

TESTING INTERNATIONAL

Vol. 19, July, 2008 Editor: Jan Bogg



PRESIDENT

Prof. Jacques Grégoire
Université catholique de Louvain
Faculté de Psychologie
Place du Cardinal Mercier, 10
1348-Louvain-la-Neuve
Belgium

PRESIDENT-ELECT

Prof. M. Born
Erasmus University Rotterdam
Institute of Psychology
Woudestein, T13-15
P.O. Box 1738
3000 DR Rotterdam

SECRETARY

Prof. R.K. Hambleton
University of Massachusetts U.S.A

TREASURER

Prof. Em. Barbara M. Byrne
School of Psychology
University of Ottawa
145 Jean-Jacques Lussier
Ottawa, Ontario
Canada K1N 6N5

PAST-PRESIDENT

Prof. Jose Muniz
University of Oviedo, Spain

COUNCIL MEMBERS

Elected Members
Prof. Fanny Cheung
Chinese University of Hong
Kong, Hong Kong
Prof. Cheryl Foxcroft
University of Port Elizabeth, S.A
Prof. Frederick Leong
Michigan State University, USA
Co-Opted Members
Dr Iain Coyne
University of Nottingham, UK
Dr David Foster
Kryterion, USA
Prof. Tom Oakland
University of Florida, USA

REPRESENTATIVES

Dr. M. Bullock, IUPsyS
American Psychological
Association, U.S.A.
Prof. D. Bartram, IAAP
SHL Group plc
U.K

EDITORS

International Journal of Testing
Prof. John Hattie
University of Auckland
School of Education
Fisher Building
18 Waterloo Quadrant
Private Bag 92019
Auckland
New Zealand

Testing International

Dr. Jan Bogg
University of Liverpool
Department of Clinical Psychology
The Whelan Building
Liverpool L69 3GB
UK

CONTENTS

ITC PRESIDENT'S LETTER	2
MEET THE ITC 2008 SCHOLARS	3
SCHOLARS ARTICLES	
Testing in the South African Context Dalray de la Harpe, SOUTH AFRICA	3
The Use and Development of Personality Testing in China Zhou Mingjie & Zhang JianXin, CHINA	5
Testing in Albania – a new field full of needs and challenges Gladiola Musabelliu, ABANIA	7
Psychological testing in South Africa current issues and challenges Thamsanqa. j. Dhladhla & François de Kock, SOUTH AFRICA	8
ARTICLES	11
The Challenges of Fostering a Culture of Psychological Testing in a Small Post-Soviet Country Albinas Bagdonas, Lithuania	11
From normal curve to slippery slope John H Court, Australia	12
The intelligence test-battery: AID 2 as a prototypical globalised test Klaus D. Kubinger, Austria	13
Large Scale Assessment at the Austrian Educational Standards: a Review Klaus D. Kubinger, Martina Frebort, Lale Khorramdel, Elisabeth Weitensfelder, Philipp Sonnleitner, Christine Hohensinn, Manuel Reif, Kathrin Gruber & Stefana Holocher-Ertl, Austria	15
DEBATES IN PSYCHOMETRICS	
Fundamental Problems in Psychometrics John Raven, United Kingdom	16
On the need to secure psychological test materials Jacques Grégoire & Thomas Oakland	18
NEWS	18
CONFERENCE WATCH	
THE INTERNATIONAL TEST COMMISSION (ITC), 7TH CONFERENCE HONG KONG 2010	20

AN OFFICIAL PUBLICATION OF THE INTERNATIONAL TEST COMMISSION

ITC PRESIDENT'S LETTER

Dear ITC Members,

I am pleased to inform you about the current activities and projects of your organization. The next important ITC activities will be the 6th ITC Conference in Liverpool (14 – 16 July 2008) and the General meeting in Berlin during the XXIXth International Congress of Psychology (July 20-25 2008). I will be pleased to meet you on these occasions. The Council already decided to organise the 7th conference of the International Test Commission in Honk Kong, July 19-21, 2010. It will be the first ITC Conference in Asia where testing is a fast emerging field.



ITC 2008 Conference

The ITC conference (14–16 July, 2008) in Liverpool, UK, has the main theme of 'The impact of testing on people and society'. The overall goal is to bring together researchers, educators, psychologists and testing experts from across the world to discuss the impact of testing on people and society.

The scientific program incorporates workshops, plenary keynote presentations, themed keynotes, individual papers, round-table discussions and posters. Workshops include structural equation modelling (Barbara Byrne), DIF (Bruno Zumbo), IRT (Craig Wells) and communicating test results (Dave Bartram). We have had over 160 submissions to the conference including 18 symposia submissions, 26 poster submissions and 4 round-table/discussions. Individuals in a number of different countries (including the US, Canada, UK, Netherlands, New Zealand, China, Israel, Turkey, Sweden, Croatia, Spain, Lithuania and Belgium) have submitted papers.

ITC web site

Initially developed by Dave Bartram, and then by Iain Coyne, the ITC web site was recently revamped. Currently hosted by a server of the Rotterdam University, the new web site was developed by Arjen Karten, from the Rotterdam University, under the supervision of Marise Born, ITC President-elect. Visit the ITC Web site to find all the last information about your organization and its activities on test and testing. You are also invited to send us International News for publication on the website. Keep the ITC web site address bookmarked: <http://www.intestcom.org/>

International Journal of Testing

For the past three years, John Hattie has been the editor of the International Journal of Testing. During that time, the Journal has met every time schedule, has developed a policy for the mission of the Journal re International focus, increased the rejection rate to about 80%, has worked to include articles in IJT in major index systems. The International Journal of Testing is becoming an important part of the international testing literature. As part of the transition of the Journal, and to further enhance the quality and focus, two new Associate-editors were appointed in October 2007: Professor Steve Sireci from the University of Massachusetts and Professor Rob Meijer from the University of Twente, Netherlands. This new editorial team will have a series of initiatives for the Journal, and along with the Editorial Board, will aim to enhance the value of IJT for all readers and the international measurement community.

Next General meeting

The next General meeting will be held during the XXIXth International Congress of Psychology in Berlin, July 20-25 2008 (for information about the Congress, <http://www.icp2008.de/>). During this meeting, we will hold the election of a new President-elect and several other ITC officers. You are kindly invited to participate to the meeting and the election. More information on the meeting is available on the ITC web-site.

Jacques Grégoire
ITC President

ITC Website

News—Guidelines
Publications—Information

www.intestcom.org

Testing International
Submit articles, news, reviews
for consideration to the Editor
Jan Bogg
jbogg@liverpool.ac.uk

ITC sponsored scholars 2008

Information about the ITC 2008 sponsored scholars features below, followed by an article from each scholar on testing in their country.

Dalray de la Harpe

Nelson Mandela Metropolitan University (NMMU)

South Africa

Dalray de la Harpe completed her MA in Research Psychology in 2002 at the University of Port Elizabeth (UPE) and her internship at the Unit for Student Counselling at UPE in the same year. Dalray is currently the Co-ordinator of Research And Development at the Student Counselling Career and Development Centre, in addition she is a member of the University Research, Technology and Innovation Committee and the Research Ethics Committee. Her present interests include health promotion and test development.

Zhou MingJie

Chinese Academy of Sciences.

China

Zhou MingJie, Ph.D., is an assistant professor in the Key Lab of Mental health, Institute of Psychology, Chinese Academy of Sciences. MingJie is involved in two psychological assessment projects, The General Assessment and Prediction System of Sub-health of Chinese and The Evaluation Index System of 0-6 Infants Phased Development. The data gained from the two projects will provide national norms.

Gladiola Musabelliu

University of Tirana

Albania

Gladiola Musabelliu has a Master in Psychological Counseling and works as an assistant professor in the Psychology Department, in the Faculty of Social Sciences. Gladiola is General Secretary of the Albanian Association of School Psychology, Tirana and is also a member of the International School Psychology Association Research Committee, (ISPA), USA. In 2006, Gladiola authored 'School Psychology in Albania' in the Sage Handbook of International School Psychology.

Thami Dhladhla

South African Military Academy

South Africa

Thami Dhladhla works in the Department of Industrial Psychology of the South African Military Academy, as a junior lecturer and is also a Masters student with a keen interest in tests and testing. Thami is working on the psychometric properties of the General Health Questionnaire (GHQ 28-item version), within an African setting.

SCHOLARS ARTICLES

Testing in the South African Context

Dalray de la Harpe

Nelson Mandela Metropolitan University (NMMU)

South Africa

Introduction

South Africa's history is one that is dominated by the legacy of apartheid. Apartheid was a system of racial segregation which existed after 1948, with major language and cultural differences fuelling the racial divide. Robinson (1999) asserts that "...with Nelson Mandela's election as South Africa's first black president, the last vestiges of the apartheid system were finally outlawed". However, the quality of life of many has not changed, and group differences still exist based on differential exposure to developmental opportunities. The system might no longer exist, but the effects still remain. These group differences have had, and continue to have, significant implications for the development of appropriate counselling and assessment practices.

Early Test Use

Early South African psychologists enthusiastically imported and adapted various psychological tools and technologies, most notably intelligence tests, for use in education and industry (Foxcroft, 2004; Louw & Foster, 1991). However, South African norms have been developed for only a limited number of psychometric tests, thereby increasing the risk of significantly underestimating the capacities of non-Western South African children (Richter & van der Walt, 2003). Furthermore, the norms of many locally developed instruments are not applicable to the total South African population (Owen, 1998).

In addition, negative perceptions have been reported regarding the role of psychologists in supporting the system of apartheid. For instance, Burnette (1994, in Owen 1998) shares the perception that "for years, South African psychologists were largely responsible for devising employment tests that were used to screen out Blacks from the workplace and from opportunities for development and higher-paying jobs". This kind of political sensitivity very nearly put an end to all psychological test practices in South Africa. Owen (1998) contemplated whether the late 90's would see the end of the road or a new beginning for psychometric testing in South Africa. The outcome of the 90's debate regarding the role of psychological testing in the construction of a new South Africa was fortunately a positive one, with the addition of new legislation as a framework for a new beginning.

