹ The minimum scores that an applicant needed to have were 110 for the TOEFL(IBT), 960 for TOEIC, and 873 for TEPS.

Although both universities choose the essay as the main method of selection, they have slightly different administration processes. University A selects applicants solely based on their writing and interview skills, while University B factors other variables such as the applicants' high school GPA and their TOEIC, TOEFL, or TEPS test scores. The language constructs that each assessment tests also slightly varies. University A wants applicants to comprehend passages and structure that understanding into a logical essay, while University B looks for applicants that can write expressively, logically, and error-free. Nevertheless, since both universities regard the essay as a main method to select their candidates, the test information that would affect test-takers the most would pertain the writing section.

III. Method

1. The Research Questions

The study addresses the questions:

- 1) Does the amount and quality of information provided by test developers prior to the assessment shape test-taker's perceptions of test accountability?
- 2) What washback effects, if there were any, did test-related information have on students as they prepared for the tests?"

To answer the questions, it is crucial to 1) investigate the kind of information test-developers are currently providing test-takers with, 2) study students' perceptions of such information, and 3) determine whether these perceptions affected their beliefs of test accountability, as well as their actions as they prepared for the tests.

2. Participants

A total of 6 participants took part in the study. All of them had taken both exams provided by University A and University B, and thus were able to relate their personal experiences in preparing for and taking the actual exams. As they were recruited through the internet, the subjects did not know each other and they had all prepared for the exams differently:

TABLE 2
Profiles of the participants

	Student 1	Student 2	Student 3	Student 4	Student 5	Student 6
1. Profile						
Type of high school	GE	GE	AB (U.S)	SP	GE	SP
English proficiency	TOEFL 119	TOEIC 975	TOEFL 114	TOEIC 975	TOEIC 970	TOEIC 980
Method of test preparation	Academy	Academy	School Teacher	Tutoring	Academy	Academy
Total length of preparation	6 months	3 months	3 months	1 month	2 months	2 months
Number of hours spent per week	25	30	20	6	10	15
Test results (University A, University B)	P, P	F, F	F, P	F, P	F,F	F, P

* Note: GE refers to a general education high school, AB refers to a school abroad, and SP refers to a special-purpose high school.

** The actual test results for each school are marked by P and F, P meaning pass (wait listed included), and F meaning fail. Pass or fail refers to the final results of the entire admissions process, not just the results of the essay section.

3. Data Collection and Analysis

Data was collected through two main methods: a questionnaire and an interview. Just before the data was collected, participants received photocopied papers of all the information that the universities had released, so that they could refresh their memory and answer the questions correctly. The questionnaire was a simple one, asking for participants' basic profiles (such as their English proficiencies and how long they prepared for the exam) and their perceived efficacy of the test-related information released by the universities. Participants were asked to rate the sufficiency and usefulness of the information on a 5-point Likert scale, 1 referring to "least helpful", and 5 referring to "most helpful". The interview, which was semi-structured and lasted at least 30 minutes for each participant, was conducted right after the questionnaire and thus served to provide a better understanding of the students' intricate thought processes

that could not possibly be captured by quantitative measures. The questions in the questionnaire and interview have been provided in the Appendix.

The data collected from the questionnaire and interview were examined for the following patterns in the participants' responses:

1) Accountability:

- Did the students think that they were given sufficient/relevant information?
- Were students able to accept the results of the test? In other words, did students believe that the selection decisions made on the basis of the test were justifiable?

2) Washback:

- How did the information affect the way they prepared for the test or their anxiety levels?

IV. Results

In order to answer the research questions, the results are organized in the following order: 1) investigate the kind of information test-developers are currently providing test-takers with, 2) study students' perceptions of such information, and 3) determine whether these perceptions affected their beliefs of test accountability, as well as their actions as they prepared for the tests.

1. What Information Were the Test-Takers Provided With?

Both Universities A and B offered various material to inform students about the task characteristics the students would be assessed on, as can be seen from Table 3. First, both universities provided a comprehensive overview of the assessment, including information such as the subject of the test, the number of students that would be selected, the date of the exam, and the basic procedures of the assessment ("1. Overview of assessment"). Next, both universities provided a constant update of how many students applied to the program, so that applicants knew the intensity of the competition for a certain major ("2. Competition"). Finally, both universities provided prompts from previous exams, as well as an explanation of the prompts ("4-1. Past Prompts", "4-2. Explanation of past prompts"). These explanations included the intention of the essay prompt and a brief scoring guide.

