מרכז ארצי לבחינות ולהערכה (ע"ר)
NATIONAL INSTITUTE FOR TESTING & EVALUATION
المركز القطري للامتحانات وللتقييم
מיסודן של האוניברסיטאות בישראל
ת"ד 26015 ירושלים 91260 טל' 02-6759555
http://www.nite.org.il , e-mail:public@nite.org.il

מוזיאון המדע ע"ש ברנרד בלומפילד ירושלים (ע.ר)
متحف العلوم على اسم بلومفيلد القدس
Bloomfield Science Museum Jerusalem

Progress Report

Development of an Exhibition on Testing and Measurement Following the International Workshop, October 2009

1. Background

A proposal put to the Bloomfield Science Museum Jerusalem by the National Institute for Testing and Evaluation (NITE) to create an exhibition on testing and measurement resulted in the establishment of a joint work team. With the aims of examining an initial proposal formulated by the team during 2009 and determining future directions for development, an international workshop was convened in Jerusalem, facilitated by a grant from the Yad HaNadiv Foundation.

The workshop was held on October 20-22, 2009, in Jerusalem at the Bloomfield Science Museum and at NITE. Participants included members of the work team, members of the project's steering committee, partners from the Franklin Institute Science Museum in Philadelphia, U.S.A., guests from ETS (Educational Testing Services), and from NAGB (National Assessment Governing Board), representatives from The National Authority for Measurement and Evaluation in Education (RAMA), representatives from the Israel Center for Medical Simulation (MSR), and representatives from the Yad HaNadiv Foundation. A full list of participants can be found in appendix 1.

1.1 The goals of the exhibition

During the development process, the following goals were formulated:

- **Exposing the public to the scientific field of testing and measurement.** Testing and measurement constitute an important scientific field with a long history. Regarded differently by different cultures, the field is founded on various theories, has a unique methodology, and a variety of applications. It is a growing field that has not yet received much public exposure.
- **Creating dialogue between the public and measurement experts.** The public's attitude to testing is usually negative. The exhibition is aimed at encouraging dialogue between measurement experts and the public, especially with regard to the connection between testing and society.
- **Creating educated clients.** In today's society, individuals are constantly required to take tests. The exhibition will convey the importance of professional testing: accuracy, standardization and fairness. This is meant to influence the public and require that test developers maintain high standards.

2. Workshop Discussions

2.1 Target audience

The exhibition's target audience was initially defined as students aged 13 and up, youth and adults (assuming that some children will visit the exhibition with their schools). A suggestion was made in the international workshop to extend the target audience to children aged 10 visiting the museum with their families. This will make it possible to cater to the families who are the museum's most regular visitors, and to define specific topics and activities for youth and other interested visitors.

2.2 The main concepts of the exhibition

The main concepts that the exhibition will attempt to convey were discussed during the workshop. Fine-tuning and clarification of these concepts is needed for the exhibition to be effectively curated.

The three main concepts discussed in the workshop were:

- **The many forms that tests take**
- **A good test teaches us new things about ourselves**
- **Culture, time, and context and their impact on testing**

In addition to these main concepts (Big Ideas) it was suggested that the exhibition will feature a testing lab dealing with the question: **what is a good test?**

2.3 Detailed exhibits

Two detailed exhibits were presented in detail at the workshop (A judge is born, Imperial examinations). The exhibits and their suitability for the exhibition were discussed.

2.4 Discussion of sensitive issues

Sensitive issues and ways to present them in the exhibition were discussed during the workshop:

- **Intelligence.** The main message: Intelligence is a multi-dimensional and multi-faceted concept. The aim of this concept is to explain certain phenomena, which are affected by time and place. It was suggested that the culturally-dependent factors in testing be presented: tests are dependent on culture and period. It is worthwhile using the intelligence issue to show that tests only useful within the context of culture and period.

- **Cheating, copying.** The main message: A good test does not enable cheating. Measurement experts are supposed to prevent cheating or detect it after the fact. It was suggested that cheating not be presented as a separate topic, but in the context of another topic such as fairness, validity, or score interpretation. Furthermore, because it is difficult to define what constitutes cheating and copying, the cultural differences in the definition of cheating must be addressed. It is possible to show various cheating methods without passing judgment on them and without conveying a didactic message regarding cheating.

- **Test anxiety.** The main message: Some of the workshop participants were opposed to addressing the topic of test anxiety in the exhibition. They believe that the exhibition should be enjoyable and that presenting test anxiety may cause visitors to develop it. Conversely, teachers who attended as session held about a month before

the workshop requested that the exhibition deal with the topic. The proposal formulated in the workshop was to include this topic in the educational activities offered to schoolchildren on class visits.