A New Beginning

Section 8 of the Employment Equity Act (55 of

1998) was designed to address the matter of psychological testing within the South African context. It states that:

Psychological testing and other similar assessments of an employee are prohibited unless the test or assessment being used – (a) has been scientifically shown to be valid and reliable, (b) can be applied fairly to all employees, and (c) is not biased against any employee or group.

Mauer (2000) states that “In earlier drafts of the Bill that eventually became the EEA, psychological testing was completely forbidden...it was clear that the abuses of the past would have been exacerbated had the legislation turned a blind eye to the situation.” However, certain criticisms have been levelled against the EEA, for instance that: (a) the Act is referring to measurement bias and as such would appear to argue then that the elimination of measurement bias would necessarily prevent unfair discrimination (Theron, 2007); (b) required qualifications of the professionals who should carry out assessment practices are not discussed (Saunders, 2008); (c) not many test users have the skills to establish whether a test is reliable and valid (Saunders, 2008); and (d) there is a critical shortage of experienced test developers in South Africa (Foxcroft, 2004).

Clearly, supportive mechanisms were needed to facilitate the move towards more valid and reliable test use. The Professional Board for Psychology was established as a juristic entity in an amendment to the Health Professions Act (56 of 1974), and then itself facilitated the establishment of a Psychometrics Committee in 1999. The Professional Board for Psychology asserted in a Press Release on 1 July 1999 soon after the formation of its mandate that “there is great uncertainty and confusion in South Africa regarding the use and possible misuse of psychological tests.” The Board stressed that:

...it is the responsibility of the developer of the test to apply to the Board for classification, and it is the responsibility of the psychometrist or psychologist to ensure that any test he/she intends to use has been classified by the Professional Board and that such a test is accompanied by a classification certificate issued by the board.

However, ensuring test developers' adherence to this classification process remains a challenge, as does creating the necessary awareness amongst practitioners regarding the pitfalls of purchasing or using tests that have not yet been classified by the Board.

Current Thinking / Future Talk

It is heartening to note that South African researchers have been becoming increasingly active in firstly, attempting to understand group differences and secondly, in attempting to develop more appropriate test instruments. South African psychology is striving to become more appropriate in its applica-

tion with different groups, in terms of both professionalism as well as research (Stead, 2002). Stead (2002) argues that there is “...a desire to develop indigenous epistemologies and not to become excessively reliant on Euro-American epistemological traditions.” An indigenous approach to counselling would also have positive implications for the development of appropriate tests and assessment practices.

An extensive national needs analysis of test use patterns and needs of psychological assessment practitioners was conducted by the Human Sciences Research Council (HSRC) in 2004 (Foxcroft, Paterson, le Roux, and Herbst, 2004). A comprehensive survey of this kind had never been conducted in South Africa before, and the aim was to assist in the establishment of an effective agenda for the development, adaptation and updating of tests as well as to establish appropriate professional developmental opportunities for practitioners. Twelve recommendations were made as an outcome of the survey, with regards to the following main issues: (a) procedures and systems for test development, classification and use (b) training regarding testing and psychometrics, (c) the adaptation and revision of existing instruments, (d) the development of new culturally and linguistically appropriate tests, (e) monitoring and regulating test use, and (f) centralised leadership and coordination. It is envisioned that the dialogue prompted by this survey will be used together with the lessons from our past to move towards a better future for us all.

References

- Foxcroft, C. (2004). Planning a psychological test in the multicultural South African context. *SA Journal of Industrial Psychology*, 30 (4), 8-15.
- Foxcroft, C., Paterson, H., Le Roux, N. & Herbst, D. (2004) Psychological assessment in South Africa: a needs analysis: the test use patterns and needs of psychological assessment practitioners: final report. Human Sciences Research Council (HSRC) Press. Retrieved 1 June 2008 from http://www.hsrc.ac.za/research/output/output_Documents/1716_Foxcroft_Psychologicalassessmentin%20SA.pdf
- Louw, J. & Foster, D. (1991). Historical perspective: Psychology and group relations in South Africa. In D. Foster & J. Louw-Potgieter (Eds.), *Social Psychology in South Africa* (pp. 57-90). Johannesburg: Lexicon.
- Mauer, K. (2000). *Psychological test use in South Africa*. University of South Africa. Retrieved 1 June 2008 from <http://www.pai.org.za/Psychological%20test%20use%20in%20South%20Africa.pdf>
- Owen, K. (1998). *The role of psychological tests in education in South Africa: Issues, controversies, and benefits*. Human Sciences Research Council (HSRC) Press. Retrieved 31 May from http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/15/cd/64.pdf
- Richter, L. & van der Walt, M. (2003). The psychological assessment of South African street children. *Children*,

Youth and Environments, 13 (1), 1-19. Retrieved 31 May 2008 from http://www.colorado.edu/journals/cye/13_1_Vol13ArticleReprints/PsychAssessmt_AfricaInsight.pdf

Robinson (1999). Apartheid, social and political policy of racial segregation and discrimination enforced by white minority governments in South Africa from 1948 to 1994. In K.A. Appiah, & H.L. Gates (Eds), *Africana: The Encyclopedia of the African and African American Experience*. Retrieved 31 May 2008 from <http://www.africanaencyclopedia.com/apartheid/apartheid.html>

Saunders, E. Assessment in South Africa: Finding method in the madness. *Workinfo.com – human resources for today's workplace*. Retrieved 31 May 2008 from <http://www.workinfo.com/free/Downloads/36.htm>

Stead, G. (2002). The transformation of psychology in a post-apartheid South Africa – an overview. *International Journal of Group Tensions*, 31 (1), 79-102.

Theron, C. (2007). Confessions, scapegoats and flying pigs: Psychometric testing and the law. *SA Journal of Industrial Psychology*, 33 (1), 102-117.

~ ~ ~

The Use and Development of Personality Testing in China

Zhou MingJie Zhang JianXin

Key Lab of Mental Health Institute of Psychology
Chinese Academy of Sciences
China

The origin of Testing in China

There has been a long time since Psychological Testing was considered and implemented in China. In an ancient medical work named "The Orthodox Medical Classic of Huang Di", the writer observed and measured five kinds of people, Excess Yin, Excess Yang, Less Yin, Less Yang, Balance of Yin and Yang. This could be regarded as the prelude of Personality Testing in ancient China. In the Epoch of Warring States (403-221 B.C.), Mencius had known that the psychological difference exists generally, and also had the concept of difference identity and difference equidistancy. He said "Only by weighing a thing, can you know its weight; and only by measuring it, can you know its length. It is so with all things, and especially so with your mind." This is the earliest description of the theory of Psychological Testing in the whole world. The imperial examination system from the Sui Dynasty to Qing Dynasty lasted over 1300 years, this is the source of Chinese modern Psychological Testing. However, from the view of scientific psychology, the beginning and implement of Personality assessment in China started when the Cultural Revolution ended in 1976.

The Introduction of personality scales

Research on Chinese personality arose from those on Chinese nationality, which resulted from philosophic consideration. With the emphasis of Positivism

paradigm, the core task of Chinese psychologists was to develop suitable scale to assess Chinese personalities. Then from early 1980s, a series of personality scales were introduced into China. In this area, Song W Z introduced MMPI into China in 1980s and then after more than 10 years, Zhang J X et al. (1999) standardized MMPI-2; Chen Z G (1983) analyzed the items of Eysenck Personality Questionnaire (EPQ); Zhu B L & Dai Z H(1988) modified Chinese norm of 16 PF. Besides these, M. H. Bond etc. first translated NEO PI-R into traditional Chinese, and then adapted for Mainland China by Yang J (1999), and Dai X Y et al. (2004,2005) confirmed the Reliability and Validity of the NEO-PI-R in China. Generally speaking, these researches showed that these scales are of good reliability and validity, but there are also some items are not suitable for Chinese context, eg. Chen Z G(1983).

The Influence that Chinese traditional culture put on Personality testing

Besides introducing personality scales from outside, Chinese scholars focused on the uniqueness of Chinese traditional culture, also on how it impacts the personality testing in China. Li M R(2007) discussed Taoist personality scale, found six dimensions in Chinese personality named as wisdom, charity, mature, sturdily, straightforward and glamorous. Yan G C(2008) reviewed the four-factors personality structure from Mencius: Ren(Benevolence), Yi (Rightouse), Li(Courtesy), Zhi(Wisdom), and analyzed the basis of five characteristics of each factor. Yang Q L, Xue C C(2006), developed the Five-Pattern Personality Inventory based on the theory of a famous ancient medical work "The Orthodox Medical Classic of Huang Di" which is mentioned before as the prelude of Psychological Testing in China.

Exploring potential structure of Chinese personality by two large indigenous personality scales

There are several personality traits theory, such as 3-factor model (EPQ), 5-factor model, even 16-factor model (16PF). However, how many factors Chinese personality are there? When introducing the exotic questionnaires into China, indigenous psychologists also developed tools to describe Chinese personality, as well as detect latent structure of Chinese personality. Some of them took psycholinguistic studies on Chinese personality from natural language, such as Chen Z G etc.(1984), Huang X T (1992), Wang D F (1995), these studies led to the Chinese Personality Scale (also named as Qing nian Zhong guo Personality Scale, QZPS,2003). which includes 7 personality factors, they are: extroversion, good-heartedness, emotionality, talent, interpersonal relations, diligence, and honesty. Besides QZPS, some other researchers, such as Cheung F. M, Leung K., Zhang JX., Song WZ.

