

TESTING INTERNATIONAL

Vol. 27, July, 2012 Editor: Jan Bogg



PRESIDENT
Prof. John Hattie
University of Auckland
New Zealand

PRESIDENT-ELECT
Prof. Fanny Cheung
Chinese University of Hong Kong
Hong Kong

SECRETARY
Prof. Dave Bartram
SHL Group, UK

TREASURER
Prof. Em. Barbara Byrne
University of Ottawa, Canada

PAST-PRESIDENT
Prof. Marise Born
Erasmus University Rotterdam
Netherlands

COUNCIL MEMBERS
Elected Members
Prof. Frederick Leong
Michigan State University, USA
Prof. Kurt Geisinger
Buros Center on Testing & University
of Nebraska, USA
Prof. Dragos Iliescu
SNSPA Bucharest, Romania
Aletta Odendaal
University of Johannesburg
South Africa

Dr. Solange Wechsler
Pontificia Universidade Catolica de
Campinas, Sao Paulo, Brazil
Co-Opted Members
Dr. David Foster, Kryterion, USA
Prof. Ron Hambleton, University of
Massachusetts, USA
Observers
Prof. Marise Born, Erasmus University
Rotterdam, Netherlands
Prof. Paula Elosua, Universidad del País
Vasco, San Sebastián, Spain

REPRESENTATIVES
IUPsyS Representative
Dr. Merry Bullock
American Psychological Association,
USA
IAAP Representative
Prof. Tom Oakland
University of Florida, USA

EDITORS
International Journal of Testing
Prof. Stephen Sireci
University of Massachusetts, USA
Prof. Rob Meijer
Groningen University, Netherlands

Testing International
Dr. Jan Bogg
University of Liverpool, UK

CONTENTS

YOUR ITC COUNCIL

2

ARTICLES

In Focus ITC Scholars 2012

2

ITC Scholars Articles on Testing in Their Country

2

Psychological Testing in the Ukraine: Peculiarities, Goals and Perspectives Khrystyna Rakhubovska, Ukraine

3

Testing in Belize

Betty-Jean Usher-Tate, Belize

4

Application and Development of Testing in China

JiETING ZHANG, MINQIANG ZHANG, WENYI ZHANG, China

6

Test Translation, Adaptation and Development: The Situation in Pakistan

Sobia Aftab, Pakistan

8

Gender Differences in Mathematics and Science Performance in Malawi

Ella Banda, Malawi

9

The Structural Validity of the Inductive Reasoning

Developmental Test for the Measurement of Developmental Stages

Hudson F. Golino Cristiano & Mauro Assis Gomes, Brazil

10

Current Developments of Clinical Assessment in Bulgaria

Zornitsa Kalibatseva, United States

11

The Philippine Educational Measurement and Evaluation Association

Ronnel B. King, Hong Kong

13

NEWS

14

President Elect Fanny Cheung Honoured with Prestigious APA Award



ITC President Elect, Professor Fanny Cheung is co-recipient of the 2012 APA Award for Distinguished Contribution to the International Advancement of Psychology, she will give an award address at the 2013 APA convention in Honolulu. Two

other Council members Fred Leong and Tom Oakland are both former recipients of this international award. The APA has produced an article on her contribution to international psychology

<http://www.apa.org/international/pi/2012/06/cheung-landis.aspx>

Teach Your Children Well: A Conference to Honour Ron Hambleton

November 9-10, 2012
University of Massachusetts
Amherst
USA



Registration deadline: September 1, 2012
Call for posters, registration information, invited speakers and other information can be found at www.umass.edu/ronference

Interested in contributing to our Committees?

The ITC has 6 Committees, each chaired by a Board member and include non-Board ITC members:

- *Publications, Communications & Marketing* - Chair Fred Leong
- *Membership, Involvement & Marketing* – Chair Dave Foster
- *Conferences Chair*—Fanny Cheung
- *Research & Guidelines* – Chair Dragos Iliescu.
- *Policies* – Chair Tom Oakland
- *Outreach* – Chair Ron Hambleton

Contact John Hattie for further information
jhattie@unimelb.edu.au

The work of the committee's will next be reported on at the AGM during the Amsterdam conference.

ITC Scholars 2012

The 2012, early career scholars (under 35) will be present at the 8th conference of the International Test Commission in Amsterdam. The scholarships enable recipients to attend ITC conferences and provide an opportunity to internationally network.

If you are attending the 2012 ITC conference in Amsterdam www.itc2012ams.com look for their posters or presentations.

Profile of the 2012 Scholars

Jieting (Jenny) Zhang has completed postgraduate and doctoral programs study in psychological measurement and statistics, at South China Normal University, Guangzhou, China and is currently undertaking a study visit to Penn State University from Sep. 2012 to Sep. 2013. Her research interests are in Latent class modeling, categorical data analysis and classification testing. Jenny's ongoing research includes college students' consuming behavior in Guangzhou, China and a comparison of methods of diagnosis of college students' mental health status.



Sobia Aftab obtained a PhD in Clinical Psychology from the Institute of Clinical Psychology, University of Karachi-Pakistan. She works as an Assistant Professor at Institute of Clinical Psychology (ICP), University of Karachi, Pakistan and is a Practicing Consultant

Clinical Psychologist. She is the Member Editorial Board for the "Pakistan Journal of Clinical Psychology" and "Pakistan Journal of Psychology" published by ICP. Her research focus is in the field of Clinical Psychology, especially various mental health issues, including Depression; Attributional Style; Cognitions; Suicide; Self-Esteem; Social Support; Marital Adjustment. Her other research interests is Psychometrics and she is working on the adaptation and development of different scales in local language.

Ella Banda has a Masters degree from the University of Massachusetts. She worked as a secondary school teacher, teaching Mathematics and Biology, prior to joining the Malawi National Examinations Board (MANEB), as a Research and Test Development officer. Ella's job involves developing and administering standardized national examinations for primary and secondary schools, as well as research in assessment and she has applied to study for a Ph.D. Her research inter-



ests are in item response theory and structural equation models.



Khrystyna Rakhubovska is from Kiev, Ukraine. She obtained a Masters degree in Psychology at the National Taras Shevchenko University of Kiev in June 2011. During her studies at University she adapted the Test of Motivational Orientation (TOM), the Big Five Questionnaire-2 (BFQ-2) and the Poppleton Allen Sales Aptitude Test 2000 (PASAT 2000). Currently a postgraduate student and an R&D manager for a test publishing company in the Ukraine, her main interests are in test development and adaptation and cultural differences.

