The Scandal of Using the Internet Language Tests and Multiple Potential Liabilities from Using the Tests

(For discussion purposes only, and any feedback is welcome)

At this information age, when a corporation has a legal matter, it most probably has a large number of documents to be reviewed and the review normally requires a great deal of billing time. This practice mode creates two problems. First, the representing law firm may run into all kinds of unpredictable conflicts of interest issues. The documents may concern all kinds of commercial transactions between the client and other corporations. The law firm might have represented another party in a prior case to sue one or more corporations that happens to the partners of the client, and its attorneys might have acquired confidential information from the opposing party of the client. Second, the sudden demand for a large number of review attorneys forces the law firm to seek the help of contract attorneys. The law firm cannot maintain a large review staff because the firm may not have enough work for them after the case is concluded. Thus, the best solution is to retain contract attorneys ("document reviewers") from one or more staffing agencies. This also helps the law firm reduce the chance of encountering conflicts-of-interest issues.

Business globalization of large corporations bring about?? yet another problem. Corporations routinely use multiple languages in their routine transactions. Whenever a corporation is under a duty to produce its documents in litigation or administrative proceeding or conducts its own internal review, it must retain review attorneys who can read foreign languages. Due to a large demand for foreign language reviewers, foreign language reviewers are paid at much higher hourly rate. They can earn considerable amount of money.

Some law firms and staffing agencies now require foreign language reviewers to take language tests on the Internet and use the scores in making their hiring decisions. They use the scores to rank all potential candidates for selection or use the test scores as a mandatory passing mark. Some staffing agencies embrace this magic solution enthusiastically. One of the most widely used test vendors is ALTA Language Services. ALTA web site (www.altalang.com) allows any employer to set up a user account for each of its candidates and automatically sends log in name and password to the email address of each candidate. The candidate then can log in the web site and take the test remotely. ALTA does nothing to verify the identity of each test-taker, nor it is possible to prevent all kinds of frauds. Those law firms and staffing agencies somehow believe that this magic test can substitute for the decades of foreign education experience in a native language environment, and that native fluency can be achieved by getting high scores from taking this "open-to-the-public" test.

As shown below, this testing practice is a scheme for discriminating against native reviewers, fostering a "foreign language" review climate, promoting massive fraud in the non-English review employment market, wasting client's legal fees and reducing review quality, and gaining unfair competitive advantages in the legal service market and staffing market.

As demonstrated below, ALTA has demonstrated no credentials in the test art and nor actual knowledge of testing art. Its language expertise is questionable. It cannot show clear test objective, well-defined test participants, a required knowledge space, a sufficiently large test sample, and objective and unbiased questions-designing method. Its test model is based upon ill-conceived foreign language in America model. ALTA most probably has never evaluated its test validity before it uses the test in the real world.

A. Staffing Agencies and Law Firms Are Joint Employers Legally

In the document review industry, when a law firm has the need to review a large number of documents, it “hires” document reviewers from one or more staffing agencies on a temporary basis. The review may be housed in the law firm premises, a rented building, or a space occupied by the staffing agency. Technically, the staffing agency is the employer because it is responsible for paychecks. It is believed that many law firms make this arrangement to mitigate risk from imputed conflicts of interest. Notwithstanding the lack of direct employment relationship between the law firms and the document reviewers, the law firms are legally employers because they have absolute right to select reviewers, and absolute right to dismiss reviewers. The law firms can dismiss reviewers without letting staffing agencies know. The law firms are also responsible for creating review instructions, distributing review instructions, controlling daily work schedules, providing detailed supervisory work in all respects, and providing review tools. In each arrangement, the staffing agency only serves as a firewall for shielding imputed conflicts of interest. They are joint employers as a matter of law.

B. ALTA Language Tests Are Incompetent, Discriminatory, and Fraud-prone.

Problems in ALTA tests can be identified in its credential, fundamental design, and discriminatory impacts.

1. ALTA is a web operator without any credential as a testing agency

ALTA is an Internet web vendor that holds out to provide “language services.” Based upon the information and belief, it can be shown that ALTA

(1) has not been recognized by any native language authorities and has no credentials for evaluating language skills;

(2) demonstrates no knowledge of the most basic concepts in the testing art such as test objectivity, test participants, test designs, test performance, and test's representative, and it has invested little in research and development efforts;

(3) fraudulently represents to prospective users and the public that its tests can measure “native fluency,” whereas none of its questions are designed to test native vocabulary, native language structures, and foreign culture, all of which are the key components of native fluency;

(4) fails to disclose that it makes absolutely no effort to ascertain the identity of test takers and thus its test scores are meaningless and its tests are responsible for creating a fraudulent employment climate;

(5) fails to understand that the test scores from a test consisting of 15 to 30 questions are meaningless lottery scores, which can never reliably represent native-level language skills that can be acquired only by the life-time learning of the person;

(6) fails to inform customers and clients that its tests are absolutely open to the public and are susceptible to fraud in a dozen ways;

(7) probably copies test questions from other English tests and translated them into English with the American culture, and thus contained inaccurate, misleading, and confusing materials, and each translated test could be used in unlimited number of times, bring in a huge profit.

ALTA should be fully aware that it has no business to say anything about other people's language ability for employment purposes. Its tests are far worse than any recognized language tests because its tests fail to meet standards in every aspect from test objectivity, intended test takers, test knowledge space, question accuracy, design fairness, and sample sufficiency. The tests have fundamental flaws because they use a misconceived language-in-wrong-context model and contain omissions, inaccuracy, errors, and awkward expressions.

2. Fundamental Flaws in ALTA Language Tests

ALTA is a business entity and its only mission is to make a profit. From the limited disclosure on its web site, its test model has a fundamental flaw. The first problem is that it can never define test purpose and expected test participants. It tests are for everyone. The potential test participants include (1) native speakers (those whose primary language is the foreign language required for foreign language document review), (2) American born persons who has acquired a foreign language, and (3) second-language speakers.

a. Undefined testing purposes and undefined knowledge space

One fundamental flaw in ALTA language tests is in its claim that it can measure “native fluency.” It is common knowledge that native fluency can be acquired by education in the country where the language is used, but cannot be acquired in a foreign culture afterward. Even though a large number of web operators offer to institute all kinds of language tests for profits, none of them have proved that they can evaluate and rank native language skills. None of them even understand the theory behind all tests.