(1993,1996,2001) combined emic and etic approach in developing an indigenous personality scale, Chinese Personality Assessment Inventory (CPAI), which disclosed a Four-Factor Model, they are: leadership; reliability; tolerance and interpersonal relations. In a joint factor analyses on data from NEO PI and CPAI, a Six Factor Model (SFM), rather than five factor model (FFM) emerged. Later studies showed that IR (Interpersonal relatedness) factor is a culturally salient personality trait factor in Chinese people, but does also exist latently among western people. Just in contrary, O (Openness to experience) is a salient factor among western people, but latent among Chinese people.

The use of personality testing in China

From 1978, psychological testing has progressed obviously in China. Some personality scales have been introduced into China and localized by Chinese psychologists, some have been developed originally. In academic research, 8 scales are often used by scholars. We have searched the personality scales in the China National Knowledge Infrastructure (CNKI), the biggest Chinese academic database, to determine use of these scales (Table 1).

Table 1: Use of personality scales in China

Scale	Scale used or mentioned	Titled
CPI	204	13
16PF	1286	46
MMPI	831	88
EPQ	1602	44
MBTI	96	17
CPAI	44	7
QZPS	119	9
NEO-PI/ NEO-FFI	43	513

In the clinical domain, MMPI is the main tool to test the psychosis/mental disorder in mental health organizations of China. In talent selection domain, MBTI and 16PF are widely used. However, some Chinese psychologists have reflected on the background and the uniqueness of Chinese culture and developed personality inventories suitable for Chinese enterprises. Xu Z C, Gan Y Q, et al.(2000) developed the Chinese Personality at Work Questionnaire (CPW), they chose items from the item pool from the Assessment and Development Centre (1996), and found 15 CPW dimensions. Cun X G (2003) did research on the construction of Chinese Vocational Personality Sorter(CVPS) in his doctoral dissertation, and tested its validity by sampling GuangZhou enterprises' employees.

Summary: The situation & future of personality testing in China

In a word, the development of personality testing in China started with introducing western personality scales, then reflected on the validity of them, and developed testing tool suitable for Chinese culture. Regardless of the distinction of every scale, they are all based on the Classical Test Theory(CTT). Some scholars have utilized new testing theory to analyze and build Chinese personality structure, such as IRT (Zhan S R,2006) and probability unfolding models (Cai S G, 2005). This will lead to a leap in research and the implementation of Chinese personality testing.

References

- Cai S G (2005). Probabilistic Unfolding Theory and Research Method in Personality Assessment (in Chinese). *Unpublished Doctoral Dissertation*. Normal University of South China.
- Chen, Z. G.(1983). Item Analysis of Eysenck Personality Questionnaire Tested in Beijing-Districts (in Chinese).. *Acta Psychologica Sinica*, 2, 211-218.
- Chen, Z. G.,& Wang, D. F. (1984). Desirability, Meaningfulness and Familiarity Ratings of 670 Personality-Trait Adjectives(in Chinese). *the resources of Psychology department*, Peking University.
- Cheung, F. M.,& Leung, K., et al.(1996). Development of the Chinese Personality Assessment Inventory. *Journal of Cross-Cultural Psychology*, 27(2), 181-199.
- Cheung, F. M.,&Leung, K., et al.(2001). Indigenous Chinese Personality Constructs: is the Five-Factor Model Complete? *Journal of Cross-Cultural Psychology*, 32 (4), 407-433.
- Cun X Q, Research on the Construction of Chinese Vocational Personality Sorter (CVPS): Sampling Guang-Zhou enterprises' employees (in Chinese), *Unpublished Doctoral Dissertation*, JiNan University.
- Dai, X. Y.,& Wu, Y.Q.(2005). A Study on NEO-PI-R Used in 16-20 Years Old People(in Chinese). *Chinese Journal of Clinical Psychology*, 13(1),14-18.
- Dai, X. Y.,& Yao, S. Q., et al.(2004). Reliability and Validity of the NEO-PI-R in Mainland China (In Chinese). *Chinese Mental Health Journal*, 18(3),171-174.
- Huang, X. T.,& Zhang. S. L.(1992). Desirability, Meaningfulness and Familiarity Ratings of 562 Personality-Trait Adjectives (in Chinese). *Psychological Science*, 5, 17-22.
- Hui, C. H., Gan. Y.Q.,& Cheng, K.(2000). The Conceptualization and Validity of Chinese Personality at Work Questionnaire(CPW)(in Chinese). *Acta Psychologica Sinica*, 32(4), 443-452.
- Li M R(2007), Research on Taoist Personality Structure and Measurement(in Chinese). *Unpublished Doctoral Dissertation*, Normal University of central China.
- Song, W.Z.(1985). Analysis of Results of MMPI of Normal Chinese Subjects(in Chinese). *Acta Psychologica Sinica*, 4,346-355.
- Song, W.Z., Cui, Q.G., Cheung, F.M.,& Kong, Y.Y.(1987). Comparison of Personality Characteristics of University Students in Beijing and HongKong: Analysis of Item Endorsement Discrepancies on the MMPI(in Chinese). *Acta Psychologica Sinica*, 3,263-269.

- Song, W. Z., & Zhang, J. X., et al. (1993). The Significance and Procedures in Constructing the Chinese Personality Assessment Inventory (CPAI) (in Chinese). *Acta Psychologica Sinica*, 25 (4), 400-407.
- Wang, D. F., & Cui, H. (2003). Processes and Preliminary Results in the Construction of the Chinese Personality Scales (QZPS) (in Chinese). *Acta Psychologica Sinica*, 35(1), 127-136.
- Wang, D. F., Fang, L., & Zuo, Y. T. (1995). A Psycho-lexical Study on Chinese Personality From Natural Language (in Chinese). *Acta Psychologica Sinica*, 27(4), 400-406.
- Yan G C (2008). Review of the Four-Factors Personality structure from Mencius. *Studies of Psychology and Behavior*. 6(1), 61-64.
- Yang, J., & McCrae, R. R., et al. (1999). Cross-Cultural Personality Assessment in Psychiatric Populations: The NEO-PI-R in the People's Republic of China (in Chinese). *Psychological Assessment*, 11 (3), 359-368.
- Yang, Q. L., & Xue, C.C. (2006). The Personality Theory of Medicopsychology of TCM and the Five-Pattern Personality Inventory. *Chinese Journal of Basic Medicine in Traditional Chinese Medicine*, 12(10), 777-779.
- Zhan S R (2006), a Study on Applying IRT in Personality tests (in Chinese). *Unpublished Master Thesis*. Jiangxi Normal University.

~ ~ ~

Testing in Albania – A New Field Full of Needs and Challenges
Gladiola Musabelliu
 University of Tirana
 Albania

For those who have never heard about this country I will take the opportunity to explain a little about it and its history, since all of this has influenced a lot in the development of society, psychology and other fields related to human sciences. Albania is an old Mediterranean country, with a history dating from the 4th century BC. Albania is located in Southern Europe and is one of more than 10 Balkan nations. It is bordered by Montenegro in the north, Kosovo in the northeast, the Republic of Macedonia in the east, and Greece in the south. The Adriatic and Ionian Seas lie to the west of Albania and provide beautiful views with mountains, hills, and beaches within an area of 28,748 square kilometers.

Albania was a closed country for 47 years (1944–1991), due to a dictatorship regime and a communist ideology. But in the 1990s, political and economic changes drastically altered Albanians' lives. Under communism, most Albanian households shared similar standards, conditions, and lifestyles, but the changes fostered differences among Albanian households. Changes within the political system and the introduction of a market economy caused radical economic and social reforms.

The changes became very visible in the way people lived and reacted toward stress, radical changes and different difficulties of life. The number of people who suffered from psychological disturbances

increased a lot and is still increasing. This situation augmented the need for professionals in the mental health field. The number of psychiatrists graduating from universities increased and new professions emerged, such as psychologists and social workers, who promoted different ways of treating mental ill people.

In 2000, the first psychologists graduated in Albania. Since that time, psychology has begun to expand its roots in the welfare of people. These were very fragile and 'roots that needed to be watered every day'. School and clinical psychologists are the only specialties that are developed in Albania.

In the last 15 years, the need for high quality psychological and diagnostic tests is increasing and is becoming a crucial point in the treatment of mental ill people. Many professionals have bought tests and have adapted them for use in clinical assessment. Training and practice has been organised more on a private basis, than organised and followed by a concrete plan of training and needs.

In 2000, some professionals who were pediatricians, general doctors or psychiatrists were trained in Swiss for using intelligence tests such as Griffith and SON-R. They were trained in continuity for about 7 years. The use of Griffith is very popular and it is considered as the only standardized test, but in fact there is doubt on the procedures that may have been used during this process.

The diagnosing and assessing tests that are used in Albania are: WISC-R in children, WAIS in adults, SDQ for the general symptomatic in children, Hamilton Rating Scale for depression, Brief Psychiatric Rating Scale, Beck Inventory for Depression, Global Assessment of Functioning, SON-R. These are all non-standardized test that lack credibility in Albania. These tests are used in psychiatry and national and local centers of mental health, mainly for children.

Tests are not used in the employment field or evaluation of ones performance. Everything is done based on national laws, which give a general sense of evaluating ones performance at work. Also there are no tests for assessing children's behavior at school or kindergarten; this is done based on individual perception.

There is no public or private entity to address psychometric problems and no psychometric experts exist in Albania. The good will of the professionals is not enough to ensure the quality and the credibility of these tests. Every one is aware of the difficulties and those who work in the mental health field perceive this as very important. As the standardisation of tests is the best way to provide adequate care for mentally ill people.

The standardisation of tests is urgently needed in Albania. This requires funding, support from the political environment and dedication from the current professionals in Albania. There are many professionals that are very interested in test standardisation, who would contribute to this work.