- **Gender and cultural differences.** The main message: The issue of gender is sensitive and controversial. However, provocative and controversial issues are not to be shied away from. These issues could actually improve the exhibition. Their underlying message is that every test has biases and a good test is not biased towards any specific gender or culture.
  Suggested exhibit: Gender detector: A short test made up of questions biased towards men, and other questions biased towards women. Visitors will get a computerized output that will guess their gender according to their answer patterns.

2.5 Exhibition size and participating partners

2.5.1 The size of the exhibition and number of exhibits

The intention is to mount the exhibition at the participating museums (the Bloomfield Science Museum Jerusalem, and the Franklin Institute Science Museum), and then at other museums. The size of traveling exhibitions in the U.S.A. ranges from huge exhibitions intended to draw a large number of visitors (over 700 square meters, or roughly 7,500 SQF) to medium size exhibitions (250-500 square meters, or roughly 2,700-5,400 SQF).
It was decided to examine the interest level at museums of various sizes. The size of the exhibition will be determined on the basis of their responses (the size of the space dedicated to the exhibition, number of exhibits, budget and fundraising methods). The museums' reactions will also determine whether two versions of the exhibition are needed, or one that will travel between the two partner museums and later to other museums.

2.5.2 Participating partners

The workshop participants (the Bloomfield Science Museum, the Franklin Institute Science Museum, NITE and ETS) are partners in the project.
In order to consolidate the collaborative process, the abilities of each partner (concept building, writing, pilot exhibit building, design, educational activities, mobility etc.) should be assessed. A work plan, which will include assignment division, timetables and an initial budget, should be agreed upon.
It is necessary to formulate a fundraising plan for the development and construction of the exhibition right from the start. The participation of other museums and institutions in the project is contingent on their ability to contribute to funding it.

3. Following the workshop

3.1 Suggested content and exhibit list

After agreement on the main concepts of the exhibition was reached, the Israeli development team began writing the theoretical background material that will be the basis for developing the exhibits. In addition, the topics that will be used to examine the main concepts were selected:

- To present the concept of everyday tests and attitudes towards testing in general, we decided to use the situation of choosing a spouse or partner. The topic: **How do we choose a spouse/partner?**

- To present the purpose of tests, we chose the example of testing job compatibility using medical simulations. This kind of exhibit was chosen because it has the potential to draw a large number of visitors. The topic: **What a great job!**
- To present the different aspects of tests in various cultures, time periods and contexts, we chose the question: how is talent tested/how can we test talent. The topic: **Are you a talented …?**
- A **test lab** section will focus on the professional process of test building: defining the attribute to be tested, building the testing tool, checking its validity, and examining its compatibility with diverse populations.

A table presenting the main concepts and topics is presented below. Each topic includes a list of exhibits, most of them discussed in the workshop. The exhibits are in various stages of development.

Table of Exhibits			
Topic	**Main concept**	**Exhibits**	
How do we choose a spouse/partner?	**There are many forms of tests**	**The exhibit**	**Related to-**
		Barnum effect (astrology)	Construct validity
		Polygraph (with partner)	Measurement reliability
		Popular tests: test yourself	Content validity and construct validity
		Interviews	Construct validity and reliability problems
		Measurement bias: love at first sight, ice cream machine, telephone conversation	Reliability problems/ measurement error
What a great job!	**A good test helps us discover new things about ourselves**	**The exhibit**	**Related to-**
		Testing hand-eye coordination (mechanical skills: sewing, dentist)	Building a battery of tests – assessment centers, which include, among other things, performance tests
		Medical simulations	
		Professional inclinations	
		Team work	
		Decision making	
		Job analysis	
Are you talented?	**Culture, time period and context influence tests**	**The exhibit**	**Related to-**
		Old intelligence tests	The effect of period and culture on the definition and measurement of a theoretical concept
		Culturally-dependent intelligence tests	Measurement error/bias
		Raven's Progressive Matrices – shape tests	Neutralizing cultural differences (to some extent)
		Women can't drive	Gender differences and their origins
		Automated essay scoring	Tests in the future
		The ancient Chinese Imperial examination system, emphasizing intelligence and leadership skills	tests in the past, multi-method measurement
Test lab	**What is a good test?**	**The exhibit**	**Related to-**
		A Judge is Born	Inter-rater correlation
		False weight	Measurement reliability
		Assessing copying/cheating	Measurement reliability
		Psychometric tests	Predictive validity
		What is my average?	Interpersonal differences
		Adaptive tests	Test technology
		Building test items	The process of building a good test item
		Teachers' exhibit	Formative assessment
		Gender detector	Predictive validity and content validity

3.2 Educational activities

Several topics suitable for educational activities emerged as we consolidated the list of exhibits:

- Preparing for the next test (what happens to the brain during a test, test anxiety)
- Choosing a career (in addition to the exhibit dealing with this topic)
- Tests in the past, present and future

3.3 Initial contact with museums in the U.S.

In order to assess how much interest the exhibition could generate from the different museums and to assess their needs in regards to the size of the exhibition, the exhibition was presented in personal conversations at the ASTC (Association of Science Technology Centers) conference in Texas, at the end of October 2009.
The museums' response was that the topic of "testing" was not attractive. Interest in the exhibition was only sparked when the idea of a wider view of testing – not just school tests – was presented, and when we explained that the intention is to incorporate medical simulation tests. Testing and measurement are relevant and important topics for testing experts and the educational system, but museum people need an attractive selling point in order to draw the crowds.