For a test to be valid, it must have a clearly defined knowledge space[1]. ALTA tests violate this fundamental rule. Testing is a unique way of “taking sampling” from a defined knowledge space for intended test takers. The knowledge space may be viewed as a collection of knowledge and skill elements (N1, N2, N3.... Nn). For a test to have any meaning, the knowledge space must be clearly defined, and a sufficient sample (S1, S2, S3.... Sn) is then taken. The

1 Different from the concept used mathematical psychology.

purpose is to draw a sample from the knowledge space to make an inference on the knowledge space. Since the knowledge elements cannot drawn like physical objects, the only way to get a sample is to ask questions and getting answers, and then from the right answer number to make an inference on the knowledge space. For any test to be valid, several fundamental requirements must be met. The knowledge space must be well defined. Obviously, the things like animal languages and personal knowledge do not belong to the knowledge space. The sample (i.e., the number of questions) must have a sufficiently large size so that the sample can represent the knowledge space, the test questions must be designed to measure the elements in the knowledge space. The sampling method must be unbiased. Since the test model is an imperfect model, it is very susceptible to abuse and bias. The test designer can easily design the test to selectively bias against certain test takers. One of the imperfections of the test model is that all test takers have strengths and weaknesses that may be correlated to their races, cultures, grow-up environments, family influences, education background, and economic conditions etc. In designing tests, the test designer may not consciously use any noted differences to favor some test takers and bias against others. A great number of cases concerning employment tests can be seen in http://www.fairtest.org/resources/employment.

When test scores are used to rate the knowledge or skill level of test takers, the knowledge space must be well defined and the test questions must be designed to measure the knowledge elements in the knowledge space. It is easy to prove the validity of administrating tests in the academic setting. When a course is presented to M students, all students are taught to learn the same knowledge space, and then test questions are designed to test only the elements in the knowledge space. Of course, the best method is to test everything that the teacher has taught, but it could require too much time. Therefore, the test containing only limited questions is designed to check only a few selected elements within the knowledge space. If the number of the knowledge elements in the test is sufficient, the scores can adequately and fairly reflect the knowledge space that the students have acquired in the class. In this case, the scores can also be used to rank their knowledge of this course. In driver's license test, traffic knowledge required is very well defined in the traffic law, and the test scores from a properly designed test can be used to predict how well the test taker has learned the traffic law. For professional license tests, an examination board has published outlines that clearly define the basic required knowledge (which is normally defined by course outlines). For high school equivalency tests, some published courses may serve as the guideline for defining the knowledge space. For a language test in the academic setting such as TOEFL, the knowledge space is the basic knowledge for living and learning at colleges. While the knowledge space for foreign language tests may be subject to debates, but the use of such tests in academia can be justified for necessity: the universities have to find a best way to determine if a foreign student has enough language skills to succeed in an American university. In each case, tests are designed to measure a well-defined knowledge space. It should be noted that when the knowledge space becomes bigger and bigger and its boundaries becomes more and more unclear, and the validity of the tests become more and more questionable. When a knowledge space is unclear or undefined, it would be difficult or impossible to design test questions to measure the unclear or undefined knowledge space. In employment cases, the knowledge space for any test must be directly related to the tasks of applied jobs. Any attempt to test something beyond the required skills must be for improper purposes.

b. Insufficient testing sample and unavoidable bias against some test takers

One obvious flaw in ALTA tests is that ALTA can never define the knowledge space for native fluency. From its early test questions, ALTA seems to have no understanding about knowledge space. Native speakers may major in any subjects such as history, art, performance art, literature, mechanics engineering, biology, biochemistry, chemistry, physics, medicine, business, accounting, investment, law, military, and aviation etc. The total number of occupations may well be in order of thousands. The distinctive vocabulary and unique expressions even in a field such as astrology, fortune-telling and ghost-driving ceremony can be overwhelming. It is not what ordinary people can understand. For the sake of convenience, the knowledge space of a language for any native speaker can be classified as basic knowledge, occupational knowledge, and extraordinary knowledge. The only component that can be compared among all native speakers is the basic skills (the ability to ask for directions and ability to attend lectures etc.). Even this portion of the knowledge space cannot be accurately defined in practice. The second component, occupational language skills, varies drastically among native speakers. As to the third component, there is no way to define them. Some native speakers may know ancient language, ancient history, ancient culture etc.; some may know enough information equivalent to half of a library (those who can win a prize in the Jeopardy show), and others may know very little beyond the basic and occupational language skills. Therefore, each of the native speakers may have a distinctive set of vocabulary, expression skills, and special knowledge. The knowledge spaces of native speakers are like various-size giant pots that contain different types and amounts of knowledge elements. Each of the pots is dramatically different from others. One should immediately conclude that no single test could be used in any conceivable way to measure their language skills and rank them. Contrary to the foreign language test case, where threshold language skills can be defined, native fluency cannot be measured for both theoretical and practical reasons. Any idea of measuring and ranking them is unattainable in practice. ALTA should have known such a basic concept before it gets into the business of ruining others life.

