



2011
A

EXHIBITION ON TESTING & MEASUREMENT

Although tests play an important role in the lives of people all over the world, the public is for the most part unaware of the principles of development and use that ensure effective and fair tests. Collaboration between experts in the field and in science museums will facilitate making this information accessible to the general public in a friendly manner.

The National Institute for Testing and Evaluation (NITE) and the Bloomfield Science Museum of Jerusalem (BSMJ) are currently developing an exhibition on the various educational and psychological aspects of testing and measurement. A development team consisting of experts from both institutions formulated an initial proposal for this innovative exhibition. The exhibition will open at the Bloomfield Science Museum with subsequent - or concurrent openings in other countries.

OBJECTIVES

- To present the field of testing and measurement to the general public. The field, which has not had much public exposure, has a broad theoretical grounding and employs unique methodologies in diverse contexts. The exhibition will stress the importance of professional test development on the basis of scientific principles.
- To create dialogue between the public and measurement professionals. Public attitudes to testing are for the most part negative. The exhibition will encourage dialogue, particularly in relation to the impact of tests on society (for example on social mobility). Visitors will be asked to consider the influence of testing on their lives and express opinions with regard to test use.

TARGET POPULATION

The exhibition is designed for children (age 10 and up), and adults. We also hope to reach teachers and students visiting the museum with their schools.

MAIN THEMES

The exhibition will be organized around three main themes:

- **Reliability** In scientific terms, reliability is consistency over time. In educational measurement and psychology, reliability refers to consistency of performance on the test, which is an indicator of the test's quality.
- **Validity** In scientific terms, validity is the degree to which a concept, inference or measurement accurately reflects reality. In testing, validity is the degree to which the test indeed measures what it is supposed to measure. There are different kinds of validity: content validity, construct validity, criterion-based validity and face validity.
- **Fairness** A test with high reliability and validity is fair. A test must be fair across different population groups. Considerations of fairness affect both the construction and use of the test.



HISTORY OF TESTING

The exhibition will present milestones in the history of testing in the form of posters, old tests, interviews with experts, etc. Among the topics that will be covered are: testing in Imperial China, IQ tests, projective personality tests, military screening tests, the development of assessment centers, and the history of computerized tests.

SOCIAL ASPECTS

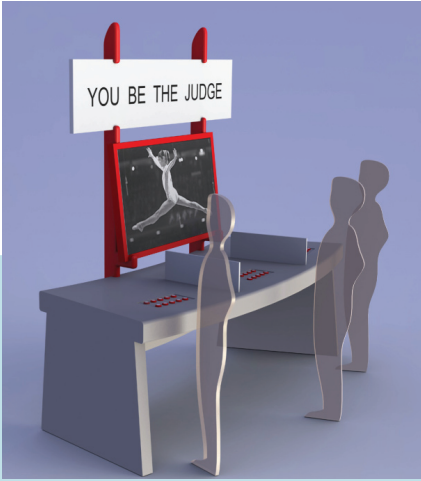
Tests can be a first-rate tool for narrowing gaps, giving people a second chance, and promoting social advancement in general. The exhibition will familiarize visitors from all strata of society with the world of testing and measurement, by means of an enjoyable and informative encounter with a variety of tests. These include educational tests, professional tests (admissions, qualifications and licensing) and other areas of ability, e.g. motor or musical. The encounter with testing by way of the exhibition has the potential to make the world of testing more accessible and familiar and thus reduce the anxiety associated with tests.



SAMPLE TOPICS AND SUB-TOPICS

| | |
|----------------------------------|---|
| History of testing | What was the nature of the imperial examinations in China? |
| Reliability | Will you get the same score from two raters? Is the test accurate? |
| Validity | How do we know if the test measures what it intends to measure? |
| Test construction | How are standardized tests (like the Psychometric) constructed? |
| Intelligence | What is intelligence and IQ? Can I improve my Intelligence score? |
| Psycho-physiological measurement | How does a polygraph work? |
| Psychological assessment | How do personality tests work? |
| Behavioral assessment | What can we learn from a medical simulator? |
| Educational measurement | How does my country rank in international comparisons? |
| Vocational assessment | Will an individual make a good medical doctor? |
| Testing and society | What is the impact of testing on society? |
| Gender and assessment | Is the test fair to both genders? |
| Fairness and bias | Is the test fair to different population groups? |
| Adaptive testing | How can a test be tailored to the ability of specific examinee? |
| Technology | Can a computer assess writing skills? |
| Measuring change | How can we assess the progress of a pupil in a class? |
| Test coaching | What is the effect of coaching on aptitude test scores? |
| Popular tests | "Who wants to be a millionaire" - good test or good entertainment? |

SAMPLE EXHIBIT: YOU BE THE JUDGE

| | | |
|------------------------------|--|---|
| Physical Structure | A table with three touch-screen computer terminals, separated by partitions. |  |
| Invitation | Would you like to judge an Olympic sporting event? | |
| Visitor Participation | Visitors will be asked to score an athletic performance (video) and compare their scores. Initially they score without instructions. Then they are given the dimensions of a successful performance as determined by professional raters. The visitors then watch the video again and give a separate score for each dimension. Agreement between scores on each dimension will be illustrated by means of numbers and user-friendly graphs. | |
| Challenge | To give as professional and fair a score as possible | |
| Messages | Inter-rater agreement is one of the conditions for reliability. A high correlation between raters is achieved by means of an agreed-upon rating scale (scoring instructions and criteria that have been checked in advance) that relates to each aspect of the behavior being tested. | |
| Implicit Message | Performance Assessment is difficult because the raters' attention is on a complex situation. It is to be expected that the raters give different scores. A large discrepancy between the scores given can make the evaluation appear unfair. An analytic rubric (which breaks the behavior being evaluated down into different dimensions) is extremely useful in narrowing the gap between rater's scores. | |

MEETINGS OF THE SCIENTIFIC COMMUNITY

In October 2009, an international planning workshop was convened in Jerusalem. It was attended by members of the development team, the steering committee and representatives of the partner organizations: the Franklin Institute in Philadelphia (TFI) and the Educational Testing Service (ETS). In addition, there were representatives from the U.S. National Assessment Governing Board (NAGB), the Israeli National Authority for Testing and Measurement (RAMA), and the Israeli Center for Medical Simulations (MSR).