Betty-Jean Usher-Tate is from Belize, the only English speaking country in Central America, with a population: 350,000. Betty travelled to the USA to study Early-Childhood Education and returned to Belize with an Associate Degree, then completed an undergraduate degree in Secondary Education. She became an English Teacher at Ecumenical Alma Mater High School, then taught at the University of Belize. Betty went on to become Vice-Principal of Delille High School. In Belize, high school admission is linked to scores from a national exam. When her government expressed a need for a Belizean Psychometrician, Betty-Jean relocated with her family to the USA to pursue a PhD in Psychometrics. Her current areas of interest include high-stakes testing, test-fairness, testing special populations, and validation.



Hudson F. Golino is a Master's student in Developmental Psychology and has a Bachelors degree in Psychology from the Universidade Federal de Minas Gerais (UFMG), Brazil. He is currently working at the Laboratory for Cognitive Architecture Mapping, UFMG and has collaborated with the Dare Institute, Cambridge, MA, USA, since 2008. Hudson has been working on the construction and validation of psychological and educational instruments. He is interested in the organization and development of the mind's architecture, developmental stages, cognitive change and psychometry.

Zornitsa Kalibatseva is from Plovdiv, Bulgaria. She has a Master's degree in Clinical Psychology from Michigan State University on assessment of depression among racial and ethnic minorities in the U.S. She is currently a postgraduate Ph.D student, looking at the influence of cultural factors on the experience and expression of depression. Zornitsa will co-chair a *symposium during the ITC Conference* on 'Cross-



cultural assessment of psychopathology' and will present her Master's research during the symposium. Her interests are in the assessment of cross-cultural psychopathology and in examining the current state of clinical assessment in her Country.

Ronnel B. King is currently a Ph.D. candidate in educational psychology at the Faculty of Education, The University of Hong Kong. He obtained his bachelor's degree in psychology from Ateneo de Manila University (Philippines) with first class honors. His doctoral dissertation focuses on the investigation on how different types of goals motivate students to be engaged in school and how they facilitate psychological adjustment and well-being. His research interests are in the investigation of student motivation and well-being and in the development and validation of instruments designed to measure psycho-educational constructs in cross-cultural contexts.



SCHOLAR PAPERS

Psychological testing in the Ukraine: Peculiarities, Goals and Perspectives Khrystyna Rakhubovska Ukraine

In order to understand clearly the situation on psychological testing in Ukraine it is important to take into account the history of Ukrainian psychodiagnostics. Before the Russian Revolution of 1917 in Russia (Ukraine was a part of Russia at that time) all existing worldwide tests were well-known and original tests were being developed. For example, there was a famous test called "Psychological Profiles" created by G. Rossolimo (Rossolimo, 1909). During the Soviet period of Ukrainian history, especially in the 1920-30s testing in educational, organizational and counselling context was very common. Foreign tests were translated, new instruments were developed and scientific research on testing was conducted. Nevertheless, in 1936 testing was criticized and considered by the government as a practice based upon unscientific and "bourgeois" ideas. Subsequently, a lot of scientific institutions and laboratories were liquidated. For a long period of time testing was abandoned in the USSR.

In the late 1960s – the early 1970s the government officially acknowledged the need of testing development in the USSR. However, academical psychology kept treating testing neutrally or even negatively. Despite that, testing was used in psychological assessment practice. At that moment of time testing was used above all in clinical setting. Although, a long-term rejection of psychological testing had a very bad influence on further development of Soviet psychodiagnostics. Thus, in the 1980s only

a small part of Soviet psychologists were familiar with concepts of test validity and reliability and procedures for their evaluation. A lot of invalid and unreliable tests were developed, various foreign tests were translated and published without any copyright. In the late 1980s the situation changed slightly for better – even though so called quasi-tests were being published, certain monographs and manuals about the history of psychodiagnostics, psychometric properties and development of tests were issued. After disintegration of the USSR in the early 1990s and over the following 21 years the situation on testing in Ukraine has not changed fundamentally.

Ukrainian psychodiagnostics remains almost the same. Test users are offered the famous foreign tests, sold without any rights of their replication and sale. This does not only infringe upon interests of the authors, but also disables an author's supervision of a test adaptation. This leads to mistakes and confusion. In addition, most of the tests sold in Ukraine are outdated and education in the field of psychological testing is taught by professors, who do not have sufficient qualification for this activity.

The use of so called quasi-tests does not only undermine the authority of psychology as a science but may also damage mentally a person taking a test and influence negatively his or her further life. Finally, a professional nihilism towards tests is formed. It is a consequence of the absence of control of testing practice. Besides, the current situation complicates the integration process of Ukrainian psychology into the international scientific community.

To change the current situation it is important to create a testing infrastructure which requires the implementation of certification of tests and test users as well as the emergence of test publishers which take full responsibility for their products and act in accordance with the copyright law. Definitely, this infrastructure must contain a specialized post-graduate education system in the field of psychological testing which is necessary for professional test usage.

Currently there are no specific resolutions and authorizations concerning test standardization and respect for test authors' rights as well as certification of test users. Nonetheless, a big contribution to development of psychological testing in Ukraine was made by Prof. Dr. Leonid Burlachuk, the author of numerous publications in the field of psychodiagnostics and the first handbook on psychological testing. He works in educational activity regarding psychological testing and has also initiated, in 2008 the foundation of the first Ukrainian test publisher, OS Ukraine, in cooperation with the Italian test publisher Giunti O.S. The first psychodiagnostic instruments which meet international requirements are

now emerging in the Ukraine. All details can be found on the website of OS Ukraine: www.osukraine.com.

Acknowledgement

I wish to express my gratitude to Prof. Dr. Leonid Burlachuk for his assistance and provision of the materials essential for preparation of this article.

References

- Burlachuk, L. (1993). О дилетантстве в психологической диагностике. *Вопросы психологии*, 5, 116.
- Burlachuk, L. (2006). *Психодиагностика: Учебник для вузов*. СПб.: Питер.
- Burlachuk, L. (2011). Психологические тесты в Украине. Причины затянувшегося кризиса. *Матеріали Міжнародної науково-практичної конференції «Актуальні проблеми юридичної та екстремальної психології»*. Макіївка.
- Burlachuk, L. (2007). *Словарь-справочник по психодиагностике*. СПб.: Питер Пресс.

Testing in Belize **Betty-Jean Usher-Tate** **Belize**

Educational System and External Examinations

The beginning of formal education in Belize (formerly British Honduras) dates back to 1816 with an effort to educate "the young and disadvantaged children" in the Honduras Free School which offered religious and academic instruction (Bennett, 2008). Under British colonial rule, the Church-State system of education was established. In twenty-first century Belize, this symbiotic partnership remains in existence mostly because Belizeans perceive religion and moral teaching as an integral part of one's education. The system has also worked conveniently for the Belize government thus far. Secondary education began with Wesley High School in 1882 operated by the Methodist Mission (Bennett, 2008). By 1900 there were five high schools in the country; all managed by the religious denominations with little or no support from the government. Unlike primary schools which were tuition-free, secondary education required a financial investment for tuition, fees, textbooks, school supplies, and uniforms. According to Bennett (2008), the only contribution made to these schools in the early 1900s was an incentive bonus to the teacher or school for each student who passed the Cambridge Local Examinations.

