

Using language tests for upgrading knowledge: The phenomenon, source and consequences

Elana Shohamy
Tel Aviv University

Abstract

This paper is divided into two sections. The first part looks at the way in which many high stakes tests have been introduced for reasons of political expediency rather than educational benefit. It examines the consequences of using tests in this way and suggests that when tests are used as disciplinary tools their introduction leads to a narrowing of the curriculum rather than an improvement in learning. The second part of the paper examines the introduction in Hong Kong of the Language Proficiency Assessment for Teachers in light of the earlier discussion and suggests that a number of questions need to be addressed if the assessment is not to have the reverse effect to that intended by the Education Department.

Introduction

Benchmarks, standards and outcomes belong to a family of terms invented by the educational profession to indicate quality. Applying these concepts in the educational arena is most often based on the assumption that if goals are clearly set, learners will achieve them and knowledge will be gained. This paper will examine the above assumption, its source and consequences with regards to mandatory and qualification tests introduced by centralized authorities. It will then look at the proposed benchmarking of teachers in Hong Kong, now known as the Language Proficiency Assessment of Teachers (LPAT), in light of the earlier discussion.

The phenomenon: Tests as disciplinary tools

In a variety of contexts tests and other assessment measures are used as disciplinary tools; that is, they are used as means for imposing behaviors on those who are subject to them. This phenomenon is based on the assumption that if standards are set and benchmarks are defined, tests are capable of dictating to test takers what they need to know, what they will learn, what they will be taught and often how these can be achieved. The application of such a procedure is possible given the power of tests, which implies that test takers

will change their behavior according to the demands of the assessment. Test takers are willing to do so in order to maximize their scores, given the importance of the results of the tests, especially in situations when the test is of high stakes for the test taker.

Using tests as disciplinary tools is usually practiced by those in authority - policy makers, governments, as well as principals and teachers. Since it is realized that test takers will change their behavior in order to succeed on tests, those in authority will use them as a means to **cause** changes in behavior in accordance with their priorities and criteria.

The notion that tests are used to cause a change of behavior and to impose discipline has been noticed by Madaus (1990) who writes that tests are a means of social technology as they can impose behaviors on individuals and groups and thus define what learners are expected to know. He also comments on how uncritical society is with regard to such uses and how much faith it has in the scientific and technological solutions given by tests.

McNamara (1997) notes that tests have become an arm of policy reform in educational systems and in the workforce. The concern for national standards of educational achievement in a competitive global economy, together with a heightened demand for accountability of government expenditure, has propelled initiatives involving assessment as an arm of government educational policy at the national, state, and district levels.

The use of tests as disciplinary tools goes beyond the individual test taker as it can affect whole societies. Tests can manipulate educational or political systems, control curricula and re-define knowledge of communities. At the school level, tests are used as disciplinary tools when principals use school-wide exams to drive teachers to teach or when teachers use tests and quizzes to motivate students to learn and to impose discipline. But tests may be introduced nationally in order to upgrade the status and prestige of specific topics or subjects.

The source

The use of tests as disciplinary tools is possible because most tests have a detrimental effect on the lives of test takers. Thus, tests are capable of instantly creating winners and losers, successes and failures, rejections and acceptances. Since test scores are often the sole indicators for making major

decisions, their results can lead to placing people in class levels, granting certificates, determining whether a person will be allowed to graduate and continue in future studies, deciding on professions, entering special classes, getting accepted to higher education and for obtaining jobs. Thus, doing well on tests can give permission for students to graduate from universities or high schools, while doing poorly on a test can deny such a diploma. Doing well on tests entitles persons to certificates to teach, while doing poorly may prevent this. Tests, then, can open and close doors, provide or eliminate opportunities and shape the lives of individuals in many different directions. Moreover, it is often performance on a single test, often on one occasion, at one point in time, that can lead to irreversible, far reaching and high stakes decisions.

The detrimental effects of tests and the high stakes decisions they lead to are described by Noam (1996:9): “How we assess can support careers and make people successful, but it can also destroy people’s careers, place unfair burden on individuals” self perception and unnecessary hurdles in the path of their achievement”. Madaus (1990:5) emphasizes the fact that a single test provides a powerful indicator of the future of individuals as it independently triggers an automatic admission, promotion, placement or graduation decision. These decisions are non-negotiable even when there are contradictory judgments from educators about what a student knows or can do.