Psychological Testing in South Africa:

Current Issues and Challenges

Thamsanqa. J. Dhladhla & François de Kock

Department of Industrial Psychology

Stellenbosch University

South Africa

In more ways than one, South Africa finds itself at the point where the developed and developing worlds meet, seen from a socio-economic perspective. The issues and challenges surrounding psychological testing as experienced in the rest of the world are exacerbated in a sharply contrasting society where aspects of development, language, culture and diversity jointly affect the practice of psychological testing.

In South Africa, psychological testing is actively used for various purposes, mostly in therapeutic, educational, sociological, and occupational applications (Owen, 1998). This article will address the issues and challenges related to psychological testing in our country, by providing a brief historical development of testing in South Africa as background, followed by a broad discussion of trends and challenges faced by psychologists involved in testing.

Development of psychological testing in South Africa

The use of psychological tests in South Africa (SA) has largely followed international trends, beginning in the 1900s, where imported tests were used with mainly white test takers (Foxcroft, 1997; Huysamen, 2002).

The way in which psychological testing was introduced stemmed from the colonial heritage, where speakers of African languages who comprise the majority of the population, were excluded from psychological testing due to the lack of African language versions of popular tests (Claassen, 1997). Between 1960 and 1984, research on the equivalence and bias of assessment instruments was non-existent, apart from some work on group differences in cognitive ability test scores (Owen, 1992). More recently, the cross-cultural equivalence of personality measures has emerged as a research trend that focuses on instruments such as the 16 PF (Meiring, 2007) and the NEO PI-R (Heuchert, Parker, Strumf, & Myburg, 2000).

Despite many decades of being considered a 'western approach' and even an 'apartheid instrument' by large segments of society, psychological testing has more recently become increasingly accepted and valued for its contribution to mental health services, business, education and economic development. The profession as a whole benefited from transformation in its regulation, practices, as well as the ethnic profile of its practitioners. The Health Professions Council of South Africa (HPCSA), which oversees the functioning of all

health-related professions in the country, also enhanced the professionalisation of psychology through the auspices of the Professional Board of Psychology, which regulates psychological testing and test use.

Recent developments in labour legislation have massively impacted testing on testing practices, where psychological testing is strictly regulated by law. Apart from our very liberal Constitution, the Employment Equity Act 55 of 1998 (Government Gazette, 1998) stipulates that:

Psychological testing and other similar assessments are prohibited unless the test or assessment being used – (a) has been scientifically shown to be valid and reliable, (b) can be applied fairly to all employees; and (c) is not biased against any employee or group. (S. 8).

In this regard, Van de Vijver and Rothman (2004) commented that South African law differs from the international trend. In most countries the fairness of psychological tests is assumed, unless proven otherwise, and discrimination and unfair treatment in psychological assessment are forbidden. In our case, test use is forbidden in the absence of acceptable psychometric evidence for test use.

From a general testing perspective, the South African situation is not dissimilar from other countries. Competing needs of various stakeholders constantly interact where, on the one hand, the users of testing (e.g., business, industry and practice) require increasingly effective, accurate, yet cost-efficient assessment instruments that are perceived as fair (Tustin, 2007). On the other hand, government and professional associations act as advocates of societal needs for the redressment of past discrimination, socio-economic upliftment, and ensuring compliance and adherence to professional and ethical guidelines. In the middle, sometimes unenviably, are the academics, researchers and test developers that attempt to satisfy both needs based on good science.

Trends in psychological testing in South Africa

South Africa shares many similar testing issues faced internationally. These include the awareness of test taker perceptions of fairness, need for ethical use of tests, increasing use of online testing, test-taker cheating, and assessment in a multi-cultural society. Being a developing country, some of these issues are more salient, such as language, illiteracy, and dealing with the effect of separate development in testing. Some have tried to address the challenge of unequal developmental opportunities with a preference for learning potential assessment and assessment for developmental purposes.

Naturally, the cross-cultural aspect of assessment is

very salient, considering the heterogeneity of the South African population. Since we have eleven official languages, and the majority of our population has a mastery of at least two languages, the study of language in testing in South Africa is literally a seedbed for research. Recent surveys (e.g., Foxcroft, Paterson, Le Roux, & Herbst, 2004; Tustin, 2007) have sought to investigate trends in testing in SA from the practitioner's perspective. Generally, practitioners saw value in testing provided that culturally appropriate, psychometrically sound, high quality tests were used. However, they were concerned that the majority of the tests being used frequently were in need of adapting for our local multi-cultural context. They expressed the need for tests that were available in all official languages were regularly updated, and that had appropriate and specific norms. Practitioners also identified the need for increased awareness of ethical practice of assessment in general, and computerised testing in particular.

The results of these surveys suggested that many practitioners questioned the success of the Professional Board for Psychology in controlling and regulating test use. The idea of establishing a new self-regulating body (e.g., a Centre of Excellence for Testing) was even mooted to monitor test use, advise practitioners, research and review tests, and to monitor and coordinate test development, adaptation, and updating. Interestingly, the same pattern seems to be emerging where the international trend is that professional organizations of psychologists established by professionals themselves regulate the profession. In this regard, Van de Vijver and Rothman (2004) mention that:

In various countries issues of bias and fairness are not primarily enacted in national laws, but in codes defined by and enforceable on their members. Although many countries have both legal and professional regulations, their enforcement shows considerable cross-cultural variation. For example, whereas in South Africa court cases are the main option available to plaintiffs, in a country like the Netherlands the ethics committee of the national association of psychologists is more likely to see a complaint being filed than is one of the courts. (p. 1).

Challenges in psychological testing

Given our history and the legacy of separate development, the task for psychology in South Africa remains huge. These challenges centre mainly around issues of regulation and compliance, the stigma associated with psychological testing, dealing with language and cultural heterogeneity, and the use of test scores.

Regulation and Compliance. Various recent structural changes of the profession have been implemented in the last decade. Apart from establishment of new bodies and legislation to regulate the profession, these changes include a revision of the scope of practice of psychologists and other practitioners involved in psychological testing, the introduction of a continuous professional development (CPD) system for registered professionals, and the revision of academic degree programmes and qualifications, and lastly, new training requirements leading up to professional registration. Usually, substantial change on this scale does not go without its growing pains. Though mostly successful, these initiatives sometimes still fall short in achieving their objective. Makgoke (2004) highlighted the concern of the Professional Board for Psychology about the growing inappropriate use of psychometric tests, such as some test distributors and registered psychologists that provide training to unregistered persons, classifying them as certified users, and also using tests that are not registered with the Board. This is clearly contrary to the ITC International Guidelines for Test Use that stipulate that ethical testing and assessment practices require that the assessment practitioner use tests appropriately, professionally, and in an ethical manner, paying due regard for the needs and rights of those involved in the testing process, and the broader context in which testing takes place (Foxcroft, 2002). This situation has a potential of discrediting all the good work that has been done in the professionalisation of the discipline.

Stigma Associated with Psychological Testing.

The view that psychological testing stems from colonial heritage is a source of resentment towards psychological testing in Africa in general (Stead, 2002). It resulted from the perception that psychological testing instruments were used to screen out blacks from the work place (Owen, 1998). With the advent of democracy in South Africa in 1994, the major overhaul of the professional landscape of psychology and increasing racial transformation of regulatory bodies and the practitioner corps have contributed to enhancing public perceptions of psychological testing in South Africa (Painter & Terreblanche, 2004).

Language and Cultural Heterogeneity. The main concerns in psychological testing in South Africa relate to language, diversity, and equity (Pretorius, 2008). Very few tests have been translated into the nine official African languages, while only 10% of citizens speak English as a first language. Moreover, the level of English language among Black South African is generally not comparable to mother-tongue speakers (SAIR, 1997). The lack of availability of parallel language versions of tests that have been appropriately adapted for the cultural

group of a particular test-taker represents a danger to good testing practices (Foxcroft, 2001). Test bias is a major concern because such tests were initially developed for whites and are not always appropriate to those whose first language is neither English nor Afrikaans (Claassen, 1997).

Mpofu (2002) lamented the fact that western practices are often applied to African communities without any cultural adaptation. Therefore, it is often suggested that assessment practitioners should be sensitive to culture in assessment. When developing and using psychological tests, the true ability of the individual should be measured without undue influence of culture and/or group affiliation (Foxcroft & Roodt, 2005)

Van de Vijver and Rothman (2004) have proposed four kinds of procedures for dealing with multicultural assessment, including establishing the equivalence of existing instruments, defining new norms, developing new instruments, and studying validity-threatening factors in multicultural assessment. A number of various recent studies seem to be addressing these identified needs.

Use of Test Scores. Ironically, the biggest issues surrounding testing involve not the tests *per se*, but rather how test scores are used. In a recent survey, South African organisations report the misuses of the results of assessment, and inconsistent use of assessments as the strongest weaknesses in psychological testing (Tustin, 2007).

The fact that the use of certain psychometric tests (e.g., cognitive tests) traditionally demonstrates adverse impact against previously disadvantaged groups (usually called 'minority groups' in Western countries) has major implications for test developers, regulating bodies, and practitioners. Internationally, the use of psychometric tests in selection for employment has been singled out for intense scrutiny from the perspective of fairness and affirmative action (Arvey & Faley, 1988). Strangely enough, very little litigation surrounding psychological testing in assessment and selection has followed in South Africa. This is surprising, considering our liberal constitution and culture of human and worker rights enshrined in legislation.

In personnel selection, more specifically, the use of psychometric tests has been regarded with an extraordinary degree of suspicion and scepticism. The unintended consequence is that test developers and practitioners are increasingly pursuing psychometric tests that are 'Employment Equity Act-compliant': they try to find or develop alternative tests that lead to less severe adverse impact (e.g., situational judgment tests) and consider innovative ways to use test scores for decision-making (e.g., banding, group norming). Theron (2007) suggests that this effort is misdirected, and is of the opinion that an inappropriate focus on compliance has obscured the hard fact

that affirmative development should be the primary priority in order to address the underlying social issues that could cause observed group differences.