ALTA tests also violate the fundamental principle that the test sample must adequately represent the knowledge space. For this reason, its tests must be bias in nature and useless. Assuming, for the sake of argument, that the knowledge space could be defined (which can never), the test designer must determine how to design a test that can adequately represent or measure the knowledge space for each of the intended test takers. The amount of materials that any language test can include in practice is just a tiny subset of the knowledge space. While the knowledge space for a language is as big as a universe, the test size is absurdly small in light of all well-known tests. For example, a driver license's test may contain 20-40 questions; foreign language test such as TOEFL may have more than 100 to 200 questions; and a typical license examination takes 2 to 3 days (but test scores still cannot predict true ability). The knowledge spaces in those tests relative to the knowledge space of a language are like a drop of water relative to a bucket of water. One can immediately see that the test sample in ALTA tests is grossly insufficient. This gross insufficiency can readily seen by considering the knowledge components of various native speakers. If one question is allocated in one technical field, only the test takers in the fifteen or thirty allocated fields can competently answer one question. The remaining test takers will find that all questions are from technical fields they do not know. If all test questions are about astrology and fortune telling, only astrologists and fortunetellers can get right answers. Therefore, the tests must be bias against some test takers and favor others under all circumstances, regardless of how test materials are selected and how the questions are designed. Even if both the employers and the vendor have no actual intention to discriminate against any class, the scores of the tests are like drawing a lottery number and will distort true qualifications of all potential reviewers. The magic use of any test is that employers can eliminate any class of employees by selecting test types in light of their backgrounds. For example, they can eliminate native speakers with technical background by selecting tests, which contain advanced history and literature. The vendors may also design test types to please the demand of discriminatory employers. By a tacit agreement that they have reached from the need and demand, they can use the tests to achieve any purposes: favor some candidates and exclude others in the employment market in violation of Title VII of the Civil Rights Act.

To cure the insufficient size and unavoidable bias, the test must have a sufficient size for adequate representation of the knowledge spaces for all native speakers. The test must include properly weighed materials perhaps in tens of thousands of fields. A test so designed may be objective and fair for all test takers in theory, but will make all of them flunk. Each test taker will fail the questions in the fields the test taker has no knowledge. Due to the magnitude of amount of information in any language knowledge space and the extreme diversity in expression skills, vocabulary, cultural inferences and dialects, it is an absurd idea to use any test scores to rate native speakers. It cannot be done and have never been done in the Untied States or any nations. If Americans, including the president, senators, house representatives, lawyers, doctors, businessmen, farmers, janitors are required to pass an English test offered by a foreign web operator like ALTA, they all would feel insulted and vigorously object.

Why should the employers use such a language test? Does everyone need to know everything in the world? Why should a history professor know the language knowledge in mechanics? One has to conclude that the tests for rating native speakers cannot be administrated fairly and there is no need for such tests. The motivation for proposing such tests for native speakers is improper. Tests have been always a convenient tool for discriminating candidates. It is time to stop. New York City firefighter cases show after an employer used a test, canceling the test is a violation of Title VII, and continuing the test also a violation of Tittle VII.

2. Design Flaws of ALTA Language Tests

a. Well-accepted design of language tests

To understand whether a language test is valid, one must understand how how language is interpreted in a context. In any language, a context permeates language and contextual assumptions affect how we understand language.[2] In any writing, most the words and phrases are NOT defined in its verbal context. Therefore, the readers understand them according to contextual assumptions. The context assumptions for an English writing are those that we understand in America (the “English in America” model)

One of the most well known foreign language tests is the Test of English as Foreign Language (TOEFL). When foreign students come to the United States to study, they need to show they can live in this language culture and have the ability to attend classes. In this test, all questions are made on the basis of how the words and phrases are understood in America.“Foreigners” must accept the assumptions even if the assumptions are absurd, contrary to their belief and practice. This convention is necessary to avoid a potential situation where a language is subject to interpretation under two or more sets of conflicting context assumptions. If any foreign authorities provide language certification services, they would apply the same native context assumptions.

Moreover, a language test must have a reasonably defined language space and the test must be directed to the knowledge space. If a test maker select testing materials in a totally arbitrary way, the test scores may have no meanings. Before a foreign language test can be administered to anyone, the testing agency should show that its test is objective and representative. Despite considerable effort done in well known tests such as TOEFL, the test scores in language are still very poor prediction of language skills. The reason might be that the test is designed to test only limited number of elements in the knowledge space. Thus, test takers can get high scores by preparing for the test. Many foreign students can achieve very high scores in TOEFL by spending one or more years of special training. It is often noted in Academia, the reading ability and speaking ability of foreigner students are not consistent with their test scores. The scores are more influenced by test preparation efforts rather than true language skills.

b. Misconceived test model: foreign language lives in America

ALTA claims its tests can measure “native fluency.” Its full score 12 under its new score protocol is said to be equivalent to the skills of native speakers. However, its test questions are not designed to measure “native fluency” at all.

2 Context in Language, by Susan M. Ervin-Tripp, In Dan I. Slobin, Julie Gerhardt, Amy Kyratzis, & Jiansheng Guo, (Eds.) Social interactions, social context, and language. Hillsdale, NJ: Lawrence Erlbaum Associates

Native fluency may be defined as the proficiency level of a person who has spoken the language from very young age, and can understand essentially everything in the native language, including sufficiently large vocabulary, complicated grammatical structures, cultural references, and dialects." Although many different definitions can be found, the key components of native fluency are the knowledge of native vocabulary, native grammar structures, cultural inferences, and dialects. Assuming that native fluency can be measured and rated without violating the statistical principle, a test for measuring native fluency must be directed to the language elements in all those key aspects.

ALTA tests have never been designed to measure native vocabulary, native language structures, cultural inferences, and directs. ALTA has no ability to develop its own foreign language tests using the native contextual assumptions. It most probably copied English tests and translated them into various foreign languages. The original questions might have passed a rigid trial and unlikely contain any inconsistency between its contextual assumptions and specific verbal context. However, after any test question is translated into a foreign language such as Japanese and Chinese, there are several conflicts. If a question does not indicate its relationship with context, a test taker may apply Japanese context. As a result of a change in the ground rules, many words and phrases will be given different meanings and some will be in conflict with its verbal context. If a test question clearly indicates its relationship with United States, a test taker may apply the American assumptions by transforming the affected words and phrases. If the test taker applies native assumptions, which may also in conflict with the verbal context. The tests are conceived on an absurd test model: it inputs American context presumptions into a foreign language without proper ground rules on how to resolve conflicts in context assumptions. The same wrong test model might be used for many foreign languages such as Japanese, Chinese, Korean, Russian, or French.