Test takers usually have no choice but to change their behaviour in line with the demands of a test, to comply with the test according to its requirements. It is this situation that tempts decision makers and policy authorities to use tests as disciplinary tools to cause test takers to change their behavior in line with specific criteria, standards and benchmarks. It is clearly believed that the imposition of tests will, in fact, upgrade the level of achievement of test takers along these lines.

The temptations

Tests offer a number of irresistible temptations for decision and policy makers to use them as disciplinary tools:

Tests are perceived as authoritative

Using tests for policy making has strong authoritative power as, in most cases, those who are affected by the tests tend to change their behavior in accordance with the demands of the test. Once a test becomes a means for

control supported by central bodies, it rarely faces any objection by those who are the subjects of the test, as results obtained from tests are viewed as final and absolute and are therefore rarely challenged. The main reason for the authority is that tests are administered on behalf of powerful organizations while test takers are independent individuals who are often powerless.

Tests allow flexible cut-off scores

A major temptation for policy makers to use tests is that they allow users to determine cut-off scores in arbitrary ways and thus create quotas in a flexible manner. Cut-off scores refer to the specific points that serve as the basis for decisions, such as the number of people who will pass a given test, enter special programs, obtain certificates or be granted entrance visas. The meaning of a low cut-off score is that more people pass the test while a high one means that fewer people pass and are therefore entitled to the privileges associated with the test. While a number of methods have been developed to determine cut-off scores accurately, the issue is still very controversial and unresolved. Cut-off scores are often used by those in authority as tools for “gatekeeping” those who are not wanted and for determining the number of people who, for example, will be certified as language teachers.

Tests are effective for defining knowledge

An additional temptation for decision makers to use tests is that they allow the manipulation and control of knowledge. Specifically, by including certain contents and topics on tests, the tester can re-define and introduce specific knowledge according to his/her perceptions. This is inevitable, since test takers tend to adopt and internalize their knowledge according to the requirements of the test in order to maximize their scores. In a number of studies (Shohamy, 1993, 1997; in press) it was found that the content of the tests became the *de facto* curriculum, by far more important than the existing curriculum. There is generally very little supervision by pedagogical bodies as to what content is included in tests. Sheppard (1991) shows how the decision as to what is included on a test is left to the testing experts and is often in contradiction with the knowledge as defined by the subject-matter experts. It is very tempting for those who introduce tests to use them as a means to create, re-define and control knowledge.

Freedman (1993, 1995) notes that examinations are popular among policy makers because they provide one of the few levers on the curriculum that policymakers can control. Not having tests means that teachers are free to create their own curricula, their own means and content of instruction. It is

therefore of no surprise that in situations where the status of teachers is low, decision makers choose to use tests to control learning and to dictate what will be taught in and out of the classroom, as well as to define and control the knowledge that teachers are expected to have.

Tests have strong appeal to the public, especially parents

The introduction of tests has a strong appeal to the public as it symbolizes social order in areas such as education, where the public normally feels a lack of control. Using tests is interpreted by the public as a sign of a serious and meaningful attitude towards education and as evidence of action on the part of government.

Bourdieu (1991) writes that it is frequently the case that tests serve the needs of certain groups in society, to perpetuate their power and dominance. Tests have wide support of parents who often do not trust schools and teachers. For them, tests symbolize control and discipline and are an indication of effective learning. It has often been observed that raising educational standards through testing appeals to the middle classes, partly as they equate testing with gaining access to better jobs for their children, and also because for some it is a code-word for restricting minority access (House, 1998).

Tests are useful for delivering ‘objective proofs’

Due to the authority and trust that tests have in the eyes of the public, evidence obtained from tests serves as proof for a whole list of arguments. The type of interaction taking place between the testers and test takers is based on the assumption that the tester has a neutral, non-participatory role in the testing process, while the test results provide evidence and proof. Reference is often made to: ‘The test demonstrated that you are a failure’ or ‘The test showed you did not study hard enough’. Such statements put ‘the burden of proof’ on the test results, while the tester remains a neutral observer, shedding all types of responsibilities. Flexible cut-off scores serve as useful gimmicks that testers often use when they employ tests as tools for proof.