TABLE 3
Test-related information provided by test developers prior to the assessment

University A		University B	
Type of information	Y/N	Type of information	Y/N
1. Overview of assessment	Y	1. Overview of assessment	Y
2. Competition	Y	2. Competition	Y
3. Simulated practice exam	Y	3. Simulated practice exam	N
4. 1) Past prompts	Y	4. 1) Past prompts	Y
2) Explanation of past prompts, including the intention of the prompt and a scoring guide	Y	2) Explanation of past prompts, including the intention of the prompt and a scoring guide	Y
3) Sample essay	Y	3) Sample essay	N

In general, however, University A offered more descriptive and a wider array of resources. The most salient differences lay in the practice exam and the sample essay. First, University A offered a simulated computer exam that allowed participants to predict any changes in the format of the test (interestingly, such information was not included in the 'overview of the assessment') as well as their relative writing abilities. For example, by taking the practice test for the 2016 admissions exam, participants understood that there would be three reading passages this year instead of two. After taking the exam, the university provided participants the essay score and their ranking. "I received a score of 96 and was ranked 18th out of roughly 200 students that took the practice exam," Student 1 said.

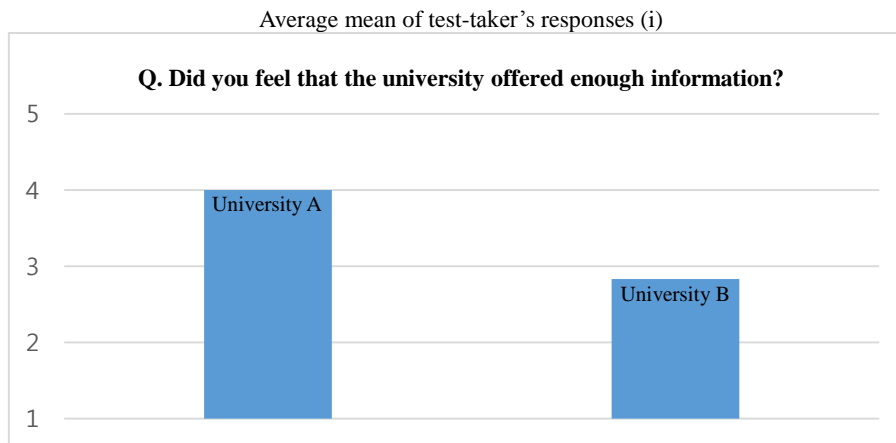
Another difference lay in the sample essay. While both universities offered scoring rubrics and a brief explanation of the intentions behind past essay questions, University A openly provided three model sample essays that were submitted by actual test-takers. However, the essays were not graded samples. As elaborated on in the next section, such information significantly helped test-takers understand the assessment construct and the test format.

2. Test-Takers' Perception of the Information

After conducting the questionnaire and the in-depth interviews, it was unsurprising that the participants generally perceived the information provided by University A as more helpful (Figure 1). When asked to answer whether universities had provided enough information for test-takers to prepare for the written exams without confusion,

the participants gave an average of 4 out of 5 for University A, while they gave an average of 2.8 out of 5 for University B. “Both universities did a fairly good job of providing me test-related information,” said Student 6. “However, I feel that University A gave me a much more concrete understanding of what the test was going to be like, and what I needed to do to prepare for the exam.”

FIGURE 1



* Note: Participants were asked to rate the sufficiency and usefulness of the information provided by each university. 1 refers to “least helpful”, while 5 refers to “most helpful”.

Indeed, examining the questionnaire and interview responses yielded some common themes. These themes could be summarized into two keywords: descriptiveness and consistency. In general, interviewees considered information sufficient and relevant when they had a concrete grasp of what they were going to be graded on, what the test was going to look like, and the quality of writing that the graders desired. First, Table 4 shows excerpts of the scoring explanations provided by both universities. Interviewees unanimously replied that the scoring rubric provided by University A felt more descriptive and concrete. Not only are there distinct percentages allotted to each criterion, allowing interviewees to see the relative importance of the scoring components, but the scoring distribution also provides further detail as to how the essays are scored. Although the explanation offered by University B also describes the scoring criteria, interviews replied that the explanation felt less helpful because it provided less information.