The improper test model in ALTA tests can be seen from its historical tests. Its context assumptions are reflected in the questions. Some questions are about Capitol Hill, American Criminal Law, American Immigration Law, American Disability Act, Fraud in the Wall Street, and Stories in the United States Employment. The questions clearly indicate that tests are not intended to measure language skills in native context. By using such a model, ALTA distorts native vocabulary. Every word and phrase has its assumed meaning in the native context where the foreign language is spoken. Some examples are office products, medical supplies, good, evil, sun, earth, moon, and human beings. They have distinctive meaning scopes in any native culture. When a foreign language is used to describe anything in the United States or anything unique in the United States, each word and phrase is subject to two different interpretations. For example, the native speakers interpreted "medical products" to include herbs, urine, and animal by-products, but non-native speakers, who know only American assumptions, may interpret it to include cosmetic devices. The word "human being" means only human beings after birth in a native foreign context, but it has a broader meaning in the United States; the word "residency" in a foreign country is totally different from state residency in the United States; Motor Vehicle Administration ("MVA") in American context is the agency that issues official ID cards, but has no business to issue official identification cards in foreign cultures; and even "sun" may mean hope in one culture but mean something else in another culture. Every word, which is not specifically defined in the verbal context, carries the assumed meanings in the native culture. By imposing the American context to a language, ALTA essentially rejects the native vocabulary in the entirety.

The sentences in ALTA tests are awkward and do not have the flow of native expressions. This poor flow may be caused by the fact that the test questions are translated from another language. Some questions may be copied from language tests such as TOEFL or GRE. Due to the constraints of the sentence structures in the original language, it is impossible to make the translated text smooth. It is unable to test native expression structures. And finally, ALTA tests pay no attempt to cultural inferences and dialects. When its questions are based upon stories in the United States, it is impossible to ask questions about the native culture. It impossible to test dialects because it is impossible to know how many dialects. So, ALTA tests fail in all key respects.

C. The Testing Scheme Has Created a Fraudulent Employment Climate

The testing scheme of using a remote Internet test such as ALTA tests has created a fraudulent employment environment. It is well known that any of the Internet tests, the true identities of test takers cannot be verified and test materials cannot be controlled. The potential frauds include (1) taking a test by a substitute person; (2) ordering a test, practicing it, and figuring out so-called "correct" answers before taking a real test, (3) asking an associated business owner to create a test account for practices before taking a real test (4) taking a test while asking for help from another person in the same room, (5) repeating the same test from different agencies, (6) acquiring an expected test from a friend or buying it from an underground vendor, (7) taking the same test several times, (8) getting real-time help from a remote person by using second communication software such Skype, (9) using real-time training program (such "remote desktop", "remove screen", and "visual conferencing call" etc.) to do test with another person, and (10) studying the test, which has been acquired by a large number of means including Print-Screen, Paint, Screen-photo taking application, camera, and email program. It can be easily simulated that a perfect fraud can be committed in less than ten seconds.

The law firms, hiring attorneys and personnel managers should have seen the overwhelming possibilities of fraud, and should have not promoted the formation of such a fraudulent employment environment. Requiring candidates to take the test in a controlled environment alone is never enough to prevent fraud because there is no way to prevent prior exposure of the same or similar questions.

Due to the open-to-the-public nature of the Internet and uncontrolled characters of the ALTA tests, fraud must be presumed and cannot be rebutted.

By spending time, a person can get a "full" score by any of the methods. Deceptive persons can get a big reward for each point of increased raw score, which is about an improvement of one rank under its new scoring scale. Therefore, the scores become meaningless and their sole utility is to discriminate against native speakers or force the honest candidates to decline the tests.

D. Mere Use of the Test Scores Has Disparate Impacts on Native Speakers

The discriminatory intent of the tests is obvious. The law firms and staff agencies find a clever way to keep out native speakers even though native fluency is critical to foreign language review. This testing scheme has changed the nature of non-English language review into "foreign language" review (a review by foreign language speakers).

When same test is administrated among native speakers and non-native speakers, this arrangement is insulting and discriminatory to native speakers. The concept is as absurd as demanding American English speakers to pass English test created by a second language speaker as a requisite. Even those who advocate the use of the test would vigorously object to any similar test if they were required to take. By using such lottery scores, the employers can

(1) disparage the qualifications of native speakers and justify hiring decisions by citing their test scores. The employers use test scores to rate all reviewers in one single scale for selection. If everyone gets a full score, the employers can select candidates by drawing. If second language speakers get higher scores, the employers can select those high-score earners and reject native speakers;

(2) increase the candidate pool size by including otherwise unqualified or less qualified candidates. By using test scores, the employers can equate "two decades of native language education" to "two-year foreign language training," "two foreign country trips," "two-foreign language courses" or any unsubstantiated foreign language qualifications and justify hiring decisions by test scores. If both native speakers and non-native speakers get same score, the employers can select the non-native speakers and justify it by scores;

(3) force some native speakers to refuse to take the tests for various reasons: the test designs are flawed and incompetent, the test method is susceptible to abuse and fraud, the test is particularly insulting to native speakers. Whenever they refuse to take the test, they will not be hired;

(4) create an arbitrary drawing method of selecting reviewers. By using test scores as mandatory passing mark, a ranking criterion or a determinant factor, the employers select reviewers who possess limited language skills, no relevant case experience, no required technical background, and no relevant legal background as long as their scores are "high." If the candidate pool contains all native speakers, the employer can select reviewers by considering the scores rather than the three kinds of relevant experiences. In the end, the test scores give the employers the total freedom to select foreign language reviewers without the need to find best fits;

(5) promote fraud by those who will do anything to get high scores. Among all native speakers, the testing scheme will favor cheaters, and among all non-native speakers, the testing scheme also favor cheaters.