Tests allow cost effective and efficient policy making

Using tests as disciplinary tools is a cost effective strategy for policy making. In comparison to introducing reforms through teacher training, development of new curricula or new textbooks, changing the test is a substantially cheaper venture. House (1998) notes, for example, that it was

difficult to see how new national educational initiatives could emerge under the Clinton government when the Department of Education had neither the money nor the agency capabilities to provide leadership. It was at that point that the primary reform strategy relied on introducing the standards-ending initiative. Thus, it became the center-piece of the Clinton administration.

Obtaining funds is often a long process that cannot possibly be achieved in the short time bureaucrats are in office. Tests therefore provide policy makers with a means to create policy in the shortest time.

Tests provide those in authority with evidence of action

The perception by bureaucrats, as well as the public at large, that introducing tests provides evidence of action is another temptation for using tests. For bureaucrats this offers a major advantage over other ways of making policy.

It is the realization that test takers will change their behavior in accordance with the demands of the tests that leads those in authority to introduce tests, in order to *make* those who are affected by tests change their behavior in accordance with their criteria set by the authorities. Decision makers are often strongly tempted to use tests for policy making as they view them as efficient and authoritative, guaranteeing control, while at the same time having strong public appeal, especially for parents. Tests are capable of re-defining knowledge, allowing flexible cut-off scores and are cost effective for policy making, and particularly for demonstrating action to the public.

The consequences

But, is this way of making policy, of attempting to change and upgrade knowledge, effective?

Can the introduction of tests improve knowledge?

A number of empirical studies have examined the effect of introducing tests for upgrading knowledge. In three such studies, Shohamy (1993, 1997, in press) demonstrated that the introduction of tests was capable of changing educational behavior. Yet, the nature of these changes was complex and occurred in a number of directions.

Specifically, the findings showed that teaching and learning, in all three cases, became ‘test-like’. While there was increased focus on the topic tested, the knowledge became narrow as the test content became the *de facto* curriculum. This was demonstrated in both the teaching methods and materials as there was a diversion of educational attention to content that had not been previously taught. In the case of a test of Arabic as a second language, the focus was on specific vocabulary at the expense of shortening the time devoted to the alphabet. In the case of an English oral test, attention was diverted to exclusively teaching ‘test like’ oral language tasks in the classroom. (See Li and Wang, this volume, for a discussion of this problem in China). Likewise with regards to a reading comprehension test, teachers began focussing on an area of L_1 that had not received explicit attention before, but the nature of the learning was identical to the content of the test, that is in terms of the texts and questions used for teaching. The degree of effect in each of the cases varied depending on the type of test and whether the test was of high or low stakes, the relevance of the subject to decision makers and to the public at large (certain subjects were considered more important than others), and the rate of failure (high rate of failure received more attention from the public and media). The different patterns of effects as well as the factors that contributed to the effects are described in detail in Shohamy et al. (1996).

These studies demonstrated how the quality of the knowledge created as a result of the test could be defined mostly as ‘test language’. They showed that the intention of creating changes in the teaching of English, Arabic and reading comprehension by introducing the tests, could not ensure meaningful changes in pedagogy. In fact, such pedagogical changes rarely took place. Furthermore, introducing these changes through tests did not mean that there were available resources to carry out such changes. Often decision makers relied exclusively on the tests to create the changes without guaranteeing available resources to ensure familiarity of methods to teach the content, or to adjust work loads for teachers to allow for the training needed to carry out the tasks. It became clear that pressure and sanctions alone were not enough to create a meaningful change in knowledge.

It was also shown how the introduction of the tests affected various components in the educational system much beyond what had been expected by those who introduced the tests. Even when intentions were not explicitly stated, the new tests resulted in major effects on individuals and in meaningful impact much beyond that expected. The test takers expressed fear, tension and high anxiety. Often test results were used to penalize and gatekeep test takers, although this was not the intention of the test.

Thus, the consequences of using the tests were mostly a greater focus on the topic but also resulted in unethical behavior, a re-definition of knowledge, punishment, gatekeeping, control of education and a narrowing of the knowledge. It became evident that the tests which were used instrumentally resulted in no more than a 'fast fix'. In none of the test specifications was there any serious discussion of what the topic tested meant; rather the tests were introduced as therapies for solving problems. In the case of the reading comprehension test, schools added reading comprehension hours at the expense of subject areas such as geography and history, rather than integrating reading comprehension into the subject areas. This represents a simplistic and instrumental solution. Similarly, with regards to the Arabic test, the complex problem of teaching Arabic was reduced to specific and defined components such as the teaching of the alphabet and some vocabulary items. This provides evidence that those who introduce the tests are interested in simplistic solutions, where gains can be seen immediately, in place of meaningful changes.