Conclusion

The answers to many of these questions lie in the ability of stakeholders to jointly find ways to resolve sometimes competing needs in mutually beneficial ways. Rather than being considered a threat, legislation should be viewed as an opportunity and catalyst to propel best practice in psychological testing. A central theme in the success of such an approach would be the education of test-takers and client organisations regarding their rights in testing, seeking greater acceptance for the tools we use, and influencing perceptions of the value and fairness of psychological testing. Unquestionably, various opportunities for learning and advancing our knowledge of psychological testing abound in South Africa.

References

- Arvey, R. D., & Faley, R. H. (1988). *Fairness in Selecting Employees* (2nd Ed.). Reading, MA: Addison-Wesley.
- Claassen, N. C. W. (1997). Cultural Differences, Politics and Test Bias in South Africa. *European Review of Applied Psychology*, 47, 297 – 307.
- Foxcroft, C. D. (1997). Psychological Testing in South Africa: Perspective Regarding Ethical and Fair Practices. *European Journal of Psychological Assessment*, 13, 229 – 235.
- Foxcroft, C. D. (2001). Reflections on Implementing the ITC's International Guidelines for Test Use. *International Journal of Testing*, 1 (3 &4), 235 – 244.
- Foxcroft, C. D. (2002). Ethical Issues Related to Psychological Testing in Africa: What I have Learned (so far). In W. J. Lonner, D .L. Dinnel, S.A. Hayes, & D. N. Sattler (Eds.), *Online Readings in Psychology and Culture* (Unit 5, Chapter 4), [<http://www.wvu.edu/~culture>], Centre for Cross-Cultural Research, Western Washington University, Bellingham, Washington, USA.
- Foxcroft, C.D., Paterson, H., Le Roux, N., & Herbst, D. (2004). *Psychological assessment in South Africa: A needs analysis. The test use patterns and needs of psychological practitioners* (Human Sciences Research Council Rep.). Pretoria: HSRC.
- Foxcroft, C., & Roodt, G. (Eds.). (2005). *An Introduction to Psychological Assessment in the South African Context* (2nd Ed). Cape Town: Oxford University Press.
- Health Professions Council of South Africa. (2006). Why Do We Classify Tests? *Form 207: List of Tests Classified as Being Psychological Tests*. [<http://www.hpcsa.co.za>] 5 June 2008.
- Heuchert, J. W. P., Parker, W. D., Strumpf, H., & Myburg, C. P. H. (2000). The Five-Factor Model for African College Students. *American Behavioural Scientist*, 44, 112 – 125.
- Huysamen, G. K. (2002). The Relevance of the New APA Standards for Educational and Psychological Testing for Employment Testing in South Africa. *South African Journal of Psychology*, 32, 26 – 33.
- Makgoke, P. (2004). *Board is Concerned About Admini-*

stration. Newsroom: Communications Officer, *Health Professions Council of South Africa*. [<http://www.hpcs.co.za/hpcs/news>] 3 June 2008.

- Meiring, D. (2007). *Bias and Equivalence of Psychological Measures in South Africa*. Ridderkerk: Labyrinth.
- Mpofu, E. (2002). Psychology in Sub-Saharan Africa: Challenges, Prospects and Promises. *International Journal of Psychology*, 37 (3), 179 – 186.
- Owen, K. (1998). *The Role of Psychological Tests in Education in South Africa: Issues, Controversies and Benefits*. Pretoria: Human Sciences Research Council.
- Painter, D., & Terreblanche, M. (2004). *Critical Psychology in South Africa: Looking Back and Looking Forward*. [<http://www.criticalmethods.org/collab/critpsy.htm>] 3 June 2008.
- Pretorius, H. G. (2008). Race, Sex and Class in Psychology: A Vision for Hope for a Fair and Just South Africa. *Journal of Psychology in Africa*, 18 (2).
- Stead, G. B. (2002). The Transformation of Psychology in a Post-Apartheid South Africa: An Overview. *International Journal of Group Tensions*, 31 (1), 79 – 102.
- South African Institute of Race Relations. (1997). *South African Survey 1996/7*. Johannesburg, South Africa: Author.
- Theron, C. C. (2007). Confessions, Scapegoats and Flying Pigs: Psychometric Testing and the Law. *SA Journal of Industrial Psychology*, 33 (1), 102 – 117.
- Tustin, D.H. (2007). *Issues facing organisations using assessment in the workplace. People Assessment in Industry (PAI)*. Pretoria: Society for Industrial and Organisational Psychology.
- Van de Vijver, A.J.R., & Rothmann, S. (2004). Assessment in Multicultural Groups: The South African Case. *SA Journal of Industrial Psychology*, 30 (4), 1 – 7.

Four years ago, in May 2004, Lithuania formally became a full member of the European Union. A more in-depth analysis of individuals, communities, public and private enterprises, however, shows that a brief process of revolutionary change of a system does not ensure the reaching and change of basic everyday life standards and practices in at a fundamental, grass-roots level. Many challenges arise when implementing democratic principles, personal and institutional responsibility, new practices in communication and relationships. A prime example of such a challenge is in the field of psychological testing and assessment, the field to which the 6th ITC Conference is devoted.

When one is trying to evaluate the situation of psychological assessment in a country (especially one in transition) it is necessary to consider such an evaluation in its total context, taking into account the development of other aspects of psychology within that context. From 1579 (the year in which Vilnius University was established) until the early 1970s psychology in Lithuania was primarily a teaching discipline for teachers, philosophers, economists etc. Only in 1969 a real comprehensive study programme in psychology began. From 1974, when professionals with a diploma in psychology began working in applied settings such as in the education and health care systems, the need for research and assessment tools appeared. The need for teaching, creating and using these tools also appeared. The closed Soviet system had been the main obstacle for receiving information and extensively developing psychodiagnostics, psychological testing and assessment.

The LSP was established at Vilnius University by the Ministry of Education in 1973 and according to an agreement with the Lithuanian Fellowship of the Blind. For 20 years this Fellowship was the main financial sponsor of LSP. The main objectives of the LSP was to carry out comparative psychological studies of persons with and without visual impairments. In the beginning we started to create our own psychological tools for research and measuring different psychological functions (visual, acoustic, haptic perception, attention, memory, reasoning etc.). We also tried to adapt different foreign tests of intelligence and personality. At one point, we think it is necessary to confess the using methods and tests developed in West countries without permission and licenses of owners. Such practice was common in the entire Soviet region (perhaps with the exclusion of Czechoslovakia which had at that time its own Center for Psychodiagnostics in Bratislava which was working according to international requirements). In the Soviet Union, such practice occurred because of: very low culture of psychological research and testing, lack of financial resources, isolation, and absence of communication with foreign

ARTICLES

The Challenges of Fostering a Culture of Psychological Testing in a Small Post-Soviet Country: The Experience of the Laboratory of Special Psychology of Vilnius University **Albinas Bagdonas** Vilnius University Lithuania

Introduction

In this short paper an overview of some the achievements and problems of institutionalisation of psychological assessment in Lithuania over the past 35 years will be given. These observations will be based mostly on my experience of working in the Laboratory of Special Psychology (LSP) of Vilnius University.

Background

Lithuania re-established its independence in 1990 and is in the process of changing its political, economical and social orientation to the outlook, traditions and standards of Western European countries.

colleagues. The practice of using tests without permission of the owners however was never used for commercial purposes. We know that such illegal practice exists in some post-soviet republics.

Development from the eighties onwards

In 1984 the LSP received permission from the Verlag für Psychologie (of Germany) to adapt and standardise two intelligence tests (IST-70 and PTV). It was the first official permission for Lithuania to use test developed in a foreign country. However this first attempt was unsuccessful and these two tests were standardised only in 2007 in the Department of General Psychology of Vilnius University.

After re-establishing independence the Lithuanian Psychological Association created the Commission for Using Methods of Testing (currently the Committee for Psychological Assessment). In 2003, the LSP sponsored the publication of the *International Guidelines on Test Use: Version 2000* (ITC, 2000). According to the Agreement between the Psychological Corporation (later Harcourt Assessment and now Pearson Education) and the LSP, the LSP adapted and standardised the Lithuanian version of the WISC-III in 2002. At the present time we are completing the standardisation of the WAIS-III and WASI. From 2009 we will begin the adaptation of WPPSI-III. According to agreement with PAR (Psychological Assessment Resources) LSP started the adaptation of the NEO PI-R and NEO-FFI.

What conclusions could be made from the experiences of the LSP on issues of adaptation and standardisation, of tools for psychological assessment?

The main problem for a small country like Lithuania is its small population and few test users. Other problems associated with population size include:

- 1.Limited human professional resources (not enough professionals for the creation and adaptation of tests)
- 2.Low general psychological culture (especially culture of psychological assessment), at all levels
- 3.Parapsychologisation of society (a very sensitive issue for a small country)
- 4.Lack of financial resources
- 5.Lack of national policy in the field of psychological testing
- 6.Difficulties in harmonising the interests and requirements of all stakeholders including: test owners, financial sponsor (usually very bureaucratic), test adaptors and test users (other issues that arise include ethical requirements, requirements of ITC, national legislation etc.)
- 7.Lack of legislation and control mechanisms for using standardised tests inside country (distributed practice to use tests illegally)
- 8.Limited number of professionals with high competence in teaching test developers and users

9.High dominance of biomedical tendencies in assessment of the person's functioning

The critical mass for developing and publishing tests where costs can be recovered on market-based principles begins for countries with populations of at least 5-7 million. For smaller countries like Lithuania there needs to be additional financial funding to help compensate the price of production. For example, the standardisation of four assessment tools – the WAIS-III, WASI, ICF-based *Scales for Assessment of Efficiency of Functioning* and the *Lithuanian Professional Interests' Inventory* needed the support from the European Structural Funds and the Lithuanian Government to make the realisation of using these tests possible. For this purpose a project was undertaken: *Development of the Methods for Assessment of Functioning (Employability, Special Needs, Professional Abilities) of Persons with Disabilities and the Recommendations for Their Application* (support under measure 2.3 of the SPD, ESF, 2005-2008).