Partly as validation of the quality of education received in Belize, students were encouraged to sit external examinations. Nevertheless, success in these examinations

translated into scholarship opportunities, university admissions, and employment benefits (increased salaries). External examinations that were advocated included but were not limited to the Cambridge Local Examinations, General Certificate Examinations, Royal Society of Arts Examinations, and the City and Guilds.

Geographically, Belize is both Caribbean and Central American. Historically, Belize identifies more with the English speaking Caribbean. In 1972 the Caribbean Examinations Council (CXC) was established with Belize as one of its 16 participating territories (CXC, 2012). British made exams were gradually replaced with subject specific tests prepared by the CXC: Caribbean Secondary Examinations (CSEC) for high school level and Caribbean Advanced Proficiency Exams (CAPE) for the tertiary / junior-college level. There are fees associated with each CXC test. In Belize, even though 25% of the nation's budget is spent on Education (Faber, 2011), students bear the full cost for the number of CXCs they take.

National Examinations

There are three major national examinations: one for secondary and two for primary school. The primary school exams are government sponsored (no cost to students).

Twelfth grade. The Association of Tertiary Level Institutions in Belize (ATLIB) prepares an exam which is intended for high school seniors applying for admission to tertiary level institutions (sixth forms / junior-colleges) within Belize. It is administered simultaneously across the nation once per year. Like the CXCs, students pay to sit this exam.

Sixth grade. The Belize Junior Achievement Test (BJAT) intended for all students in sixth grade. It is not offered simultaneously. BJAT is designed to inform and bears minimal negative consequences; it is low stakes.

Eighth grade. The Primary School Examinations (PSE) is also a paper-based exam. Items are based on the contents outlined in the national primary school curriculum. Four constructs are covered: English Language Competence, Math Competence, Science Competence, and Social Studies Competence. Each of the four content competencies is worth 100 points. The English Language Competence score is derived from three subtests (letter writing 20, composition 30, and comprehensive multiple choice items 50). The Math Competence score is derived from two subtests measuring math ability (multiple choice format 50 and student-generated responses 50). The Science and Social Studies are stand-alone tests with multiple choice formats.

PSE is administered in two days (about 4 weeks apart). High school facilities are used as testing centers and high school teachers invigilate

the examinations under the auspices of the Ministry of Education. To date, there have been very few instances of test-accommodations. The PSE is a high stakes test because one's composite score (averaged competencies) has direct impact on high school selection and admission. Some scholarship awards are also dependent on PSE scores. To date, published analyses of PSE results have been limited to descriptive statistics. Because the PSE carries such high stakes, validity of scores is very important. Invariance testing using structural equation modeling (SEM) techniques allows researchers to simultaneously estimate relationships among multiple variables for different groups such as gender and ethnicity (Motl et al., 2002). Structural equivalence, or invariance, concerns whether the theoretical structure of the instrument is the same for members of different groups (Byrne & Watkins, 2003).

In an unpublished SEM term paper, Usher-Tate and Anderson (2012) used scores from the English Language and Math Competence subtests of the 2010 PSE to conduct measurement invariance tests. There were 6,525 students represented in the dataset. For gender, the model tested operated the same for boys and girls. However, significant differences/variances were observed across language and school location categories, indicating a need for further investigation. On the other hand, Catholic managed institutions accounted for approximately half of the sample, but invariance testing showed insignificant difference. This indicates that whether the church-state partnership was Catholic or Non-Catholic, it made little difference in student achievement on the 2010 PSE English and Math. Future research with analyses using item level datasets will be useful.

Reference

- Bennett, J.A. (2008). *Education in Belize: A historical perspective*. Belize City: The Angelus Press Ltd.
- Byrne, B. M. & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, 34(2), 155-175.
- Faber, P. (2011). Education week address 2011 by hon. patrick faber part1. Retrieved from <http://www.youtube.com/watch?v=FCBQvdGv2-M>
- CXC. (2012). About the council. [the official website]. Retrieved from <http://www.cxc.org/about-us/about-council>
- Motl, R. W., Dishman, R. K., Ward, D. S., Saunders, R. P., Dowda, M., Felton, G., & Pate, R. R. (2002). Examining social-cognitive determinants of intention and physical activity among black and white adolescent girls using structural equation modeling. *Health Psychology*, 21(5), 459-467.
- Usher-Tate, B. & Anderson, N. (2012). Invariance testing on the belizean primary school exam

2010. [unpublished term paper].University of Nebraska, Lincoln.

Application and Development of Testing in China

**Jieting Zhang, Minqiang Zhang, Wenyi Zhang,
South China Normal University
Guangzhou
China**

On the Chinese Mainland, candidates of any test are numerous because of a huge population. The number of College Entrance Examination takers has reached ten million, and the number of candidates participating in Graduate Admission Test has increased to 1.2~1.5 million. Moreover, there are a large number of candidates in other examinations, such as Judicial Examination, Qualified Doctor Practitioner Examination, CPA examination, Civil Service Examination, and even in tests for foreign candidates, such as HSK. Since the recovery of the College Entrance Examination in 1970s, promotion of testing has taken place in Chinese Mainland, through the efforts of Houcan Zhang and other pioneering psychologists. Nowadays, a psychometric committee has been established under Chinese Psychological Society, and a branch for educational measurement and statistics has also been set up under Chinese Educational Society, with approximately 1,000 members. These professionals active in all kinds of fields are prompting the theory and application studies in psychometrics, leading to great achievement in theory and application of Classical Testing Theory (CTT), Item Response Theory (IRT) and Generalizability Theory (GT). Plenty of scales with conformance to psychometric rules have been widely used for different kinds of population, which play an important role in mental health improvement and mental disease prevention for Chinese people.

In recent years, we have investigated the application and theoretical development of psychological and educational testing, by reviewing the related articles published in five distinguished Chinese journals on psychology and education, from 2000 to 2009. The application and development of testing in China are discussed in accordance with these results.

Application of Psychological and Educational Testing

A multitude of important psychological scales have been revised in China mainland. Besides localization of foreign scales, many scales with considerably high reliability and validity have been newly compiled in accordance with the practical condition and demand of domestic application and research, such as Irrational Belief

Scale of Middle-School Student and Computer Game Addiction Inventory for Chinese College Students. By the end of the last century, a handbook of scales for mental hygiene has been published with approximately 140 scales of common use. Furthermore, about 60 psychological inventories have been translated or compiled in sport science alone.

Testing is mainly used in four aspects in China mainland. Firstly, routine test of ability or achievement is applied to evaluate the performance of students and teachers as well as the effect of teaching and learning. Secondly, matriculation exams and vocational qualification exams not only promote efficiency of talent recruiting, but ensure equity in recruitment and prompt rational allocation in human resource. Thirdly, scales regarding mental health and personality are widely used in psychological counseling in order to take intervening measure or therapy beforehand. Last but not least, academic achievement tests and psychological scales are also scientific research tool for theoretical or empirical studies in psychology and education.