Such uses of tests re-create the contents in testing terms, test-like material, test-like teaching and narrow the scope of the topic being taught. The knowledge created through the tests is often referred to as 'institutionalized knowledge'; its main characteristics are that it is narrow, simplistic and often in contradiction to experts' knowledge. After all, the information included on tests is only a representation of real knowledge; it is monologic, based on one instrument, (a test), on one occasion, detached from meaningful context and usually with no feedback given to test takers for improvement.

Arguments against using standards and testing as the engine of reform have been documented repeatedly (House, 1998). It has been consistently shown how similar attempts have failed in the past. By teaching to tests, instruction is narrowed and test scores are artificially inflated to the point of questionable validity.

House (ibid) also claimed that the notion of collecting data in the belief that it would force improvement in education is common, yet very questionable. Such thinking seems far removed from how learning functions. It is similar to measuring temperature as a cure for an illness. The rationale behind such an agenda is generally that teachers and students are not trying hard enough, and that students are not sufficiently motivated to achieve in school. If one puts pressure on them through threats or failures, teachers and students will try harder and achieve more. Yet, many students cannot attain

these standards, even after extra years of instruction in special classes. And a certain percentage of the population will never attain the prescribed standard.

Turning tests into a means for change, into instrumental devices for promoting agendas, narrows the process of education. While the introduction of a test can be influential in terms of changing the focus and increasing awareness, the educational effectiveness of tests introduced in such ways cannot be very high, because the approach narrows the process of education, making it merely instrumental and not meaningful.

Freedman (1995), in studying the effect of writing tests, demonstrates how high stakes tests undermined students' work and attitudes in both obvious and subtle ways and that they did not lead to improved instruction. In a study comparing American and British writing classes, she found that in Britain, where exams were part of the system, the exam classes had to adhere to requirements that inhibited the teachers' abilities to build a coherent curriculum. The exams also restricted the amount and kind of writing the students did and, in effect, the pressure of the examinations took over. While students who were not part of the exams were motivated by their own decision-making and also by being part of a community working together, when writing for the exam, this rarely happened. These issues were even more problematic for teaching bilingual writers, who may have learned from specific corrections and by having a series of errors pointed out in detail. The conclusion reached by the teachers in the United States was that any kind of high-stakes examinations, with associated curricula, would harm their students' writing development. Freedman claimed, therefore, that a system of high-stakes examinations, even well-designed performance-based examinations, provides a flawed foundation on which to build a national educational reform movement. In fact, exams have the potential to move away rather than toward, the point we all want to reach. (Freedman, 1995:26)

Testing should follow, not lead a reform effort. With regards to the area of writing, she showed that when national exams took control of something as personal as writing, a distant examiner, rather than the teacher and students, ended up owning the writing. Moreover, when the exam was high stakes, when it had an effect on students' future lives and careers, it put formidable pressures on the classroom and the exam was particularly apt to have negative effects on the curriculum.

It is important to note that these observations were not limited to objective type tests; they also applied to performance type-tasks, whereby tests turned

the tasks into narrow objective statements, which eventually led to a narrowing of the learning. Thus, when reforms were introduced through tests, the tests were no more than a ‘quick fix’ that overlooked the need to attain meaningful comprehension and deep insight of a topic, especially with regards to an external test that could represent no more than a limited body of knowledge on any subject. Freedman concludes that “the path to curriculum reform through examinations, though tempting, remains elusive” (1995:29).

Using tests as *de facto* knowledge provides only ‘a quick fix’, an instant solution that overlooks the complexities of subject matter and is not meaningful for repair. Weiss (1977) differentiates between the instrumental impact of tests, characterized as short-range and goal oriented, and conceptual impact which is long range and meaningful, followed by discussions on the nature of the trait, methods of teaching, and agreed upon criteria of quality. In none of the tests reported above, was there any serious discussion with teachers or students about the tested topics, whether they were learnable or measurable. For bureaucrats, these simplistic and instant solutions are very attractive in that they offer instant evidence of impact in their usually short term in office. As Freire states: “The more bureaucratic the evaluators are, not just from an administrative point of view but above all from an intellectual view, the narrower and more inspection like the evaluation will be” (1985:23-24).