The workshop addressed: 1) the objectives and scope of the exhibition (target population, physical size); 2) the contents of the exhibits and the visitors' experience of them; and 3) the nature of collaboration between the institutions.

The concept of this exhibition has been presented at various scientific conferences: AEA-Europe (Malta, 2009); NCME (Denver, Colorado, May 2010); ITC (Hong King, July 2010); IAEA (Bangkok, August 2010). The concept was presented with assistance from our U.S. partners, the Franklin Institute and ETS, and some global authorities in the field.

PARTNERSHIP

In December 2010, the four partners : NITE, BSMJ, ETS and TFI submitted a funding proposal to the National Science Foundation (NSF), informal education track.



FUTURE PLANS

The development team is advancing the project, with the assistance of experts in various fields. The topics being presented are undergoing review in cooperation with interested parties such as teachers and students.

INTERESTED PARTIES

Science museums, educational institutions, organization and individuals dealing with testing and measurement, and para-educational organizations are encouraged to contact us. Fundraising in support of the exhibition will commence shortly.

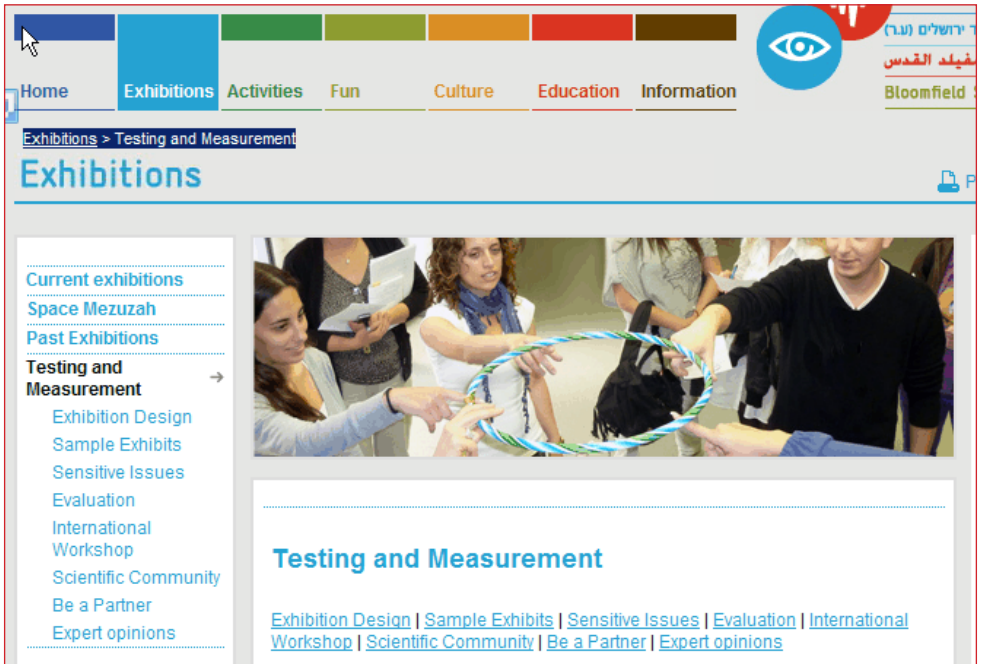
Steering Committee

Diana Alderoqui-Pinus
Avi Allalouf
Maya Halevi
Yoav Cohen

Development Team

Anat Bleich
Hanan Cohen
Hagar Lerman
Nahum Shemi
David Ziegler

Visit the exhibition website:
www.mada.org.il/testing



The screenshot shows the website's navigation bar with links for Home, Exhibitions, Activities, Fun, Culture, Education, and Information. The 'Exhibitions' link is highlighted. Below the navigation bar, the breadcrumb trail reads 'Exhibitions > Testing and Measurement'. The main heading is 'Exhibitions'. On the left, a sidebar menu lists 'Current exhibitions', 'Space Mezuzah', 'Past Exhibitions', and 'Testing and Measurement' (which is selected and has a right-pointing arrow). Under 'Testing and Measurement', the following items are listed: Exhibition Design, Sample Exhibits, Sensitive Issues, Evaluation, International Workshop, Scientific Community, Be a Partner, and Expert opinions. The main content area features a photograph of people interacting with a colorful, circular, woven object. Below the photo, the section is titled 'Testing and Measurement' and contains a list of links: [Exhibition Design](#) | [Sample Exhibits](#) | [Sensitive Issues](#) | [Evaluation](#) | [International Workshop](#) | [Scientific Community](#) | [Be a Partner](#) | [Expert opinions](#).

Contact info

NITE, Avi Allalouf: avi@nite.org.il
BSMJ, Diana Alderoqui-Pinus: dianap@mada.org.il