The impacts of the scores on different classes of reviewers are different. Among all potential reviewers, the most qualified reviewers are native speakers, considering the subject matters to be reviewed and the performance actually seen at review sites. Among none-native reviewers, the true qualifications would depend on the vocabulary and knowledge of each individual. When all reviewers are rated in lottery-type scores, the native reviewers with the highest qualifications would be impacted the most. The scores play a role of disparaging native language education receiving from foreign countries and put them at an equal footing with none-native reviewers. The disparate impact can also be seen from the probability point of view. Under a rational selection model, native speakers should have a far better chance to be selected but non-native speakers should have a lower chance to be selected. By using such lottery test scores, everyone would get the same or similar chance (assuming that the tests are not bias against any group). Therefore, native speakers get a reduced chance while the non-native speakers get an improved chance. Those who are otherwise ineligible will get far better chance to be selected. Those who are engaged in deception would also get better chances. The disparate impact can be seen also from the change in the candidate pool size. Native speakers will have a reduced chance due to an increased candidate pool. The end result is that the selection method favors non-native reviewers, but bias against foreign-educated native reviewers. The employers have engaged in discrimination in violation of the Title VII of the Civil Rights Act by mere use the tests even if the test were not bias against native speakers.

Staffing agencies are employment experts, and they know the fundamental unfairness of this selection method. Whether the test is fair is a question that can be resolved by asking how the rest of the employment world evaluates candidates' language skills. They should see the problem by asking a few hypothetical questions: What if U. S. senators are asked to take an English test created by a foreigner, and selected on the basis of their scores? What if the executives of the staffing agencies are asked to take an English test as a requirement? What if all English document reviewers are required to pass an English test created by a foreign entity? What if the hiring attorneys are required to pass a foreigner-designed English test? Anyone, even with the minimum intelligence, should immediately see the absurdity of this testing scheme and its gross inequity. So, it is not a poor judgment or a mistake. It is intended to gradually exclude foreign-educated foreign language reviewers from getting this high-pay job and place them as replaceable employees. The targeted members are new immigrants who may be Chinese, French, German, Japanese, Korean, Malaysia, Mexico, Mongolians, Netherlands, Russian, Taiwanese, Thai, and Vietnamese. This testing scheme discriminates against all of them even though many of them are United States citizens or permanent residents.

Some staffing agencies claim that it is difficult to evaluate foreign language skills. However, foreign education qualifications have been routinely evaluated in the academia and all corporations. The law firms themselves do not use such practice in hiring associates. Language qualifications can be determined by examining candidates' official transcripts. Even degrees obtained from foreign countries are fully recognized and can be reliably evaluated, and no credible argument can be made that somehow the selection method for foreign language reviewers needs to be different. Somehow, native fluency can be determined by 15 uncontrolled tests. The effects of this testing scheme are to substitute test scores for the academic records created by the institutions in foreign countries. Although Americans may reject credentials established by foreign authorities, rejecting native language education in the context of assessing native fluency is extraordinarily absurd. All foreign languages are created, developed, used, and maintained in foreign countries. All recognized authorities in each foreign language are in foreign country. If there is anything that Americans ever need to give deference to foreign authorities, foreign language qualification is one. What gives ALTA the right to think that it has a better ability to pass the judgment on foreign language qualifications? ALTA is ignorant.

Additional impact comes from the insulting nature of the tests. Some foreign-educated reviewers with exceptional skills routinely refuse to submit to such absurd, humiliating, fraudulent, and discriminatory tests and thus are to be excluded from getting this high-pay job, others may decline to join the fraudulent scheme of hurting end clients or out of the concern that they may be sued by the client for participating the fraud. Those incidences also reduce employment chance of native speakers.

E. The Test Model Discriminates Against Native Speakers

When ALTA tests scores are used to as an employment criterion, the test undermines the true qualifications of native speakers. By using the unreliable and fraud-prone scores, the employers can reject two-decade native education solely on the basis of the ranking of the test scores. Moreover, the effects of the tests go way beyond. The tests, due to the obvious flaws in its model and designs, also discriminate against native speakers directly.

ALTA tests are illegal because they fail to measure any of the relevant qualifications. They are not intended to measure the language skills: native vocabulary, native expression structure, cultural inferences, and dialects, which are critical to the job performance. ALTA tests are also illegal because their test subject matters are irrelevant to the review subjects. Some subjects encountered in foreign document reviews are commercial bribery in foreign countries, software business in China, Taiwan, Japan, and Korean, storage devices business in China, chip-developing business in Japan, mining operations in Mongolia and Russian, research and development activities in various foreign countries, storage tank business in Africa, drug development operations in China, equity fund lending business in foreign countries, securities trades in foreign countries, anti-trust conduct in foreign countries, corporate looting activities in foreign countries, and export violations committed in foreign countries. In each case, the tasks are to conduct factual findings from foreign language documents, which may contain many distinctive subsets of highly technical vocabulary. They are essentially all written by native speakers. For example, documents for large equipment installation business may include a large number of organic chemicals names, foreign client names, engineering concepts (even the meteorology parameters can be overwhelming), international secured transactions, foreign tax law and custom convention. It could have anything but U. S. Immigration law, U. S. Wall Street fraud, U. S. disability law, U. S. tax law, and U. S. capitol structure. A different case involving software development and sales business could require completely different background knowledge. Based on the recent ruling in NYC Firefighter Test case[3], ALTA tests must be illegal as a matter of law. If an employer is truly concerned with foreign language qualifications, it should identify all main subject matters and identify candidates who have three categories of relevant knowledge. This can be achieved easily.

ALTA not just fails to make any effort to ascertain “native fluency” as it claims, but achieves exactly the opposite: the tests penalize the candidates for their native fluency. Its ill-conceived test model is bias against native speakers. When a foreign language test is designed with activities and events in the United States, the test questions take the assumptions recognized in the United States as the context assumptions. The context assumptions may be in irreconcilable conflicts with the native context assumptions in light of the verbal context. The following table shows several situations where native speakers are in the clash between two sets of context assumptions. The conflicts in context assumptions can be found for all words and phrases, all social concepts, and all concepts concerning the culture, business practices, science practices, governmental organizations, and legal environment. This analysis shows that a test applying native language into foreign context assumptions will create irreconcilable conflicts in virtually every word and phrase.