~ ~ ~

From Normal Curve to Slippery Slope

John H Court Ph.D. (retired)

Jhcourt@bigpond.net.au

Australia

In 1997 Boyer wrote “*We urge that student evaluation be used in making decisions about tenure and promotion. But for this to work, procedures must be well designed and students must be well prepared.*” (p.40). In the nineties, a national initiative across Australian universities resulted in development of the Graduate Course Evaluation Questionnaire (GCEQ) to determine student satisfaction following graduation and to benchmark all participating universities. The 25-item instrument, in paper and pencil format was well researched for psychometric properties. So far, so good.

Taking up the challenge to improve student learning experiences, one university responded by introducing in-house evaluations for all undergraduate courses, borrowing from the national instrument. A series of steps over a five year period indicates how easy it is to move from a well constructed instrument with a well-defined purpose, to the accumulation of numbers with no credibility, used for different purposes and in different ways, yet leading to significant decisions way beyond the original purpose. An interview with the academic responsible for these developments is revealing, and brief excerpts are included below.

First the GCEQ was shortened from 25 items to 10 by a committee. “No study has shown that this selection correlates with the larger item pool, and we don't know the validity of the currently used instruments.” The paper and pencil administration in class was followed initially, but gradually changed across

to requests for staff to encourage online responding. The rationale for this was “to maximize anonymity and because we can’t afford paper and pencil administration”. This then moved further to a compulsory use of the online mode.

This major shift calls into question any residual validity or reliability. It means that the instrument is undertaken at various times around the end of a course (invariably before final assessment is completed, though questions relate to that), and the percentage of students deciding to respond could be expected to plummet. In fact comparisons across several courses were possible, showing an average response rate for paper and pencil of 59%, with online averaging 11%. Recognising the problems, a researcher spent a year trying to enhance response levels and reported achieving an increase of 17%, which actually meant a shift from 11 % to nearly 13%. This might suggest there is an inherent problem in the approach. Under such circumstances all assumptions about a normal distribution have to be discarded. Academics recognise that this small number is largely composed of the disaffected and the very enthusiastic in classes, so this bimodal distribution is of little help in identifying good learning, or in recommending changes.

The shift of modes raises an ethical dilemma which has been addressed by Susan Whiston (2000) viz. “If the instrument is an adaptation of a paper and pencil instrument, then the evaluation of the psychometric qualities must include an analysis of the equivalency of the two forms of the instrument.” (p.352). Nonetheless the data have continued to be a source for evaluating course outcomes. In addition, they have moved to become compulsory for all courses at all levels, forming a component of evaluations for academic promotion.

After five years of usage, do we have evidence of improvements in teaching arising from these data? “No”. Do the undergraduate data generally conform to the normal curve of distribution? “We haven’t looked at that”. A psychologist considering the student experience might ask what is going to happen if students are invited at the end of every course in their degrees to respond to an online satisfaction survey. Since the responses are anonymous there is no capacity to offer reinforcement for responding, so one could predict decreasing interest through the years of the first degree. Students continuing on to higher levels may encounter 30 or more such evaluations. Not surprisingly therefore, with contingent reinforcement absent, in graduate programs response numbers are often in single figures, heading for extinction of the response. This is a pity since pre-existing paper and pencil evaluations ran at better than 85%. Remarkably, therefore, the next step was to make these data available as essential information for academics seeking promotion. It would seem to follow that high ratings might be viewed fa-

vorably. However, responding to the comment “Presumably data from high quality instructors is negatively skewed” the response was “They should be but we don’t know”.

These steps away from use of psychometrics to an ad hoc application of numbers are not presented as a criticism of the particular location where these decisions were made. They are identified as an example of what can happen all too easily where a results-based institution is looking for ways to demonstrate its attention to accountability. This often links to funding decisions as well as perceived reputation, so the pressures for numbers are great. So are the hazards of adapting, abbreviating, and modifying procedures without developing evidence on the effects such changes have on response patterns.

This commentary is intended as a working example of the issues raised by the ITC conference addressing issues in test usage. There are times when a misguided commitment to test data that appear to have a respectable pedigree can lead to a situation where bad data are worse than no data. Psychologists required to participate in such procedures face interesting ethical dilemmas, poised between their code of professional conduct and the expectations of employers.

References

- Whiston, S. C. (2000). *Principles and applications of assessment in counseling*. Brooks/Cole Thomson Learning.
- Boyer, Ernest L. (1997). *Scholarship reconsidered: Evaluation of the professoriate*. San Francisco: Jossey-Bass.

~ ~ ~

The Intelligence Test-Battery AID 2 as a Prototypical Globalised Test

Klaus D. Kubinger

University of Vienna, Faculty of Psychology
Division of Psychological

Assessment and Applied Psychometrics
Austria

The AID 2 (*Adaptive Intelligence Diagnosticum* - Version 2.2; Kubinger, 2008) is a German language intelligence test-battery for 6 to 16 years, in practical use since 1985. It consists mainly of adaptive tests using branched testing design based on a Rasch model item calibration. It is intended as a consulting instrument for school psychology as well as for (neuro-) clinical psychology, applicable for instance in making curricular decisions or identifying learning disabilities. A detailed presentation of AID 2 is given by Kubinger (2004); its theoretical embedment into traditional intelligence theories is described by Kubinger, Litzenberger and Mrakotsky (2006).

The last edition of AID 2 now has an addition: the test-battery AID 2-Turkish. This is not just an adaptation for the Turkish population, but a globalised

test version. As a matter of fact, German speaking countries (Germany and Austria, excluding Switzerland and Southern Tyrol) have about 90 million people, more than 2.5 million of whom are Turkish. Hence, Turkish immigrants are of substantial social relevance in this area. And of course, children with Turkish as a mother tongue need the same kind of psychological consulting as children with German as a mother tongue, or even additional psychological assessment and intervention. Yet translation of pertinent intelligence tests does not solve the problem. Firstly, there may be cultural differences handicapping children with Turkish as a mother tongue in regard to certain items or even to entire subtests. Secondly, children with Turkish as a mother tongue living in German-speaking countries may differ with respect to their cultural background to those still living in Turkey. Thirdly, children with Turkish as a mother tongue represent different generations of immigrants; there is the third, the second, and the first generation of immigrants. And last but not least, children with Turkish as a mother tongue differ with respect to their German language competence – bear in mind, that the language of instruction is always exclusively German. Therefore, there are children who are proficient in German in an ‘academic’ context (that is within written, read, and spoken communication on school subjects), but who are not used to speaking (reading or writing) Turkish in that context, while at home and privately with their peers, they speak primarily Turkish. These children would indeed be better tested in German. There are, however, other children who are socially well-integrated into their German peer group and therefore, particularly if the same is true for their parents use German for everyday communication, while they still do their ‘academic’ job better in Turkish. Finally, the sketched polarisation is not at all universal; children with Turkish as a mother tongue sometimes prefer Turkish within one ‘academic’ context, but German within another.

For this reason, the AID 2-Turkish was designed as follows: Because most of the subtests are to be administered adaptively, the item pool of those subtests is grouped in 5-item-sets. Such 5-item-sets are classified according their difficulty into eight stages and, after every 5-item-set, the next level is chosen according to the number of solved items. If a child with Turkish as a mother tongue solves no more than a single item of the first 5-item-set administered in German, then that 5-item-set is administered again in Turkish. Depending on the test result, the next 5-item-set is chosen according to the proper level of difficulty and the administration language may be changed as well. This means that there are multiple checks of language competence through test achievement and, as a consequence, possibly also multiple administration language changes. The most relevant practical issue is that

there is no need for a native speaker, nor does the test administrator have to be fully conversant with Turkish; all that is needed is a test administrator who is well-trained to articulate the verbally administered items in Turkish and to differentiate between correct and incorrect Turkish responses.

Of course, an equivalence study was done. Analyses using the Rasch model (1 PL-model) disclosed that just a single subtest measures a different psychological dimension for children with Turkish as a mother tongue than for children with German as a mother tongue, but all the other subtests measure the same dimension in both populations. This is true allowing for the fact that for the subtest Everyday Knowledge, 5 of 60 items in the Turkish version had to be deleted because they proved to be much more difficult for children with Turkish as a mother tongue in relation to the other items. Comparison of the resulting Rasch model ability parameters showed that these had a substantially smaller mean in the Turkish population. For this reason, a separate standardization took place.

To summarise, as concerns a special immigrated population within a given geographical socio-cultural community, the intelligence test-battery AID 2 serves as a prototypical globalised test. Though substantial differences in mean scores occur between the immigrant population and the native one, a strategy for testing the immigrants in a fair manner as concerns their various language competencies exists and involves no more than well-trained test administrators, with just minimal foreign language competence. Most importantly, psychometric equivalency analyses have proven that the same abilities were measured in both populations.

References

- Kubinger, K.D. (2004). On a practitioner's need of further development of Wechsler scales: Adaptive Intelligence Diagnosticum (AID 2). *Spanish Journal of Psychology*, 7, 101-111.
- Kubinger, K.D. (2008, in print). *Adaptives Intelligenz Diagnostikum - Version 2.2 (AID 2) samt AID 2-Türkisch* [Adaptive Intelligence Diagnosticum, AID 2-Turkey included]. Göttingen: Beltz.
- Kubinger, K.D., Litzenberger, M. & Mrakotsky, C. (2006). Practised intelligence testing based on a modern test conceptualization and its reference to the common intelligence theories. *Learning and Individual Differences*, 16, 175-193.