According to the review study, scales are used quite extensively in psychology studies, accounting for about 15% of that total. However, it is not that common in educational studies, which only accounts for up to 4%. Among the application, about 40% used the scales developed during their temporal studies, which reflects the great needs of developing new scales in various psychology studies. However, reliability and validity for these self-developed scales requires further investigation and deserve more attention on the conclusion, since they are lack of long-term verification. The further investigation on scale property is also beneficial for the generalization of these scales. Moreover, the inappropriate "reliability induction" is quite common when applying the scales developed by previous researchers (Jiao, Huang, Zhang, Wu, & Wang, 2010; Jiao, Zhang, Zhang, Wu, L., & Zhang, 2011; Jiao, Zhang, Wu, & Zhang, 2010). Research Organizations in China have to strengthen the effort in proposing researchers to report the reliability coefficient of the samples at hand.

Development of Testing Theory and Method

Due to the expanding scale of application in multiple tests and the rising demand of test organizing, studies on theory and method of testing has been strongly facilitated in China mainland. Measuring application under CTT framework has been popular, and some researchers have proposed certain improving methods specific to the defects of CTT. For instance, it is advised that item analysis and reliability analysis should be conducted in sub-group of test sample, rather than in the whole sample, so as to avoid the over-rigid

premises when calculating difficulty or reliability coefficients.

Widespread application and development of computer technique and statistical measurement software have offered necessary technical support for the inland research on measurement and statistics theory. Modern measurement theory, with moderately rapid development in mainland, is gradually applied in all kinds of tests. Over 500 academic theses on IRT have been published since 2001. Among those, Computerized Adaptive Test (CAT) is a hot spot in IRT application, many research projects have been set up on the system of CAT, such as College English Test, which is in the experimental stage. Additionally, item banks of a certain important tests are constructed on the basis of IRT, adopting polytomous scoring and Samejima graded response model; some measuring experts make use of IRT in the test item analysis (Chen, Ding, 2011). Combination among different measuring theories or techniques is no longer rare nowadays, such as integration of CAT and cognitive diagnostic model (Tang, Ding, & Yu, 2012). Recently, professors in South China Normal University are working with Computer adaptive test system for elementary and secondary school students; moreover, cognitive diagnostic model are used to provide further information on the structure of students' ability.

There is increasing research and application of generalizability theory (GT) since it was adopted in mainland by Houcan Zhang in 1988. Searching in CNKI database alone, we could find more than 130 theoretical or applied research achievements in the last ten years. For one thing, GT is mainly used in assessing error of subjective evaluation, such as structured interview, performance evaluation and vocational-technical examination, and improving design of evaluation (Li, & Zhang, 2009; Xu, Zhang, & Li, 2009). For another, GT is used in analyzing reliability and validity of a certain large-scale test (e.g. College Entrance Examination, language test) and psychological scales (Zheng, Gu, & Zhu, 2009). Moreover, some theoretical researches in this realm are acting on international convention by degree; GT combined with IRT are been studied.

In recent years, many new statistics methods have been recommended inland and applied in measurement, which make the analysis and interpretation of measurement more objective, specific and effective. For example, latent class analysis has been applied in categorical data processing, the result of which is served as the basis of further statistical analysis. For instance, we can define target groups referring to the classification, and analyze discrimination and difficulty in accordance to the difference among the sub-groups (Jiao, Zhang, Guan, & Zhang, 2010). However, compared with other fields of psychology and education, studies on testing have not received

enough attention. The review of journals indicates that studies of testing take only 5% of psychology and educational study, and they mainly focus on scale development and IRT. Furthermore, 76% scale studies are conducted based on CTT, and others are based on IRT or GT. Consequently, we should invest more resources in researches of testing, as the basic science of psychology and education.

In conclusion, due to the increasing application and demand of testing as well as the growing academic communication, the testing methods in China are developing rapidly and applied extensively. Nonetheless, we have to make more efforts in recommending new methods in application, so as to stress the importance and influence of testing in psychological and educational studies.

References

- Chen, X., Ding, S. (2011). The Variable-weighted Item Selection Strategy for Computerized Adaptive Testing with the Generalized Partial Credit Model. *Psychological Science*. 2011, 34(4), 965-969.
- Jiao, C., Huang, Z., Zhang, M., Wu, L., & Wang, X. (2010). Meta-analysis of Statistical Method Application in Psychological Studies—Taking Acta Psychologica Sinica and Psychological Science (1998-2008) as Examples. *Psychological Science*, 33(1), 48-54.
- Jiao, C., Zhang, J., Guan, D., & Zhang, M. (2010). Analysis of psychology professional test section in postgraduate entrance examination based on latent class model. *China examinations*, (4), 3-12.
- Jiao, C., Zhang, J., Wu, L., & Zhang, M. (2010). Scores Reliability from the Minnesota Multiphasic Personality Inventory in China, *Journal of South China Normal University (Social Science Edition)* (4), 48-52.
- Jiao, C., Zhang, M., Zhang, J., Wu, L., & Zhang, W. (2011). A Cross-cultural Comparison of the Reliability Generalization Studies on the Eysenck Personality Questionnaire and Its Implications. *Psychological Science*, (6), 1488-1495.
- Li, G., & Zhang, M. (2009). Using Generalizability Theory to Analyze the Influencing Factors in Competency Evaluation with the Checklist Scale. *Psychological Science*. 32(2), 420- 422.
- Tang, X., Ding, S., & Yu, Z. (2012). Application of Computerized Adaptive Testing in Cognitive Diagnosis. *Advances in Psychological Science*, 20(4), 616-626.
- Xu, S., Zhang, M., & Li, G. (2009). Based on Generalizability Theory and Many-facet Rasch Measurement To Analyze the Data of Structured Interview. *Psychological Exploration*, 29(5), 77-82.
- Zheng, X., Gu, H., & Zhu, P. (2009). A Study of the Emotional Intelligence Scale (EIS) by the Multi-

variate Generalizability Theory. *Psychological Science*. 32(1), 181- 183.

**Test Translation, Adaptation and Development:
The Situation in Pakistan
Sobia Aftab
University of Karachi
Pakistan**

In Pakistan, in the last few years the trends have been changed rapidly with regard to test translation, adaptation, and development. Initially there was lack of empirical researches which may stem from the fact that there was shortage of questionnaires /instruments/tests in National language (i.e. Urdu). Researchers were using questionnaires that were in English language and were developed in western culture/countries. With the passage of time, researchers have started realizing that tests standardized in one culture cannot be applied or predict anything with same accuracy to another culture. Epidemiological studies have emphasized upon the significance of using research tools in studies in one's own native language. Others have also stressed the implications of translations of instruments into one's own native language, For instance, Johanson and Malak (2007), have suggested that it is very essential that every culture has its own translated or adapted or developed scales in local languages to conduct studies, so that internationally comparable brilliant quality studies could be generated. Likewise, the importance of translation and adaptation studies is also stressed by Milton (2009).