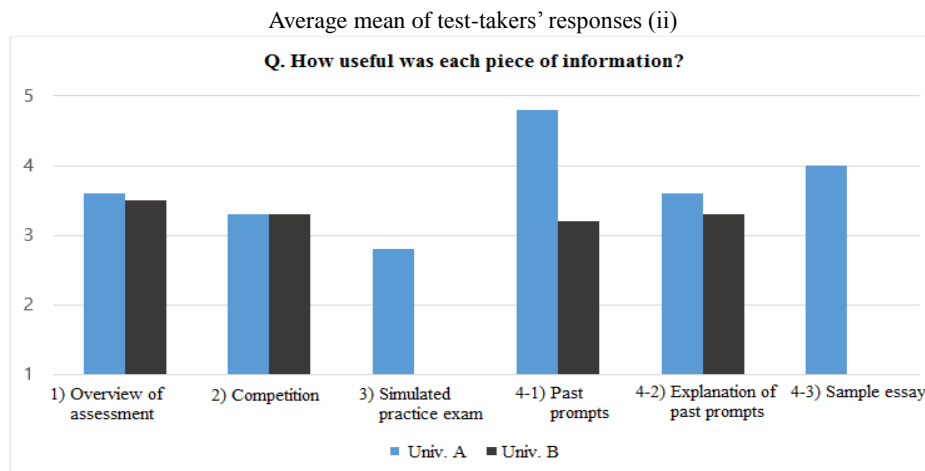
TABLE 4
Comparison of descriptiveness
between the scoring explanations of University A and University B

1) University A				
(1) Scoring Criteria				
Length	Relevance of Content, Paragraph Structure and Development	Original Ideas and Logical Discourse	Adherence to Essay Format	Language Use
Points deducted with discretion	50%	30%	10%	10%
(2) Distribution of Scores				
A+ (100-97)	A (96-91)	B+ (90-84)	B (83-80)	Fail (59-50)
The essay meets all of the criteria at an exceptional standard	The essay provides relevant content and in-depth analysis. It meets almost all of the criteria	The essay provides adequate content that lacks depth but meets most of the criteria	The essay is poor in content and fails to meet the criteria	The essay does not answer the question. It is less than 10 lines
2) University B				
(1) Scoring Criteria				
Students will be given scores within the range of A, B, C, D, and E according to how faithfully they fulfill the requirements of content and language use. First, pertaining to content, students receive higher scores if they include [...]. In terms of language use, students receive higher scores when the grammar (such as spelling, tense, subject-verb agreement, articles) is accurate and the expressions (such as stylistic and register choices, cohesion between sentences, the allocation of paragraphs) are used effectively. [...]				

However, an explanation of the scoring criteria was not enough for students to get a concrete understanding of the writing construct the university was looking for. “In the end, the scoring rubric too is filled with vague language,” said Student 3. “Just looking at

that doesn't really help. That's why the sample model essay provided by University A was so useful—by reading the student essay, I could get a clear understanding of what kind of output I needed to make at the actual test. I used the model essay as a guideline for structuring the essay and how to reason out my ideas.” Other participants thought so too. When questioned about the usefulness of each piece of information provided by the two universities, the participants gave the highest ratings to the actual prompts and the sample model essay (Figure 2). In the interview, the participants replied that seeing actual past prompts and model essays allowed them to precisely gauge the difficulty and look of the tests, as well as the quality of the writing that they needed to produce on exam day.

FIGURE 2



* Note: University B does not provide simulated practice exams nor sample essays.

It is interesting to note, however, that University A's simulated practice exam and University B's past prompt scored the lowest within the categories of information that each university offered, at an average of 2.7 and 3.2 points respectively. Why did the two pieces of information score so low when they were extremely concrete examples of the test? "They didn't represent this year's test format or level of difficulty," replied Student 1. "As a result, I felt that they weren't much help."

Student 1's reply sheds light on a second theme that interviewees agreed on: the need for consistency between the information provided and the actual test that test-takers will take. Descriptiveness is useless if the information fails to reflect the task characteristics of the actual test. In the case of University B, most interviewees replied that past prompts were not as helpful as they had expected because the test characteristics

changed last year and this year, without prior notice. In 2014, the exam requested test-takers to write a news article, instead of the usual essay that past prompts had provided. "I didn't know how to structure the response when I received the prompt," said Student 1, who needed to retake the exams this year after failing to enter his dream university last year. "I had never written a news article before." This year, the university changed the prompt into an essay. "But no one knew what the format would be like before we took the test, so I familiarized myself with writing in both the news article format and the essay format," said Student 2. "It was really frustrating." In the case of University A, the test format was consistent between the simulated practice exam and the actual exam, but all of the interviewees felt that they had been fooled by the easiness of the practice exam. Student 4 said that the practice exam was "so easy that I thought I would do equally well on the actual test." However, the actual exam turned out to be much more difficult.