Re-defining de facto knowledge contradicts existing knowledge

The negative consequences of using tests in such ways are that tests are capable of redefining knowledge and that knowledge is in contradiction to existing knowledge as defined by the curriculum. Thus, the redefinition of knowledge through tests creates two parallel systems, one manifested through the curriculum or policy documents, the other reflecting bureaucratic aspirations through tests. These two systems are often in contradiction with one another. There are many examples to demonstrate this phenomenon. Bernstein (1986) refers to the two systems as primary and secondary; primary is talk, while secondary is practice, *de facto*, and more relevant since it has the enforcing power. There is therefore an ‘official story’ and a ‘real story’, which is exercised by tests and pushed by bureaucrats, and is often not known to the public. It is clearly the testing policy which is the *de facto* policy as “tests become targets for contending parties who seek to maintain or establish a particular vision of what education and society should be” (Noah and Eckstein, 1992:14).

Unethical and undemocratic ways of making policy

Using tests to create *de facto* policies can be viewed as an undemocratic and unethical procedure as it represents those in power, often not declared publicly and not discussed openly with the subjects of the tests. It is dictated from above without including those who are affected by the tests – teachers and test takers.

Implementing policy in such ways is based on threats, fear, myths and power, by convincing people that without tests learning will not occur. It is an unethical way of making policy and an inappropriate use of force in a democratic society. When tests are used to manipulate and control education, they become the devices through which educational priorities are communicated to principals, teachers and students. In the reading comprehension test mentioned above, by applying the categories of Foucault (1979), it was clear that the uses of the test led to unethical behavior as the test was used to quantify, observe, normalize, standardize, classify, punish, judge, and for surveillance.

Using tests to make policy and shape educational systems is not ethical since the test taker does not really benefit from such testing. If tests do not provide test takers with diagnostic information about his/her performance, there are no clear benefits for the test taker. In fact, in these situations it is the test taker who is paying the price for the benefits which the authorities in power gain. Test takers are being used as they become the ammunition through which bureaucratic manipulations are performed. Test takers pay the price as they undergo pressure, tension, change in behavior and anxiety. Not providing meaningful information to students can be viewed as avoiding responsibility for the effect of the exam on learning. It can be considered an unjust use of the information, as it is used for different purposes than those known to the test taker.

Thus, the negative consequences of using tests for improving knowledge have been demonstrated. It has been shown that the use of tests does not lead to improved learning or to higher achievement. It creates parallel systems in which tests become the *de facto* knowledge, often in contradiction to more sophisticated systems and it is an artificial way of granting power. It is also claimed that such use of tests is, in fact, an unethical and undemocratic way of making policy.

The case of the Language Proficiency Assessment of Teachers

The proposed introduction of the benchmarking system in Hong Kong for examining the proficiency of English teachers is a similar act to those described earlier in this paper (see Falvey and Coniam, this volume, for an overview). The benchmark tests are being introduced by policy makers in order to upgrade the level of English proficiency of English teachers. It is doubtful whether such an attempt will actually achieve its goals.

Benchmarking, with some slight variations is no different from other forms of assessment. It is imposed by authorities, such as school principals or a Department of Education, in an attempt to find out how much another group, teachers, know about a certain topic. Benchmarks define what those who are assessed should know, how they should demonstrate that knowledge, all in line with what those who introduce the measurement believe is the right type of knowledge. The proposed introduction of the benchmark tests (LPAT) in Hong Kong is accompanied by the wish to upgrade the level of English teaching. But can this goal be achieved through the introduction of these tests? Is the goal realistic?

What makes it perhaps different is that there are criteria that have to be reached and each person has to be tagged as belonging to one of these criteria. But there is no difference between these criteria and other criteria used for any test. When 20 % are benchmarked as high and 30% are benchmarked as low, there is a clear comparison of the 20% with the 30%. The passing mark or the different benchmarks are as biased and as subjective as any other criteria.

In examining the use of the benchmarks in social and political contexts, questions need to be asked regarding their introduction:

- What is the rationale and motivation for introducing the benchmark tests?
- What are the agendas and whose agendas are these?
- What will the impact be of introducing the test?
- What will its outcomes be?
- How will it affect learning and teaching?

The benchmarks are a classic example of using tests to force people to learn something that those who introduce them believe is important. The assumption is that if they cannot learn something in a natural way, those who introduce the test will force them to do it. It assumes that teachers are not motivated, that they do not think knowing English is important, and only by having pressure

put on them, in this case by the introduction of the benchmarks, will they take steps to reach the level that those who introduce the benchmarks believe is necessary.