Consequently, the researchers in Pakistan have turned towards the translation, adaptation, and validation processes of the western instruments to make their use possible within our culture. Several research tools were either translated or adapted to assess various areas including, psychological, social, behavioral, emotional, etc. Examples of few of translated and adapted research tools include: Culture Free Self-Esteem Inventories-3 (Imran, 2012), State Self-Esteem Scale (Aslam, 2012), Rosenberg Self-Esteem Scale (Sardar, 2000), Paranoia Scale (Kamrani & Ali, 2010) Perceived Social Self-Efficacy Scale (Riaz, Yasien, & Khanam, 2011), Social Provision Scale (Rizwan & Syed, 2010), Interaction Anxious Scale (Riaz & Bano, 2011), etc.

Even so, epidemiological studies have focused on the cultural relevance and clinical usefulness of translation of western scales into non-western languages. This translation is to be an interim until the equivalents of such instruments are developed indigenously (as cited in Khan, 1998). Meaning and understanding of different psychological constructs vary from one culture to another i.e., a construct perceived in one culture

may not be perceived in the same manner in another culture. It is a fact that accurate measurement of any construct depends on the excellence and cultural relevance of the measurement tools. Nevertheless, often in various situations, this is not possible to develop native scale. The reason may be that the test development requires investment of time and money and expertise also. Hence, because of lack of assets and somehow talent also the researchers preferred to translate and adapt western research instruments. Hence, there was little innovative work done in the field of scale construction.

Another reason to prefer translation and adaptation of tools was so far, there was no regulatory body and also there was no defined copyright policy within Pakistan. So, copyright violation was very common and some researchers were translating and adapting research instruments without taking permissions from concerned authors and publishers. The reason was long procedural requirements to take permission to use scales from foreign publishers and the lack of resources to purchase instruments and questionnaires from foreign publishers; the cost of which was not bearable for everyone.

However, in 2007 Higher Education Commission (HEC) tried to create awareness among people by introducing the copyright and plagiarism policy following the international guidelines. Initially, the researchers didn't accept this reform but gradually the acceptance among research community is developed. Now all these policies are the part of the curriculum in Higher Education Studies in Pakistan. More recently another positive step was taken to upgrade fair testing and research situation in Pakistan and to make it equivalent to international standards i.e. the President of Pakistan signed a Copyright Ordinance 2012 which specifies punishments for different types of copyright violation. These reforms have created positive changes and now construction/development of research tools is also receiving increased research attention. In order to avoid the long procedures to take permission to translate or adapt scales and unbearable cost to purchase these, researchers in Pakistan prefer to develop research tools within their own cultural context following American Psychological Association (APA) and International Test Commission (ITC) guidelines. Another positive step taken by Pakistan Psychological Association (PPA) to strengthen the research and testing situation in Pakistan is the formation of sub-divisions, including, Ethics and Research Development. Hence, these steps will not help in reducing the rate of plagiarism and copyright violation but also will strengthen the status of testing and research in Pakistan.

References

- Aslam, S. (2012). *Urdu Adaptation and Validation of State Self Esteem Scale in Young Adults*. Unpublished M. Phil Thesis. Institute of Clinical Psychology, University of Karachi, Pakistan.
- Imran, H. (2012). *Self-Esteem as a function of gender and socio-economic status and its relationship with academic achievement in children: Use of CFSEI-3 with Urdu Adaptation*. Unpublished PhD Dissertation. Institute of Clinical Psychology, University of Karachi, Pakistan.
- Johanson, I., & Malak, B. (2007). *Translation and National Adaptations of TIMSS 2007 Assessment and Questionnaires*. Timss and Pirls International Study Center, Lynch School of Education, Boston College.
- Kamrani, F., & Ali, U. (2011). Urdu Translation and Adaptation of Fenigstein Paranoia Scale. *International Journal of Business and Social Science*, 2(16), 228-237
- Khan, S. (1998). *Moderator Variables in Vulnerability to Anxiety: Social Support and Resilience*. Unpublished PhD Dissertation. Institute of Clinical Psychology, University of Karachi, Pakistan.
- Milton, J. (2009). Translation studies and Adaptation Studies. Intercultural Studies Groups In Pym, A. & Perekrestenko, A. (eds) *Translation Research Projects 2*. Tarragona, 51–58.
- Riaz, Z., & Bano, N. (2011). Translation and Adaptation of IAS and its reliability assessment. *The International Journal of Educational and Psychological Assessment*, 9 (1).
- Riaz, Z., Yasien, S., & Khanam, S. J. (2011). Translation and Adaptation of Perceived Social Self Efficacy Scale (PSSE). *The International Journal of Educational and Psychological Assessment*, 2 (42), 79-92.
- Rizwan, M., & Syed, N. (2010). Urdu Translation and Psychometric Properties of Social Provision Scale. *The International Journal of Educational and Psychological Assessment*, 4, 33-47.
- Sardar, S. (1998). *Study of relationships among childhood paternal loss, sex-role orientation, self-esteem and locus of control in male and female students*. Unpublished PhD Dissertation. Institute of Clinical Psychology, University of Karachi, Pakistan

Gender Differences in Mathematics and Science Performance in Malawi

Ella Banda
Malawi

The difference in levels of achievement between boys and girls in mathematics and science is an issue that has generated a great deal of

concern and debate over the past few years. The Science and Technology policy for Malawi (2001) acknowledges that although 52% of the Malawi population is female, the participation in and utilization of Science and Technology by women has not received the attention it deserves. Efforts have been made to encourage girls' enrollment and participation in the sciences both by the government and the private sector. In 1986 a National Commission on Women in development was set up with the aim of coming up with strategies that would help change the attitudes of teachers, parents and the community at large about girls education and motivate girls to study Mathematics and Science subjects (Kadzamira, 1986). Female scientists and educationists are also actively involved in highlighting the gender disparities in the Science and Technology disciplines with the view to influence policy direction in favor of women and girls.

This paper discusses the impact of such interventions in closing the gap in the performance of girls in mathematics and science. This is done by identifying the extent and prevalence of differences in performance between girls and boys in relation to subject areas and type of school.

Aims - The purpose of the study was to find out if there are any significant differences in performance among girls and boys from the different school categories in science and mathematics at MSCE (i.e. senior secondary school) level.

Participants - Scores of all internal candidates in 2004 MSCE were used. The scores were for Agriculture, Biology, Mathematics and Physical Science. Performances of girls from three categories of secondary schools were compared. These were Community day (CDSS), private (PS) and conventional secondary school (CSS).