In June 2008 International Journal of Testing

- **Objective Standard-Setting for Judge-Mediated Examinations**
- **Generalisability of GMAT® Validity to Programmes outside the U.S.**
- **Clarifying the Measurement of a Self-Structural Process Variable: The Case of Self-Complexity**
- **The Multidimensionality of Verbal Analogy Items**

**Large Scale Assessment at
the Austrian Educational Standards: a Review**
Klaus D. Kubinger, Martina Frebort, Lale Khorramdel, Elisabeth Weitensfelder, Philipp Sonnleitner, Christine Hohensinn, Manuel Reif, Kathrin Gruber & Stefana Holocher-Ertl
University of Vienna, Faculty of Psychology
Division of Psychological
Assessment and Applied Psychometrics
Austria

After the Austrian Educational Standards were settled in 2005, the *Center of Testing and Consulting* (Faculty of Psychology, University of Vienna) was authorized to supervise item development and to calibrate the standard tests for mathematics as well as for reading, both for 4th and 8th grade students. Starting from prototypical examples given as a support for teachers' standards-based instruction, the center and its team of psychologists developed item-generating rules used by a group of teachers accompanied by some didacticians to create item pools.

The general framework was determined by the government and encompassed: a) three years of piloting the tests; b) a large scale assessment but not a survey of the entire population (each year there were about 7000 students from about 50 pilot schools for the 8th grade, and about 2000 for the 4th); c) paper and pencil test administration, though the team of psychologist has meanwhile conceptualized some secure internet testing using mobile computer systems; d) testing at schools in classrooms, using teachers from other schools who had been carefully trained in psycho-educational group testing by the team of psychologists as test administrators; e) feedback of the test results via an internet platform.

It was the responsibility of the team of psychologists to design the entire procedure. This resulted in the following decisions: a) each item pool should ultimately consist of about 500 items, but additional items are to be administered as is obligatory for large scale assessments, for calibration and future substitution, due either to alterations of item contents in adaptation to society's changes, or to the psychological half-life of the items; b) there should be several booklets, but in contrast to many other large-scale assessment tests, every student is to be tested with each subtest of the standard test; c) the latter was implemented so that every student could get feedback on his/her test results, not only as a sign of appreciation, but more importantly to enhance achievement motivation; d) apart from the students themselves, feedback should be given on an aggregated level (only the scores' distribution and mean of course in comparison with the reference population) to three administrative authorities: teachers, heads of school, and the supervisory school authority; e) the calibration of the tests

should be done using the Rasch model (1-PL model), as this allows for the simplest scoring rule (to count the items being solved) and is therefore more likely to hold empirically, than models for multi-categorically benefited item responses. Furthermore, the Rasch model allows conditional maximum likelihood parameter estimations and as a consequence of which specific objective measurement (of the items' difficulty); f) test calibration should occur using state of the art techniques, that is, above all, item deletion (due to differential item functioning) and cross validation (cf. Kubinger, 2005); g) four types of response format should be used: 1) a free text response format though this means more time and effort for administration and scoring and is therefore used very rarely (about 2 items of 35), it seemed necessary as a signal that the educational standards aim to promote complex problem solving and arguing, 2) a free response format with corresponding boxes for the solution's digits or letters, 3) a multiple choice format with six response options, a single correct one and five distractors (1 from 6), and 4) a multiple choice format with five response options, two of them being correct, the other three as distractors (2 from 5), the increased number of distractors and correct options, respectively, should help in minimizing guessing effects.

In order to enable linked item parameter estimations, the booklets needed a multiple, incomplete, balanced block design of items and groups of students. The design was 'multiple', because there are several subtests, for instance four subtests pertaining to four mathematical abilities: *Modeling, Operating and Calculating, Interpretation and Documentation, and Argumentation*. We tried to use a computer algorithm to construct the design, but this seemed for the time being, more difficult than working by hand. A number of parameters must be taken into account: i) four different content areas (*Numbers and Measures, Variables and Functions; Figures and Stereometric Corpora, and Statistics*), which should be equally distributed throughout the booklets, ii) four different response formats, which should be equally distributed as well, iii) three a-priori levels of difficulty, which should be differently distributed according to the given ability levels of the groups of students, iv) three clusters of items being either already calibrated (twice or once), or needing to be calibrated now, v) prevention of pairs of items with exactly the same topic but different problems to solve or just a different response format being in the same booklet.

The carefully developed item-generating rules and the repeated interactive improvement of each item by the group of teachers, the didacticians, and the team of psychologists, proved to be worthwhile: At most 10 % of the items needed to be deleted, and cross-validation always confirmed the calibration results. Multiple choice items in the format '1 from 6'

disclosed considerably lower difficulties than items with all the other response formats, and multiple choice items in the format '2 from 5' showed almost the same difficulty as those with either of the free response formats (details are given in the forthcoming paper of Kubinger et al.).

References

- Kubinger, K.D. (2005). Psychological Test Calibration using the Rasch Model - Some Critical Suggestions on Traditional Approaches. *International Journal of Testing*, 5, 377-394.
- Kubinger, K.D., Holocher-Ertl, S., Reif, M., Hohensinn, C. & Frebort, M. (submitted). On minimizing guessing effects on multiple choice items: A 2-solutions-and-3-distractors item format being better than a 1-solution-and-5-distractors item format.

DEBATES IN PSYCHOMETRICS

Fundamental Problems in Psychometrics

John Raven

jraven@ednet.co.uk

Edinburgh

United Kingdom

The ITC aims, among other things, to “promote responsible and valid tests and testing”. Unfortunately, many widely accepted, indeed prescribed, methods and practices in testing cannot be regarded as anything other than unscientific and unethical. The dilemma was highlighted by Spearman almost a century ago. He argued that the tests from which his *g* had emerged “had no place in schools” because they did not encourage teachers to identify and nurture the diverse talents of their pupils. To underline the point, he went on to assert that all pupils were geniuses at something but that this could not be demonstrated using current psychometric procedures. The evidence we have accumulated¹ suggests that he was right on all counts. Failure to develop a more appropriate psychometric framework has even more serious consequences than failing to help parents, teachers, managers, and others to identify, develop, utilise, and reward the huge variety of talents that are available - thereby stunting most people’s individual growth and depriving them of opportunities to gain recognition for their talents. The most serious consequence is that, because the neglected talents are the very ones that are required to transform our society in such a way that *homo sapiens* will have any chance of surviving as a species, continued reliance on the current testing framework contributes directly to our extinction and probably that of most other species at the same time². What could be more unethical? The deleterious effects of this process in itself are exacerbated

by the publication of numerous studies which, while purporting to contribute to “evidence based practice” in education and health care, are, in reality, incapable, not only of documenting the diverse ways in which people change as a result of involvement in developmental activities³, but even the overall, desired and desirable, and undesired and undesirable, effects of the programmes evaluated⁴. An example may help to make the point:

Many of those involved in “progressive” education seek to nurture qualities like self-confidence, problem-solving ability, initiative, and the ability to understand and intervene in organisations and society. Furthermore, they try to help each of their pupils to develop their idiosyncratic talents⁵. Since there are no good measures of such outcomes, most comparative evaluations utilise only traditional measures, mostly just of “the basics”, such as reading. Since the “progressive” teachers did not set out to produce higher reading scores (at least as conventionally measured), their pupils do no better on these tests than pupils who have studied in other programmes. Politicians take this as a signal to close the programmes. Worse, the destructive effects of “traditional” education do not show up. The failure of these studies to document pupils’ personal development (or deterioration) in a wide variety of *different* directions is a still more serious defect that there is not space to pursue here⁶. These problems could be ameliorated if the ITC *Standards* insisted that evaluations of both individuals and programmes be *comprehensive*. But, while such a move would be important, it would not be sufficient, because the way we have tried to “measure” individual differences is off beam. To see this, let us substitute the word “creative” for “genius” in Spearman’s claim. It would then read “Everyone is creative at something: The question is not ‘*How* creative are they?’ but ‘*At what* are they creative?’”. Think about it. Is someone who is highly creative at causing disruption in his or her classroom or work organisation likely to display that creativity if a psychologist gives him or her a box of wooden blocks and asks them to “be as creative as possible”?

In fact, creativity, thinking, initiating “experimental interactions with the environment” and learning from the effects of those actions, persisting, and so on are all difficult and demanding activities that people will not display unless they are engaged in activities that are of great concern to them⁷. It follows that these qualities cannot be meaningfully “measured” unless one has first identified the kind of activity the individual is strongly predisposed to undertake and then created a situation in which one can investigate which talents they bring to bear whilst undertaking activities they care about. Yet all of these talents, better termed components of competence or high-

level executive functions, are crucial to effective action. So, how to think about this situation? An analogy may help.

Dogs, hawks, and whales all need hearts, brains, eyes, lungs, and blood to function.

But it would not make sense to try to base our main framework for differentiating between animals on variance in their heartiness, braininess, or quality of their perceptual system. Nor would it make sense to rate all animals on scales 'measuring' dogginess, hawkishness, whaleiness, or snakeiness. What are the implications? The analogy suggests that we first need a branching *descriptive* classification, or framework, similar to that used in biology to help us identify the *kind* of person we are dealing with ... the kinds of things at which he or she is likely to be a genius (putting people at ease, creating political turbulence, pursuing adventurous research, etc.). And then we need to determine *which* components of competence ('intuitively' grasping the situation, initiating 'experimental interactions with the environment', learning from the effects of those actions, enlisting the help of other people, persisting etc) the individual brings to bear to undertake his or her 'chosen' activities. (Perhaps, in a second stage, one might assess how good they are at doing each of these things in the context of their chosen activity). A subset⁸ of the transformative processes that occur in some homes, schools, workplaces and adult developmental activities would then be understood as arising mainly from people finding themselves in environments that tap and harness their motives and lead them to utilise, develop, and display high level components of competence. When this analogy is pursued, it becomes clear that the way we have sought to model and study the interactions between people and their environments has also been way off beam. For what is required is some kind of ecological mapping of the multiple feedback loops and interactions between people and their environments.