Yet, as has been argued above, introducing tests as disciplinary tools to force test takers to improve does not work. It was argued that in a large number of studies that examined what happens to language knowledge when it is driven by tests, the results were very consistent. They showed repeatedly that when the consequences of the test could lead to important decisions (in this case, getting or keeping a job or obtaining promotion) candidates will change their behavior only in line with the demands of the test and the most they will acquire is what is known as ‘test language’.

Thus, relying on a test to create the desired change, such as upgrading the achievements in English of both pre and in-service Hong Kong English teachers, will only lead to them acquiring the English needed to pass the test. There will not be a transfer to the real language needed in the classroom.

There is no evidence to show that using the pressure of the test actually improves learning. In fact the opposite is true as it narrows the domain and restricts knowledge. The test will be successful in diverting the teachers’ attention to studying English in order to prepare for the test, given the harsh consequences for those who fail, but the type of language they will acquire will be limited and narrow and will have no use beyond the test itself. Furthermore, in a very short time the English courses and textbooks that will be developed to teach teachers English will be ‘test preparation’ courses that will include the very material that will be on the test.

It is unlikely that the desired goals will be achieved. However, it is likely that the introduction of the benchmarks will create hostility on the part of the teachers, humiliation, resentment and a feeling that they are servants of the system, that they have the responsibility of teaching but no authority, and this in an era where teachers’ organizations are gaining more power.

But those who introduced the tests will be satisfied, as they will have demonstrated to the public (and to themselves) that they have taken steps and action to change an unfortunate situation. And the public will think the benchmarks are good as they symbolize discipline, control and action. Specifically steps are being taken to solve the problem of learning of English in Hong Kong. What the public does not know is that tests are not the right solution to this problem.

But, even if it is believed for a moment that the benchmark tests can be considered to be appropriate solutions for improving the level of English in Hong Kong, and even if we assume, as the Education Department probably does, that the source of the problems of the level of English in Hong Kong is the teaching and the teachers, one needs to ask whether what is being assessed on the LPAT, language proficiency, is in fact, related to effective teaching and whether it can lead to better teaching.

To answer the above questions, we need to ask what the current views are on effective language teaching. Current theories of good language teaching inform us that there are at least three major factors that make good teachers

1. knowledge of the language
2. knowledge about the language, what is known as meta-language, and
3. knowledge of how to teach the language.

While in the early days of language teaching there was a wide-spread belief that knowledge of the language was the most important component and, in reality, the only factor that made a language teacher, there have been big changes in that area since language teaching has become a scientific discipline. In recent years there has been a decrease in the weight given to the first factor mentioned above, knowledge of the language, in favor of the other two factors. This is especially true for the English language and its central role in the world. It is clear today that there is no one type of English, and one can speak of a Singaporean English and a Nigerian English, an Israeli English and a Hong Kong English.

Consequently, in most teacher education programs the emphasis is on developing knowledge of good teaching methods as well as good meta-language. Good teaching methods include also how to become a facilitator, how to listen to tapes of English language. There is no doubt that the English teacher needs to know how to use the language, but a much lower level of English proficiency than the one previously believed, is sufficient. Similarly, with regards to meta-language, this is an area where native speakers often lack knowledge. Since they acquired the language in a natural way, they are not familiar with the pedagogical rules of the language that they need to know in order to teach it. In fact in some teacher training programs native and non-native teachers go through a very different program. But it is important to understand that the most important component in effective language teaching is the methods of teaching it.

Yet, the benchmark tests emphasize almost exclusively knowledge of language proficiency, the component which is considered least important in its contribution to effective language teaching. There are therefore two issues that need to be raised; the inability of tests to upgrade the quality of learning, and the doubt that measuring language proficiency is central to effective English teaching.