Procedure - The scores of both boys and girls from the three types of schools were extracted and compared using t-tests. Cohen D statistic was used to determine the effect size. Cohen D value of 0.2 is said to be small, of 0.5 is said to be moderate and of 0.8 is said to be large (Howell, 2002). The null hypothesis was that there would be no difference between the mean scores of the candidates from these three types of schools.

Girls from the Conventional schools are doing better in all the science subjects while those from CDSS are doing worse in all subjects. A major discrepancy is observed in Physical Science. The pattern of the performance of boys across the school types is similar to that of girls in that boys from CDSS perform worse than those from the other schools while boys from conventional schools are performing better than the rest. The differences are however larger for girls than for boys.

Differences in performance between boys and girls within a school are generally moderate except in Physical science. The achievement gaps are relatively higher between boys and girls in the CDSS's in all the subjects. One explanation of this observation could be the effect of the girls' attitude, motivation and self perception in relation to these subjects. Studies have shown that girls do not see the usefulness of mathematics and science for their future life and are therefore not motivated to work hard in these subjects. This may be true for the girls in CDSS's because these schools are characterized by mostly under qualified teachers, inadequate resources, and larger proportions of students from low socio-economic status. The conventional schools however are relatively better equipped and have more qualified teachers than the CDSS's. The private schools are mixed in terms of resources and teacher qualifications. There is strong relationship between the characteristics of a particular school and achievement. This could also explain the relatively low performance of all students in the CDSS's considering their poor characteristics.

Most of the students who go to the conventional schools are those that did well at the PSLCE i.e *the cream*, while the rest are selected for CDSS's and those that can afford to pay higher school fees go to the private schools. Ideally the students from the CDSS's are the ones that would require more resources and increased efforts by the government and the NGO's to motivate and help them to catch up with their counter parts in the conventional schools.

References

- Howell, D.C. (2002) *Statistical Methods for Psychology* (5th ed.). California: Wadsworth.
- Kadzamira, E.T. (1986) *Closing the Gender Gap*. Unpublished: Malawi National Examinations Board.
- Ministry of Education Science and Technology, (2001) *The Science and Technology Policy for Malawi*. Lilongwe: Malawi Government Print.

The Structural Validity of the Inductive Reasoning Developmental Test for the Measurement of Developmental Stages

Hudson F. Golino Cristiano
& Mauro Assis Gomes
Laboratory for Cognitive Architecture Mapping
(LaiCo)
Universidade Federal de Minas Gerais
Brazil

The Inductive Reasoning Developmental Test – IRTD (Gomes & Golino, 2009) is a pencil-and-paper instrument designed to assess developmentally sequenced and hierarchically organized inductive reasoning. The sequence was constructed based on the Model of Hierarchical Complexity (Commons, Trudeau, Stein, Richards, & Krause, 1998; Commons & Pekker, 2008). The MHC is a general measurement theory, and as such is part of the normal Mathematical Theory of Measurement (Krantz, Luce, Suppes, & Tversky, 1971) applied to the phenomenon of difficulty. Commons and Pekker (2008) demonstrated, in axiomatic terms, that task difficulty or complexity, beyond other sources, increases in two ways: horizontally and vertically. The former refers to the accumulation of informational bits necessary to successfully complete a task (Commons, 2008), e.g. $5 + 6 + 7$ is less complex than $5 + 6 + 7 + 8$, because the first differs from the second in the number of times addition was executed, and does not differ in the organization of the addition itself; that is, both have the same *hierarchical (or vertical) complexity*. Vertical complexity, or *hierarchical complexity*, refers to the organization of information in the form of action in two or more subtasks, in a coordinated way. A good example of vertical complexity is the distributive property, that requires the two actions of addition and multiplication to be performed in a certain order, thus, coordinated.

According to the MHC, constructing calibrated tests for developmental stage identification requires a specific design that involves: 1) grouping items with same hierarchical complexity [$h(i_1) = h(i_2) = h(i_3) = \dots h(i_n)$] within stages; and 2) using items with increasing hierarchical complexity [$h(\text{Stage } 1) < h(\text{Stage } 2) < h(\text{Stage } 3) < \dots h(\text{Stage } k)$] between stages. The first deals with item or task equivalence, important in order to avoid the elaboration of an anomalous scale that confuses its analysis (Fischer & Rose, 1999). The second makes possible the identification of discontinuous, stage-like development, with gaps between different orders.

The IRDT has eight items specifically designed to identify each stage, covering seven stages (from Pre-operational or Single Representations to Metasystematic or Single Principles) in a total of 56 items. Each item is composed of four letters, or se-

quence of letters, with a specific rule (correct items), plus one letter or sequence with a different rule (exception). The task is to discover which letter or sequence is the exception. The IRTD was administered to a convenience sample composed by 1,193 Brazilian people (53% women, 47% men) aged between 6 to 86 years ($M = 15.02$, $SD = 12.30$). The sample was intentionally broad, and had a distribution of 34.20% from 6 to 11 years, 53.50% from 12 to 18 years, 6.40% from 19 to 49 years, and 5.90% from 50 to 86 years. All the participants were from the city of Belo Horizonte, state of Minas Gerais, Brazil.

In the first part of the data analysis the dichotomous Rasch Model is used, using the software Winsteps 3.70.1 (Linacre, 2012). It produces linear measures, gives estimates of precision, allows the detection of lack of fit or misfit, enables the parameters' separation of the object being measured and of the measurement instrument (Panayides, Robinson & Tymms, 2010) and enables the verification of hierarchical sequences of both item and person, being especially relevant to developmental stage identification (Dawson, Xie & Wilson, 2003). After verifying the fit to the dichotomous Rasch Model, the structure of the items' difficulties is visually verified through the Wright (variable) map. It is expected that items constructed to identify a particular stage form a well-defined cluster. Each cluster of items should be separated by a gap. The clusters visually verified can be checked analytically through hidden markov model, using the `depmixS4` package of the R software.

The result of the Rasch analysis shows a reliability of .99 for the 56 items and an infit meansquare mean of .90 ($SD = .23$), having 75% of the items ranging its infit from .70 to 1.20, only 3 items with an infit less than .55 and no item with infit greater than 1.35. The person reliability was .80. The principal contrast showed that the raw variance explained by measures (modeled) is of 73.2%, and that the unexplained variance in the first contrast (modeled) is 6.9%, suggesting that the instrument can be thought of as unidimensional. The items' difficulties presented 7 clusters with gaps between them, as visually verified by the Wright map. The clusters were confirmed by the hidden markov model, which presented a solution with 7 classes (log Lik.= -26.28, df=62, AIC: 176.57, BIC: 302.14).