To summarize, more information was generally deemed better. However, the quality of the information was what participants thought were also important. Test-takers' perceptions of the information quantity and quality were determined by two factors: descriptiveness and consistency.

3. The Effects of Test-Takers' Perception in Terms of Accountability and Washback

As hinted at in the previous section, the amount and quality of the information provided directly affected test-takers' perceptions of test accountability. In the interview, five out of six interviewees replied that they perceived University A's writing exam as accountable—they understood and accepted their test results, and they felt that the intended uses for the assessment were justified. Although the participants expressed anxiety and concern about subjective grading, they mostly felt that the university was doing its best to grade fairly and select the best-fitting candidates as much as possible. For University B, four out of six interviewees replied that the writing exam was accountable.

Further evidence supports the relationship between the good communication of test-related information and perceived test accountability. Interestingly, all the interviewees who had taken the University A exam, regardless of whether they passed or failed, said that although they trusted the results of the written exam, they had no faith in the process and the results of the interview process. "I have no idea why they selected me in the interview," said Student 1. "No one knows what they're going to be asked in the interview, and what exactly they're graded upon. The interviewers asked me two simple questions: How did you get here, and why do you want to get into our university. I don't

see how that could accurately capture the abilities of interviewees.” The dearth of information for the interview process cut down University A’s overall test accountability.

On the other hand, washback on a more extensive level proved to be much more difficult to investigate. Did the amount and quality of information released prior to the assessments shape the way students prepared for the tests? The answer was yes and no. On the one hand, students definitely relied on past exam types to prepare for the next one. This is why students were enraged when University B changed its test format from essay to newspaper writing in 2014, without warning. Students had prepared for the essay format, and felt that their preparations had become useless. On the other hand, although a correlation between test anxiety and the amount of information released by the universities seemed to exist, too many variables influenced the way and intensity of how students prepared for the exams. As the interviews proceeded, it became evident that the information released about the test, the difficulty of the test, test-takers’ study style, parents’ interventions, and financial as well as geographical access to private institutions all played significant roles in determining the test preparation process. Most students relied on private institutions, albeit for different reasons. For example, Student 6 said that she relied on academies in order to fill any information gaps she had about the exam. Student 1, on the other hand, said that he depended on academies because he was used to that study environment. As a result, more research would be needed to tease out the workings of these different factors.

V. Discussion

Along with the CSAT, the admissions officer’s processes in Korea are high-stakes processes that hold heavy consequences for test-takers, teachers, schools, private-institutions, and the society as a whole. However, compared to the national exam, there is little information about the task characteristics of the admissions officer’s processes, resulting in a bulk of students to prepare for the assessment in bewilderment, or to rely on private institutions. This clearly violates the two types of access proposed by Kunnan that test-takers are entitled to: educational access and testing conditions and materials.

Two themes surfaced repeatedly during the research process regarding the information that test developers need to provide prior to the test: descriptiveness and consistency. Test-takers need to have a concrete understanding of what is required of them, such as being provided with a graded sample essay or a precise scoring rubric. Furthermore, the information conveyed prior to the test must be consistent with the test material of the actual exam. These two factors directly determine the test-takers’ perceived accountability of the test.

Meanwhile, it seems difficult to conclude whether changing the amount and quality

of information released prior to the test influences the way students prepare for entrance exams. The interviews suggest that students continued to rely on private institutions regardless of the information given, although this is only a tentative conclusion since a single interview cannot capture the complexities of the washback that an assessment has. The short duration of study, as a result, is a limitation of this study.

Despite the limitations, this paper demonstrates that universities need to take a much more proactive role in making its tests fair throughout the entire process of test development. At the moment, the status quo is that students in Korea rely on very scarce pieces of information to prepare for high-stakes tests in the admissions officer system. Although the fairness framework shows that this undermines a test taker's access to the test, students are not aware that they have a right to request for more information. Severe gaps in information throw students into a whirlpool of confusion and anxiety. Secondly, more and better information raises test accountability, which universities should be concerned about. In this paper, test-takers accepted the intended uses of the assessment when they knew what to expect. Whenever the information was unclear, or inconsistent, as was with University A's interview and University B's written exam, test-takers tended to think that the test was unfair. Universities should be more concerned with how their tests are viewed by the most important stakeholder of their tests, the students.