A number of questions need to be posed in relation to the introduction of the LPAT in order to avoid the major obstacles of introducing tests as disciplinary tools outlined above:

- How will the benchmark tests be used and what will be their consequences?
- Is the construct of effective language teaching used in the LPAT sufficient?
- How do teaching methods relate to the quality of language?
- Will they actually lead to improved learning?
- What are the difficulties that exist in arriving at the criteria for the benchmarking? (this issue is still extremely sensitive with regards to test cut-off scores, and it is bound to be even more complex with regards to benchmarking)
- Have the motivation and intentions, especially the political ones, that brought about this project been examined carefully?
- Will the impact of the LPAT on the whole language system, in fact lead to improved teaching?
- How will the results be used? Is there a possibility that they will lead to detrimental decisions for individual teachers without justification?
- Will the LPAT be relied on as the single factor for assessing the quality of teaching?
- Will the LPAT be relied on as the only means for improving English teaching and the level of English in general in Hong Kong?
- How will the LPAT affect teacher training institutions? Will they, as a result of the benchmarks, pay less attention to methods of teaching, which is their mandate, and more attention to language proficiency?
- Will other creative methods be employed for affecting the upgrading of language teaching in Hong Kong?

- How will the LPAT affect the self confidence of the teachers?
- Can strategies of a less powerful nature be introduced, such as diagnostic and constructive components?
- What are the rights of the test takers, that is the teachers, to bring counter evidence to that obtained by the LPAT? And what are their attitudes (especially of practicing teachers) to this kind of system?
- What additional evidence will be counted as valid if there is disagreement with the results of the LPAT?
- What are the responsibilities of those who are developing the LPAT to ensure the ethical uses of what they have delivered to the Education Department?

In conclusion, and with regards to the discussion in this paper, the introduction of the language proficiency assessment of teachers (LPAT) requires close re-examination and re-consideration as it cannot be assured that it will lead to the outcomes expected by those who introduced it. Additionally, there may be further side effects in the form of fear, anxiety and decreased trust in the system and in the profession. All these should be examined, deliberated and, if need arises, remedied.

References

- Bernstein, B. (1986). On pedagogical discourse. In: Richardson, J. (ed.) *Handbook of Theory and Research for the Sociology of Education* (pp.205-240). New York: Greenwood Press.
- Bourdieu, P. (1991). *Language and Symbolic Power*. Cambridge MA: Harvard University Press
- Falvey, P. & Coniam, D. (2000). Establishing writing benchmarks for primary and secondary English language teachers in Hong Kong. This volume.
- Foucault, M. (1979). *Discipline and Punish*. New York: Vintage Books.
- Freedman W. S. (1993). Linking large-scale testing and classroom portfolio assessments of student writing. *Educational Assessment* 1(1):27-52.

- Freedman W. S. (1995). Exam-based reform stifles student writing in the U.K. *Educational Leadership*, 52 (6):26-29.
- Freire, P. (1985). *The Politics of Education*. South Hadley, MA: Bergin and Garvey Publishers, Inc.
- House, E.R. (1998). *Schools for Sale: Why Free Market Policies Won't Improve America's Schools, and What Will*. New York, NY: Teacher College Press.
- Li X. & Wang, Y. (2000) Testing oral English on a mass scale: is it feasible? The oral component of the MET in China. This volume.
- Madaus, G. (1990). Testing as a social technology. The Inaugural Annual Boise Lecture On Education and Public Policy, Boston College, MA
- McNamara, T. (1997). Policy and social considerations in language assessment. In: Grabe, B. (ed.) *Annual Review of Applied Linguistics*, 18: 304-319.
- Noah, E., & Eckstein, M. (1992). Comparing national systems of secondary school leaving examinations. In: Noah, E., and Eckstein, M. (eds.) *Examinations: Comparative and International Studies* (pp. 3-22). Oxford: Pergamon Press.
- Noam, G. (1996). Assessment at a Crossroads: Conversation. *Harvard Educational Review*, 66:631-657.
- Shepard, L. (1991). Psychometricians' beliefs about learning. *Educational Researcher* 20 (7):2-9.
- Shohamy, E. (1993). *The Power of Tests: The Impact of Language Tests on Teaching and Learning. NFLC Occasional Papers*. Washington DC: The National Foreign Language Center.
- Shohamy, E. (1997). Testing methods, testing consequences: are they ethical? Are they fair? *Language Testing* 14 (3):340-349.
- Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation* 24 (4):331-345.

Shohamy, E. In press. *The Power of Tests: A Critical Perspective of the Use of Language Tests*. London: Pearson.

Shohamy, E., Donitsa-Schmidt, S. & Ferman, I. (1996). Test impact revisited: washback effect over time. *Language Testing* 13(3):298-317.

Weiss, C.H. (ed.) (1977) *Using Social Research in Public Policy Making*. Lexington, MA:Heath