The IRTD presented a good fit to the dichotomous Rasch Model, and a high reliability for the items, and a moderately high reliability for the people measured. The hypothesized stages were all identified in the structure of the items' difficulties, and confirmed through latent class analysis. So, it is possible to empirically verify developmental stages using the MHC as a framework for constructing instruments, applying contemporary psychometric models that properly deals with the issues involved in the measurement of cognitive abilities, and employing a robust method to verify stages in the structure of the items.

References

- Commons, M. L. & Pekker, A. (2008). Presenting the formal theory of hierarchical complexity. *World Futures*, 64, 375-382.
- Commons, M., Trudeau, E., Stein, S., Richards, F., & Krause, S. R. (1998). Hierarchical complexity of tasks shows the existence of developmental stages. *Developmental Review*, 18(3), 237-278.
- Dawson, T. L., Xie, Y., & Wilson, M. (2003). Domain-general and domain-specific developmental assessments: Do they measure the same thing? *Cognitive Development*, 18, 61-78.
- Gomes, C.M.A., & Golino, H.F. (2009). Estudo exploratório sobre o Teste de Desenvolvimento do Raciocínio Indutivo (TDRi). In D. Colinvaux. *Anais do VII Congresso Brasileiro de Psicologia do Desenvolvimento: Desenvolvimento e Direitos Humanos*. (pp. 77-79). Rio de Janeiro: UERJ. Available in <http://www.abpd.psc.br/files/congressosAnteriores/AnaisVII CBPD.pdf>
- Linacre, J. M. (2012). Winsteps® Rasch measurement computer program. Beaverton, Oregon: Winsteps.
- Panayides, P., Robinson, C., & Tymms, P. (2010). The assessment revolution that has passed England by: Rasch measurement, *British Educational Research Journal*, 36(4), 611-626.

Current Developments of Clinical Assessment in Bulgaria Zornitsa Kalibatseva Michigan State University United States

Clinical psychology is a relatively new discipline in Bulgaria that emerged in the last five decades. The profession "clinical psychologist" became established in the early 1990s. Currently, degrees in clinical psychology are offered only at the Master's level. Yet, the field of clinical psychology may be disregarded more than other psychology disciplines because it is not listed as one of the studied areas by the Department of Psychology in the Institute for Population and Human Studies in Bulgaria. Clinical assessment and psychotherapy are central components of clinical psychology. Clinical assessment (also called "psychodiagnostics" or "psychological evaluation") is an emerging field in Bulgaria and refers to the process of testing with different techniques to understand a person's behavior, personality, and abilities. It is most often conducted with the use of a clinical interview, observation, and standardized psychological tests.

The aim of this paper is to discuss the current use of psychological tests for the assessment of psychopathology in Bulgaria. Bulgaria, a country in Eastern Europe, joined the European Union (EU) in 2007. As a former communist country and a geographical crossroad of the East and the West throughout the centuries, Bulgaria has increasingly received western

influences and is an emerging democracy since 1989. Mental health in Bulgaria has been influenced by a wide range of historical, sociocultural, and economic factors (Tomov, Mladenova, Lazarova, Sotirov, & Okoliyski, 2004). The slow economy and higher rates of poverty, crime, and homelessness (as compared to 1989) have been associated with mental health problems and an increased need for psychological services. Yet, the previous policy neglected mental health for decades and only recently the country's Mental Health Programme was devised (Tomov et al., 2004). Still, there was little epidemiological data on which to base the programme's objectives (Tomov et al.). Specifically, one of the identified issues is the lack of process and outcome indicators to monitor mental health service effectiveness. One possibility to address this issue is to incorporate validated measures of psychopathology that would assist health professionals in tracking progress.

The first national representative psychiatric epidemiological survey (EPIBUL) in Bulgaria was conducted between 2002 and 2006. EPIBUL is part of the World Mental Health Initiative of the World Health Organization. The instrument for this survey was the structured clinical interview CIDI (Composite International Diagnostic Interview) WMH 3.0 (Tomov et al., 2008). The CIDI included questions that inform the diagnoses of psychological disorders from the International Classification of Diseases (ICD-10) and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). The preparation and adaptation of the Bulgarian version of the instrument followed the guidelines of the WHO. First, the original English version was translated by a professional Bulgarian translator, who consulted with psychiatry experts. Second, selected parts of the instrument were back-translated to English. Third, discrepancies were identified and changes were made to accommodate for specific expressions in Bulgarian. In addition, jargon related to psychiatric disorders in Bulgarian was removed. The final wording was approved by several expert Bulgarian psychiatrists. A different panel of psychiatrists evaluated the adequacy of the language and found information about particular medications, religions, professions, and income that are specific to Bulgaria. Fourth, the instrument was piloted with 50 respondents and approved for further use (Tomov et al., 2008). The described adaptation ensured linguistic and possibly conceptual equivalence of the instrument. However, no information has been provided about the metric and functional equivalence, which are important components of establishing cross-cultural measurement invariance (Leong, Leung, & Cheung, 2010).

Additionally, it is important to note that the contemporary comprehension of psychopathology as described in ICD-10 and DSM-IV reflects the

views of health and disease that are predominantly held within Western culture. Thus, it is critical to recognize that current views of psychiatric nosology may be biased towards presenting the etiology, expression, diagnosis, and treatment of mental disorders in ways that are mostly congruent with Western culture.

The overall trends regarding the standards for creating and adapting existing psychological tests are very promising. During the 3rd National Congress of Psychology in Bulgaria in 2005, a round table discussion generated several objectives for the future use and adaptation of psychological tests. Notably, the standards of the International Test Commission (2010) and the American Psychological Association for test adaptation were discussed. Subsequently, the National Test Commission (NTC) under the auspices of the Bulgarian Psychological Society was created. Its mission and role include preparing and providing information about the quality of psychological tests in Bulgaria, the qualifications for administering such tests, the standards for creating and adapting tests, the competencies necessary for professionals to use and interpret the tests, and the standards for certification programs (NTC, 2012). Additionally, there is a registry of existing adapted psychological tests in Bulgaria. Some of the tests used for clinical assessment that have been adapted include the Minnesota Multiphasic Personality Inventory – 2 (MMPI-2) and the State Trait Anxiety Inventory (STAI-Y). Three original tests that measure aggression and anxiety among adolescents and elementary school students have been developed.

The field of clinical assessment has been progressing slowly and various researchers and practitioners have developed original tests or used the ITC guidelines to or translate and adapt existing tests for research and clinical practice. However, there is a dearth of instruments that have been used in the field and there is a need to develop more emic-derived clinical assessment tools. Specifically, original personality and mental health instruments may be important constructs to explore further. Also, researchers are strongly encouraged to describe in detail any adaptations of clinical instruments that they undertake.

References

- Institute for Population and Human Studies (2012). Department of Psychology. Retrieved from <http://www.iphs.eu/en/departmentspsycho.html> on June 12, 2012.
- International Test Commission (2010). International Test Commission Guidelines for Translating and Adapting Tests. Retrieved from <http://www.intestcom.org/upload/sitefiles/40.pdf> on June 13, 2012.
- Leong, F. T. L., Leung, K., & Cheung, F. M. (2010).