Agreements with the Franklin Institute

At a meeting between Maya Halevy (Bloomfield Science Museum), Troy Collins and Steve Snyder (Franklin Institute) several points of agreement were reached:

- The exhibition needs a different, more attractive name that will represent a wider view of the concept of testing.
- The exhibition is not suitable for large museums because they are interested in huge and profitable exhibitions drawing hundreds of thousands of visitors.
- The exhibition may be appropriate for medium-sized museums that are interested in inexpensive exhibitions (there are many such museums).
- Initial documents must be prepared as soon as possible in order to begin fundraising. This should be done at this point in the development process.
- The Franklin Institute team will be responsible for raising funds from foundations in the U.S. because some of theses foundations allocate funds only to U.S. institutions.

Conclusions

- The total floor area of the exhibition should be about 200-300 square meters.
- Full funding for the development and production of the exhibition must be obtained (because rental fees cannot be depended upon).
- The exhibition's rental fees should not be too high so as to remain attractive to the relevant museums.
- It seems appropriate to build only one version of the exhibition, which will first be exhibited at the Bloomfield Science Museum, then at the Franklin Institute and later will begin traveling in the U.S. A second version will only be built if demand warrants.

3.4 The next steps

3.4.1 A fundraising document

- The Israeli team will prepare a comprehensive document about the exhibition, which will serve as the basis for a fundraising document.
- The Franklin Institute will prepare the basic structure of the budget and this will be added to the fundraising document.
- An initial list of foundations and institutions that can be approached regarding fundraising will be made together with Alan Friedman from NAGB (based on the workshop's summary and the comprehensive document)

3.4.2 Work plan

The Franklin Institute will prepare the framework for a work plan: phases, tentative time table (it is estimated that the process of building the exhibition will take up to two years). Furthermore, every institution will present its abilities to facilitate division of labor among them.

3.4.3 Engaging the scientific community

In coordination with the Bloomfield Science Museum, the National Institute for Testing and Evaluation will continue its attempts to generate interest in the exhibition and its contents, within the relevant scientific community in Israel and around the world. The exhibition was presented at the AEA-Europe conference (an educational measurement conference) that took place in Malta in November 2009, and generated some interest. The exhibition will also be presented at the Israeli Psychometric Association conference in February 2010.

In August 2009, a proposal to present the exhibition at the NCME (National Council for Measurement in Education) conference in the U.S. was submitted and accepted.
The exhibition will receive a lot of attention at this annual conference (This time in Denver, Colorado), which brings together nearly a thousand measurement and assessment experts (mainly from the U.S.) for five sessions. During the conference, the exhibition will be presented in a 90-minute symposium. It is expected that first rate measurement experts from around the world will participate in the symposium, along with representatives of the project's development team and the participating institutions.

The steering Committee:
Diana Alderoqui Pinus, Avi Allalouf, Maya Halevi & Yoav Cohen

The exhibition team:
Avivit arvatz, Anat Bleich, Hanan Cohen, Tali Frenkel, Hagar Lerman, David Ziegler.

December 8, 2009

Appendix 1 – Participants' List

Name	Affiliation
Bennett, Randy, E.	ETS – Educational Testing Service
Cizek, Gregory, J.	University of North Carolina, NAGB
Collins, Troy, M.	The Franklin Institute, Philadelphia
Snyder, Steven L.	
Beller, Michal	RAMA – National Authority for Educational Measurement and Evaluation
Rapp, Yoel	
Ziv, Amitai	MSR- The Israel Center for Medical Simulation
Shlissel, Miri	Yad Hanadiv
Darmon, Avital	Director at initiative for applied education research
Ben Ami, Uriel	Ormedia
Alderoqui Pinus, Diana	Bloomfield Science Museum, Jerusalem
Ben Shalom, Amir	
Brokman, Dea	
Cohen, Hanan	
Halevy, Maya	
Hillman, Peter	
Snir, Nurit	
Allalouf, Avi	NITE – National Institute for Testing & Evaluation
Arvatz, Avivit	
Bleich, Anat	
Cohen, Yoav	
Frenkel, Tali	
Gafni, Naomi	
Lerman, Hagar	
Ziegler, David	