One can also note that when a foreign language document is written with a verbal context consistent with native assumptions, an American reader will always feel it does not comply with native English writing convention. One can easily see why accurate translation of foreign documents will result in a translation that is not elegant and smooth. This is because every term concerning culture, business, government, and legal system may be different from what is known in the United States. Many of the concepts cannot be expressed in an American way. A demand for elegant and smooth English style translation would require massive distortions of original meanings. The law firms should take notice and should educate their attorneys in language basics before they run foreign language review projects.

The effects of context assumptions are best shown in an example concerning an article about tax liability and tax mitigation. Many foreign countries use a value-added tax system while the United States uses progressive

3. In his decision, U.S. District Judge Nicholas Garaufis found that the written tests had "discriminatory effects and little relationship to the job of a firefighter."

income tax system. When a foreign article about tax is read by an English reader who knows only the United States tax system, the reader will think the translation makes little sense unless the reader understands how the value added tax system works. When an article on U. S. tax liability and mitigation strategy is translated into a foreign language, it would create all kinds of confusion, depending upon who reads it. First, the translation makes perfect sense to an American reader who understands the U. S. tax system. Second, the translation makes no sense to a native reviewer who knows only the value-added tax system. Third, an America reader who understands both tax systems, may have to determine whether the company is a foreign company or United States company and which tax system controls. Besides all those problems, poorly selected words and subtle distortions in the translation further complicate the analysis.

Table 1-Different Effects of Context Assumptions and Verbal Context ("VC") on Native Speakers and Non-native Speakers

Type of Words or Concepts	Native Context Assumption	U. S. Context Assumption	Verbal Context (VC)	Effects on Native Speakers	Effects on Non-Native Speakers
"Office products"	Narrow scope	Broader scope	VC is consistent with native assumption	No problem.	May note it is narrower than U. S. counterpart
"Office products"	Narrow scope	Broader scope	VC is consistent with U. S. assumptions	<u>Native speakers cannot understand</u>	No problem
Social concept, "residency"	Residency is a very complicated issue.	Residency can be changed freely.	VC is consistent with native assumption.	No problem.	Note it is different
Social concept, "residency"	Residency is a very complicated issue.	Residency can be changed freely.	VC is consistent with U. S. assumptions	<u>Native speakers cannot understand due to the conflict in assumptions</u>	No problem

Business operation	X way to do business	Y way to do business	VC is consist with X way	No problem.	Note the different way
Business operation	X way to do business	Y way to do business	VC is consistent with Y way	Native speakers cannot understand.	No problem.
Secondary meaning	Secondary meaning arising from special term	None	VC is consistent with the secondary meaning	No problem.	Hard to understand.
Secondary meaning	Secondary meaning arising from special term	None	VC is inconsistent with the secondary meaning	Native speakers will be confused if there is sufficient verbal context to see the secondary meaning	Hard to understand.
Foreign humor	Based upon native assumptions	Lacking	VC is consistent with native assumptions	No problem (understand the humor).	Difficult to see the humor.
Foreign humor	Based upon native assumptions	Lacking	VC is inconsistent with native assumptions	Native speakers may see problems in the humor	As if no attempt were made.
U. S. humor	lacking	U. S. assumptions	VC is consistent with U. S. assumptions	Native speakers need the U.S. assumptions to see the humor	No problem (See humor).
U. S. humor	lacking	U. S. assumpti	VC is inconsiste	Native speakers	Cannot understand

		ons	nt with U. S. assumptions	cannot understand.	

When a language finds its way into a new culture, the language undergoes transformation. It must go through certain changes. When a foreign word "residence" is used to mean state residence in the United States, it acquires a new meaning, which is completely different form the original meaning; when the foreign word, marriage, is used to mean marriage in the United States, it acquires a broader meaning of marriage in the United States. Changes must be made to almost every word and phrase to different degrees (except systematic chemical names). By this process, the language is "adapted" to the new culture with different business practices, science practices, legal environment, and human interactions. By this process, new context assumptions replace the original native context assumptions. It is obvious that it is more difficult for the native speakers to adjust to the context assumptions. In comparison, non-native speakers do not carry native context assumptions, thus can quickly accept the context assumptions they already know. Moreover, many second language speakers learn foreign language vocabulary by using his or her primary language as a bridge. They usefully pay little attention to the differences in context assumptions and so can accept new context assumptions. Therefore, ALTA tests are extremely bias against true native speakers because the tests place them in a clash between two sets of context assumptions. It is grossly unfair to force native speakers to resolve such conflicts during the tests.

The conflicts in the ill-conceived test model are sufficient to change test scores even if fraud, prior exposure, and coach preparation have not been committed by anyone. In an ALTA test, each question is followed by four choices of possible answers. Usually, two of the choices can be excluded easily, but another two choices are very close. When the native speakers have two conflicting context assumptions, the answer from the two choices would depend on how the two context assumptions interfere with the analysis in light of the verbal context. Also, due to omissions, distortions, and errors introduced in translation, some questions may be completely improper. When such a test is administrated among none-native speakers and native speakers, the test discriminates against native speakers. This explains why some native speakers, the best native speaker, cannot pass its tests. The test has more negative impacts on the native speakers who have expert knowledge. The native speakers who carry only very limited native assumptions are in a better position to accept new context assumptions, thereby achieving "better" scores. The tests are truly a reproach to fair employment practices.

Due to the change of context assumptions, the tests are also bias against foreign-educated test takers who have substantive knowledge in one or more fields. For example, matter is classified in one way in philosophy, but in a different ways in other natural science. When a question is concerned with "matter," the intended answer may be in conflict with the classification in both philosophy and natural sciences. The test takers have to resolve the conflicts between the intended answer and the assumptions used in the substantive fields. When a test contains some terms inconsistent or in conflict with the teachings in substantive fields, the test is bias against those who have substantive knowledge, but favors those who know nothing in the field. Assuming that test scores are true representative of language skills and no fraud has been committed, the testing scheme not only eliminates the true qualifications advantages of native speakers, but also makes their scores lower so that the employers can conveniently exclude them. It also discriminates against those who have knowledge that is in conflict with the intended answer.