Integrating cross-cultural psychology research methods into ethnic minority psychology. *Cultural Diversity and Ethnic Minority Psychology*, 16(4), 590-597.

National Test Committee of the Bulgarian Psychological Society (2012). Mission and role.

Retrieved from <http://ntcbg.org> on June 13, 2012.

Tomov, T., Hinkov, H., Zarkov, Z., Mladenova, M., Vasilev, S., & Okoliyski, M. (2008).

National representative epidemiological study of common mental disorders in Bulgaria (2002-2006): Instruments, methodology, field procedures, and evaluation. *Social Medicine (Socialna Medicina)*, 4. Retrieved online from http://www.proecta.org/download/Statia_SOCIALNAM_4_2008.pdf on June 15, 2012.

Tomov, T., Mladenova, M., Lazarova, I., Sotirov, V., & Okoliyski, M. (2004). Bulgaria mental health country profile. *International Review of Psychiatry*, 16, 93-106.

The Philippine Educational Measurement and Evaluation Association (PEMEA)

Ronnel B. King

**The University of Hong Kong
Hong Kong**

The Philippine Educational Measurement and Evaluation Association (PEMEA) is the premier organization dedicated to testing in the Philippines. It was inaugurated in August 2008 during the first National Conference on Educational Measurement (NCEME). It is currently an affiliate member of the International Test Commission.

PEMEA aims to achieve the following goals:

- Promote standards in various areas of education through appropriate and proper assessment
- Provide technical assistance to educational institutions and process of attaining standards
- Enhance and maintain the proper practice of measurement and evaluation in both local and international level
- Enrich the theory, practice and research in evaluation and measurement in the Philippines

Publications

The *Educational Measurement and Evaluation Review (EMER)* (ISSN: 2094-5876) is the official publication of the association. The journal publishes "scholarly reports about contemporary theory and practice in the field of education and social science that highlights measurement, assessment, and evaluation. It welcomes articles that are about test and scale development, quantitative models of a construct, evaluation studies, best practices in evaluation, issues in assessment, contemporary approaches in educational and psychological

measurement, and other studies with direct implication to assessment in education, social science, and related fields." It is published annually and abstracted in the Asian Education Index, Social Science Research Network, Google Scholar, Open-J Gate, and New Jour. Thus far, two issues have been released covering a wide range of topics in measurement and evaluation. The editor-in-chief is Prof. Rose Marie Salazar-Clemaña from De La Salle University, Manila. The executive editor is Dr. Carlo Magno also from De La Salle University, Manila. Associate editors include Dr. Richard Gonzales and Ms. Belen Chu.

Another publication affiliated to the association is the *Assessment Handbook* (ISSN: 2094-1412). This journal is currently abstracted in the Asian Education Index and the Social Science Research Network. The journal publishes articles related to assessment, evaluation, measurement, psychometrics, and psychological testing that have been presented in the National Conference on Educational Measurement and Evaluation. Each issue is usually managed by a guest editor. The managing editors are Dr. Carlo Magno, Laramie Tolentino, Dr. Richard Gonzales, and Belen Chu.

Conferences

The first National Conference on Educational Measurement and Evaluation (NCEME) was held in August 2008 at the De La Salle College of St. Benilde (DLS-CSB). It was organized by the Center for Learning and Performance Assessment of DLS-CSB. The theme of the conference was "Developing a Culture of Assessment in Learning Organizations." It was a 2-day conference attended by over 300 people. The keynote speaker was Dr. Letticia Asuzano, President of the Asian Psychological Assessment Services Corp. A special session was headed by Mr. Neil Paríñas (Special Assistant to the President of DLS-CSB) which discussed the possibility of establishing a professional organization dedicated to educational measurement and evaluation. The initial draft of the constitution and by-laws was also presented and PEMEA was officially inaugurated. The second NCEME was held in 2010 was held on August 2010 in the College of Saint Benilde International Conference Center. The theme was "Educational Assessment in Knowledge Society." The keynote speaker was Dr. Anders Jonsen.

In August 2012, PEMEA will be hosting the International Conference on Educational Measurement and Evaluation. The theme is "Educational Assessment in a Multicultural Learning Environment." ITC Council member Dr. Thomas Oakland has been invited as the keynote speaker.

For more on PEMEA visit <http://www.pemea.org/>

NEWS

Symposium on *The assessment and teaching of 21st century skills* at the 2012 ITC conference in Amsterdam, the Netherlands

Since 2009 ATC21S has been working with teachers from Australia, Costa Rica, Finland, the Netherlands, Singapore, and the United States to pilot and test an extensive library of innovative resources (lessons, supplemental materials, formative assessments) aimed at assessing and teaching the specific 21st century skills of collaborative problem solving and digital literacy. ATC21S is a global private/public partnership founded and funded by Cisco, Intel and Microsoft. The project headquarters are at the University of Melbourne, directed by Professor Patrick Griffin, Associate Dean in the Melbourne Graduate School of Education. Upon completion in July 2012, the project will release its cutting-edge assessment tools and the accompanying teaching, technology, and policy-related materials.

The ATC21S symposium is scheduled for Thursday, July 5th, from 10.30-12.00. In addition, the project team is organizing a *launch event* for delegates during the conference. The ATC21S team will:

Demonstrate the cloud-based tasks commissioned for the project in the areas of collaborative problem solving (CPS) and ICT Digital Literacy. These tasks illustrate the state of the art in the assessment of collaborative problem solving and the assessment of learning through digital networks.

Show the reports that the ATC21S CPS and ICT Digital Literacy tasks can generate. These reports will be invaluable to schools, teachers, and individual students in assessing and tracking student proficiency in 21st century skills such as CPS and ICT Digital Literacy.

Explain how the ATC21S assessment tasks can be accessed and used once they are officially released in July 2012.

The original white papers commissioned for ATC21S can be downloaded from the ATC21S website and an edited, peer reviewed volume of the white papers is available from Springer Science + Business Media.

Further information

The project <http://atc21s.org/>

The project white papers <http://atc21s.org/index.php/resources/white-papers/>

The peer reviewed volume <http://www.springer.com/education+%26+language/book/978-94-007-2323-8>

LATEST ISSUE

International Journal of Testing
Volume 12, Issue 2, 2012

Special Issue: International
Approaches to Test Reviewing

Contacts

ITC President President@intestcom.org

ITC Secretary Secretary@intestcom.org

ITC Treasurer treasurer@intestcom.org

Editors

International Journal of Testing
ijteditor@intestcom.org

Testing International
tieditor@intestcom.org

ITC Website:
www.intestcom.org

Contribute to Testing International
Send your articles for consideration
to the Editor Jan Bogg