The result of this paper raises additional questions. For example, what other aspects of the fairness framework, besides access, must be considered in high-stakes tests in Korea? Does the transparency—or the opacity—of information released in other admissions programs affect test-takers' perceptions of such exams? What washback effects that test-related information have on students and classrooms? Such issues need to be followed up with further research.

REFERENCES

- 김승태(2010). 입학사정관제도와 사교육의 연관성에 대한 소고. 국거버넌스학회보, 17(1), 237-264.
- 이윤미(2009). 입학사정관제도 확대를 우려하며. 우리교육, 234, 72-79.
- 박상복(2013). 대입 영어특기자 전형 단순화 방안. 한국교육과정평가원 연구정책세미나, 1-104
- 황여정 & 김경근 (2012). 입학사정관 제도에 대한 정보접근성 영향요인 및 그 계층적 함의. 한국교육학연구(구 안암교육학연구), 18(3), 183-211
- Alderson, J. C., & Wall, D. (1993). Does washback exist?. *Applied linguistics*, 14(2), 115-129.

-
- Amrein, A.L. & Berliner, D.C. (2002). High-stakes testing, uncertainty, and student learning, *Education Policy Analysis Archives*, 10(18). 1-74.
<http://epaa.asu.edu/epaa/v10n18/>.
- Cheng, L., & Watanabe, Y., & Curtis, A. (Eds.) (2003). *Washback in Language Testing: Research Contexts and Methods*. New Jersey: Routledge.
- Finch, A. E. (2009). European language education models: Implications for Korea. *English Teaching*, 95.
- Hamp-Lyons, L. (2001). Ethics, fairness(es) and developments in language testing. In C. Elder et al. (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies*, (pp. 222-227). Cambridge: Cambridge University Press.
- Jensen, H. R. (1980). *Bias in mental testing*. New York: Free Press
- Kunnan, A. J. (2000). Fairness and justice for all. *Fairness and Validation in Language Assessment*, 9, 1-14.
- Kunnan, A. J. (2010). Statistical analyses for test fairness. *Revue Française de Linguistique Appliquée*, 15(1), 39-48.
- Kwon, S. K., Lee, M. B., & Shin, D. K. (2015). Educational assessment in the Republic of Korea: lights and shadows of high-stake exam based education system, *Assessment in Education: Principles, Policy & Practice*, 1(18), 1-18.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational researcher*, 18(2), 5-11.
- Qi, L. (2007) Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China, *Assessment in Education: Principles, Policy & Practice*, 14(1), 51-74.
- Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment*, (pp.291–302). Dordrecht: Kluwer Academic.
- Xi, X.(2010) How do we go about investigating test fairness? *Language Testing*, 27(2), 147-170.

APPENDIX

1. Questionnaire Questions

1) Profile

- (1) What kind of high school are you attending at the moment?
- (2) What is your English proficiency level, according to either TOEFL or TOEIC?
- (3) How did you study for the exam?
- (4) What was your total length of preparation?
- (5) How many hours per week did you spend for the exam?
- (6) What were your test results for both universities?

2) Overall, did you feel that each university offered enough information?

3) How helpful was each piece of information to you?

- (1) Overview of assessment
- (2) Competition
- (3) Simulated Practice Exam
- (4) Past Prompts
 - (4-1) Explanation of past prompts
 - (4-2) Sample essay

2. Interview Questions

1) Was this a high-stakes test for you?

2) How did you prepare for the test?

- (1) Why did you choose to study alone or depend on someone else?
- (2) Did you proactively try to find relevant information provided by the universities?

3) Did you feel that the information that the two schools provided were enough?

- (1) Why did you find the information sufficient/lacking?
- (2) Did the amount and quality of information affect your preparation process?
- (3) What did you feel as you prepared for the test?

4) How was the actual test?

- (1) Did you feel that the information provided helped you during the actual test?
- (2) Did the results of the test make sense to you?
- (3) Do you believe that the test was administered in a fair, transparent way?

5) Did you try to ask for additional feedback to each of the schools? Or were there any instances of trying to contact the admissions office for some reason?

Choi, In-hye

Dept. of English Language Education at Seoul National University

1 Gwanak-ro, Gwanak-gu, Seoul

Email: c.inhye@gmail.com

Received on 31 October 2016

Reviewed on 15 November 2016

Revised version received on 30 November 2016

Accepted on 24 December 2016