There are stories that the best foreign language reviewers cannot "pass" language test. The real problem is that the tests themselves are misconceived on an improper model, are totally flawed in designs, incompetent to measure anything, and clearly discriminatory against native speakers.

F. Selection Method of Using The Testing Scheme Injures Clients

The testing scheme injures end clients in four ways. In a vast number of cases, the activities under review take place in foreign countries. Some of the common subjects are commercial bribery in foreign countries, research and development activities in foreign countries, anti-trust conduct in foreign countries, securities frauds committed in foreign countries, corporate looting activities in foreign countries, and export violations committed in foreign countries. If the client of a document review is a foreign client, the need for retaining native reviewers is indispensable. The client has a strong interest in accurately reviewing their documents. Failure to identify problems will cause the client to miss the opportunity to correct problems. This could lead to terminal sanctions in a later administrative action. If the client is a United States company, foreign-language documents are always concerned with activities and events in foreign countries. The ability to understand foreign language in foreign context assumptions is the key. The client will have the same level of interest to discover what happened in foreign countries in order to find right solutions to the problems. In all foreign language review projects, knowledge of foreign culture, foreign business practices, foreign science practices, human interactions, geographic locations, and people name conventions, are critical. Exclusion of native reviewers will seriously diminish review quality. The impact, however, may vary. For uncontested cases and low-risk cases, review qualify may have no impact at all. For contested and high-risk cases, only hundreds of documents in over a million may be critical. Failure to identity by the whole review team would be fatal error. In rare cases, only one to a few documents may dispositively affect the disposition of the cases.

A review by all non-native reviewers increases review costs. It is well known that native fluency requires a native education environment. Any person who has both native language and a second language knows the great differences in reading speeds and comprehensive ability. They can read documents in native language much faster than in a second language. This inherent nature cannot be changed, perhaps, in their life times. Native fluency cannot be achieved from taking two courses, eight-week training, or a two-year visit in a native language country. It cannot be achieved by scoring high in ALTA tests. Such high scores will not help anyone perform better. There are specific reasons for performance differences. Non-naive reviewers have to discuss native vocabulary, special terms and phrases, people names, cities and locations, cultural concepts, business practices, science practices, and legal environment, histories from time and time. This is just way to much to learn in a review site. A native reviewer can instantly resolve a language issue what might cause a group of non-native speakers to debate for tens of minutes. At any multiple language review sites, it is a recurring scene that non-native reviewers may have to struggle from time to time.

The impact of the testing scheme has far-reaching effects. When everyone knows this magic way to establish native-fluency qualifications according to the ALTA claim, a fraudulent climate form where everyone tries to establish their native fluency instantly. More and more high-score achievers will replace native speakers. The final score that a person can achieve depends on how far the person is willing to try. This will create a climate for racing for the bottom. More and more non-native reviewers will replace native reviewers. When a non-English review becomes second language review, it is the client who will be injured seriously.

By using test scores, the employers can make no effort to find the best fits. The tests helps them form an employment culture where document reviewers are not selected on the basis of best fits. ALTA test subject matters have absolutely no relevance to review skills. ALTA seemed to try to measure the knowledge of American Law in foreign language. No foreign language review ever has anything to do with the transactions and activities in the United States. The main tasks are to understand foreign transactions in foreign counties. Unreliable, fraud-prone, and discriminatory test scores are used to replace true qualifications. By using such an arbitrary selection process, reviewers without software and patent backgrounds may be selected to review software patent matter; reviewers without accounting background may be select to review accounting matters; reviewers without secured transactions background may be selected to review international business matters.... The documents from a business in mining operations, waster processing, manufacturing machinery, software and hardware, may involve many sets of unique vocabulary in several fields. Many review teams comprise reviewers without required qualifications. Bar licenses and language scores become the only qualifications for finding foreign language reviewers. This selection method is contrary to the well-accepted hiring principle used by law firms, staffing agencies, and the rest of the world.

The test scores place experienced reviewers and new reviewers on a same footing. When law firms are presented with experienced reviewers and new reviewers who all have similar test scores, the law firms rather select the new reviewers because it is far less burdensome to conduct conflict clearance. By increasing inexperienced reviewers and reducing experienced reviewers, the review product may have increased risk of expose additional liabilities. When a client has multiple fields of business, its documents may reflect all kinds of potential liabilities in criminal law, marketing law, expert control, foreign corrupt practice, securities fraud, wire fraud, patent infringement, and trade mark infringement etc. Only those who have specific experience in handling particular risky issue can detect the issue. No single lawyer can know all risk issues. If the review team comprises all new reviewers, the chance to catch those documents for careful analysis and proper treatment is substantially reduced. In a recent suit between Victor and DLA on legal fees, a $22 million punitive claim was based upon email exchanges between DLA attorneys discussing legal bills. It takes only one document to expose massive liability.

Using the fraud-prone Internet language test to select foreign language reviewers may create potential liability on the law firms, staffing agencies, and documents reviewers. M-J manufacture case is an example that all parties can be sued in a malpractice claim arising from e-discovery. A foreign client who has lost a case needs to establish only two additional elements to allege a malpractice claim. One of the elements is a bad work product, and this can be established in nearly all review productions. Some of common translation problems include distorted meanings, lost meanings, added meanings, distortions from cultural shifts (failure to do necessary adjustment due to interpreting context shifts), confusions caused by word multiplicity, translation inconsistency, failure to translate important documents, failure to note implied meanings, failure to add omitted assumptions, and failure to ascertain special usages.... Common coding errors include failure to catch privileged documents, disclosure of irrelevant and injurious documents, and failure to redact harmful and irrelevant substances. In the worst cases, when documents are translated out of language contexts, translations may be “accurate” and “elegant” nonsense. There are hundreds types of potential errors can be found in abundance. The problems alone are not sufficient to allege a malpractice claim. If this testing scheme is used to select reviewers, the client can attribute the problem to this lottery-type and fraud-prone Internet tests. The client may be able to demonstrate how the testing scheme and the fraudulent climate has favored the selection of the deceptive and the incompetent in place of true native reviewers and thinking reviewers. The number of problems and their serious degrees may depend upon the qualifications of the review team. This would be a very strong ground for imposing liability.

The rest of elements for a negligence claim can be established as in any malpractice claim. M-J manufacturing claim clearly teaches that staffing agencies and law firms need to do much more than just using bar license and meaningless test score in selecting reviewers. Another angle to look at this problem is whether a reasonable person would object to the method of selecting attorneys on the basis of the scores of an English test. It is well known that attorney’s performance has little to do with their English skills, and law firms normally consider specific experience, relevant case experience, and relevant legal knowledge in retaining associates. No creditable argument can ever be made that document reviewers are not important when in fact every case is decided on evidence. The testing scheme is indefensible. The Qualcomm case shows serious consequence from violating discovery rules in a contest case.

G. The Testing Scheme Injures Non-Participating Law Firms and Staffing Agencies

When a small number of law firms and staffing agencies use this testing scheme, they might have represented to their clients their "best effort" to find the "best native reviewers." Some clients are misinformed of what actually happens, they may also ask for a language test in selecting reviewers. The clients may have not fully informed of the ill-conceived test model, the fundamental flaws, the lack of representation of the test scores, the susceptibility to numerous frauds, the incompetence of the testing vendors, and the nature of racing for the bottom. Those law firms and the staffing agencies might have gained unfair and unjustified competitive advantages from promoting this fraudulent testing scheme in securing clients business or staffing contracts. Those law firms and staffing agencies that consistently refused to use the tests are in the disadvantaged positions. They might have lost clients and staffing contracts. Only time will tell how many of the law firms and staffing agencies might have been injured by this fraudulent testing scheme.

H. Participants of the Testing Scheme May Violate Ethical Rule

In academia, testing environment is fully controlled to make its academic records to have any meaning. When the law firms and the hiring attorneys advocate for this testing scheme, knowing all obvious fraud and abuse, their actions actually encourage the dishonest conduct of some reviewers, such conduct may be in violation of ethical rule. Knowing the fraudulent, incompetent, absurd, discriminatory, and anti-competitive nature, the hiring attorneys cannot continue to support the testing scheme without violating ethical rule. The retention of the incompetent and deceptive reviewers may also have a negative impact on client cause, and therefore, the law firms and hiring attorneys also have an ethical duty to stop this testing scheme in order to advance the interest of their clients.

The testing scheme also places other fellow document reviewers at the risk to be sued. Most of the methods of committing frauds can be neither proved, nor disproved. However, some of the deceptive conduct can be readily proved in due course. Many foreign language reviewers may have applied for foreign language review to many staffing agencies. They have been asked to take the same test again and again without disclosing their prior exposure. If a review production becomes a subject of a malpractice claim, and if any of reviewers are named in the claim, their multiple-test history would be the subject of discovery. One could easily establish from their job applications and test histories how many times they have taken, without disclosing the same. It would be not difficult to establish that some reviewers might have discussed test questions with others in private settings. Strictly speaking, such conduct is clearly deceptive in nature and is inconsistent with the characters required for practice of law. However, they are compelled to do so and many of them indeed have the required language skills, making their deception immaterial. However, what may change the nature is that if a reviewer has been selected on the basis of a higher score from repeating the same test while another reviewer lose an opportunity. In this case, the true language skills are not an issue, but the score tainted by deception is. Even though, the native reviewers are the subjects of discrimination, they may still violate ethical rule.

G. The Testing Scheme Casts Doubt on the Credibility of Law Enforcement Actions

United States now run many enforcement programs from FCPA, export control, securities law, mail fraud, wire fraud to criminal law. Many of the regulated persons are foreign persons and foreign corporations. Whenever an investigation is concluded, the person may be ordered to pay substantial amount of fines and penalties. Most of violations happen on foreign soil or the countries where the foreign languages are spoken. All cases are decided on the documents and translations produced by the regulated persons. If the documents are not properly reviewed, and translations do not reflect the reality of what actually happened in foreign nations, the decisions would be subject to attacks. Due to the inherent difficulty to handle language problems, any of review products may have all kinds of problems. Some of the problems are inherent and unavoidable while others are caused by selection of review panels and the decisions made by the attorneys who do not understand the basics of language art. Some attorneys routinely judge translations quality by looking at how smooth and elegant the translations are. By using such criteria, they rather accept massive distortions than try to find truth. The documents of most large corporations contain highly complex subjects in many technical fields. When complex documents are in front of a non-native reviewer who lacks specific case experience, relevant technical knowledge, and relevant legal knowledge, what can the reviewer understand? The only thing the reviewer can do is to make a best guess. The unwritten rule is that if a reviewer cannot see the reason for a document to being responsive, mark it as non-responsive.

If the governmental agency enforces law based upon such foreign language review production, translations, and certification based upon such a foreign language review, its validity is questionable. Some regulated persons may escape penalties whole others may be punished unjustly. The clients may raise the question of whether its fines and penalties are justified on the basis of what have been found from their documents. It is also questionable whether a certification by a law firm after conducting this kind of foreign language review can be relied by governmental officials in making final decisions.

Conclusion

ALTA tests are uncontrolled, fraud-prone, incompetent, unrepresentative of language knowledge, and frivolous, and the mere use of the test has disparate impact on native speakers through its effects on the candidate pool, qualification distortions, altered probability due to fraud and abuse. Due to its wrong language-and-context model, ALTA tests do not gauge native fluency, but are grossly bias against native speakers. The law firms and staffing agencies responsible for the formation of this fraudulent, discriminatory, anti-competitive testing scheme should carefully review their actions and immediately cease to support this scandalous testing scheme. They should take affirmative actions to remedy past wrongs and do whatever necessary to prevent similar conduct from injuring native speakers, their own clients, non-participating law firms, and non-participating staffing agencies. ALTA should be enjoined from administering language tests for employment purposes because it has no credential in testing art and languages.