To underline the points that have been made in this brief article let us ask: "Where would biologists have got to if they had sought to summarise the variance between animals in terms of 1, 2, 5, or 16 "variables", the variance in their environments in terms of 10, and the interactions between the two sets of variables as a series of multiple regression weights?"

The problems hinted at above will be discussed in a symposium entitled *Serious Errors in the Evaluation of Individuals and Programmes arising from the use of tests yielding Arbitrary Metrics and from the deployment of Arbitrary selections of Measures*, at the ITC conference in Liverpool and in a 'Virtual

Lab Meeting' on *Progressing a Paradigm Shift in Psychometrics*, to which readers are encouraged to contribute (on PsychWiki) http://www.psychwiki.com/wiki/Progressing_a_Paradigm_Shift_in_Psychometrics

Notes

1. See eg Raven (1994) and Raven, J., & Stephenson, J. (Eds.). (2001)
2. Raven, J. (2008)
3. Stephenson, J. (2001), Kazdin, A. (2006)
4. Raven, J. (1991)
5. Raven, J. (1994)
6. See Notes 3 and 8.
7. Raven, J., & Stephenson, J. (Eds.). (2001).
8. There is more that needs to be said about the seriously misleading unethical errors that have been made in the evaluation of transformative programmes in adult education, drugs based health care, and psychotherapy especially when these are presented as contributing to 'evidence based treatment' and 'payment by results' (see *Psychwiki*).

References

- Raven, J. (1991). *The Tragic Illusion: Educational Testing*. New York: Trillium Press. www.rfwp.com
- Raven, J. (1994). *Managing Education for Effective Schooling: The Most Important Problem is to Come to Terms with Values*. Unionville, New York: Trillium Press. www.rfwp.com
- Raven, J. (2008). *Intelligence, engineered invisibility, and the destruction of life on earth*. In J. Raven, & J. Raven, (Eds.), *Uses and Abuses of Intelligence: Studies Advancing Spearman and Raven's Quest for Non-Arbitrary Metrics*. Unionville, New York: Royal Fireworks Press; Edinburgh, Scotland: Competency Motivation Project; Budapest, Hungary: EDGE 2000; Cluj Napoca, Romania: Romanian Psychological Testing Services SRL.
- Raven, J., & Stephenson, J. (Eds.). (2001). *Competence in the Learning Society*. New York: Peter Lang.
- Stephenson, J. (2001). Inputs and outcomes: The experience of independent study at NELP. Chapter 21 in J. Raven & J. Stephenson (Eds.), *Competence in the Learning Society*. New York: Peter Lang.
- Kazdin, A. E. (2006). Arbitrary metrics: Implications for identifying evidence-based treatments. *American Psychologist*, 61, 42-49.

**Get the Guidelines on
Internet and CBT Testing
www.intestcom.org/**

On the need to secure psychological test materials

Jacques Grégoire

President of the Internal Test Commission

Thomas Oakland

ITC Council member and

President-elect of the IAAP Division of Psychological Assessment and Evaluation

Psychologists play a leading role in developing and using psychological tests that serve the public and professions. Psychologists also are committed to maintaining the integrity and security of test materials and other assessment methods, knowing that their unauthorized release to the public jeopardizes test integrity, results in test use by unqualified persons, and thus harms the public. Many professional associations are committed to this principle as reflected in their ethics codes. For example, the International Test Commission *Guidelines for Test Use* reflect this widely accepted commitment to maintain test security:

Standard 1.4.3. Protect the integrity of the test by not coaching individuals on actual test materials or other practice materials that might unfairly influence their test performance.

Standard 1.4.4. Ensure that test techniques are not described publicly in such a way that their usefulness is impaired.

In contrast to the need to maintain test security, some psychologists are selling tests through unauthorised sources to the general public. For example, the International Test Commission learned recently that tests are being sold on eBay. Such sales jeopardize test integrity, harm the public, and violate accepted practice.

The International Test Commission urges psychologists to become aware of this possible practice in their countries and to take steps to stop such unauthorized sales. We encourage national psychological associations to inform their members of this problem and to take preventative measures, including the revision of their ethics codes to help prevent this and similar unauthorised releases of tests to others. National psychological associations also are encouraged to develop standards that promote the safe disposal of outdated tests. We also urge eBay and other companies to establish and maintain standards that prevent the unauthorised sale of tests and other professionally protected materials.

~ ~ ~

John Keats Death

Prof. Emeritus John Keats died on New Year's Day. He was 86. John Keats was appointed Foundation Professor of Psychology at the University of Newcastle (Australia) in 1965, a position he held until his retirement in 1986. He served the International Test Commission for several years. He was President from 1995 to 1998. During his presidency, Professor Keats worked to promote ITC and to strengthen the links with the International Union of Psychological Science. He was also concerned to associate the developing countries, and especially the Asian countries to ITC. At the beginning of the nineties, he actively supported the participation of China, perceiving the high potential of this country. Professor Keats was a distinguish psychometrician. He published with Frederick Lord an influential article ("A theoretical distribution of mental test scores"), often cited, and several other important papers. Until recently, he continued to be very active, developing an ordinal test theory with Norman Cliff. The officers of the ITC Council thank Professor Keats for all what he did for their organisation.

~ ~ ~

Anne Anastasi @ 100: Her legacy for psychometrics

Harold Takooshian and Howard T. Everson

Fordham University

USA

The year 2008 marks the centenary of the birth of Anne Anastasi (1908-2001). If we date modern psychometrics from 1890, when Cattell introduced the term "mental test," Anastasi's diverse and prolific 71-year career spans well over half of this 118 years. Like the International Test Commission itself, Anastasi pioneered responsible cross-national testing in several ways. This brief salute reviews Anastasi's career, and its impacts on psychological testing world-wide. Anne's brilliant career began as a student, earning her BA with honors at Columbia in 1928, and her PhD in 1930 at age 21. Anne was a petit Italian-American woman whose long life of 91 years was entirely in a 12-mile radius within New York City. Yet her 71-year career from 1930 till 2001 was as diverse as it was long. She seemed to do everything, and did it with panache: a respected scientist, prolific author, dedicated teacher, no-nonsense administrator, much-sought consultant, leader of her discipline, and visionary architect of twenty-first century psychometrics (Hogan, 2003). She was ever forthright in her sometimes "dangerous ideas:" about the ethics and limitations of testing, the importance of



cultural factors, and her notion of mutable “developed abilities.” Of special relevance to ITC are her contributions to both cross-cultural and international assessment.

While still a student, Anne joined what has been called “the most impressive gathering of psychologists in the history of the discipline” (Hogan, 2000). At this Ninth International Congress of Psychology at Yale University on 1-7 September, 1929, young Anne at age 20 was one of 826 participants from 21 nations, rubbing shoulders for one week with such luminaries as Ivan Pavlov and Alexander Luria from Russia, Kurt Lewin and William Stern from Germany, Jean Piaget from Switzerland, Charles Spearman from U.K., and James McKeen Cattell from the USA. Her early research on the cognitive correlates of bilingualism soon segued into the first of her three major books, *Differential Psychology* (1937). With trademark thoroughness, clarity, and total command of her subject, Anne’s 868-page magnum opus literally created a new field blending quantitative psychology with anthropology and sociology. She succinctly defined differential psychology as “the scientific study of group differences,” and went on to offer 24 research-based chapters on group differences in ability and personality based on heredity, anatomy, age, education, family, gender, race, ethnicity, language, SES. She not only offers a panoramic review of this data on such group differences, but thoughtfully describes the methodological challenges with this data. While European Nazism was discrediting serious attempts to scientifically study group differences, Anne’s tome dismissed such efforts in a crisp sentence: “The array of evidence in support of this [Aryan supremacy] is incomplete and one-side at its best and fantastic and mythical at its worst” (Anastasi & Foley, 1949, p. 690). Still, this post-fascist stigma propelled the liberal-minded Anne to segue away from group differences to the less controversial and more practical focus on individual difference, with the debut of her classic *Psychological testing* in 1954. Across its seven editions, this was THE classic on testing for half a century—clear, even-handed, thorough. Anne was 87 when she co-authored the 7th edition with her alumna Susana Urbina in 1996.

Anne’s *Psychological Testing* had long been officially translated into most major languages for use on every continent as the primary reference on psychometrics. This includes even the most unlikely languages like Russian (where the CPSU had outlawed bourgeois “testy” in 1936) and Pharsi (where its translator was reportedly executed). It is hard to imagine a psychometrician with greater impact on world psychology than Anne, through *Psychological Testing* and her related writings. How odd that things come full-circle: When Peter Merenda (2005) delivered the annual Anastasi Lecture, he outlined a crisis in U.S. psychometrics, that in 2004 only 0.7%

or 22 of 3,200 psychology doctorates were in psychometrics. Indeed, Fordham is one of the few universities to maintain a doctoral-level program in psychometrics, with a large percentage of its psychometrics students drawn from around the globe to study at “Anastasi U.” In 2008, teams of overseas educators (starting with Ukraine) have begun coming to Fordham, with the possibility of government-funded Fordham training of indigenous psychology students to return and establish evidence-based educational selection programs in their own nation, a bold dream worthy of Anne Anastasi herself.

Largely due to *Psychological Testing*, Anne was revered by psychologists world-wide. Throughout the second half of the 20th century, a quiet stream of psychologists from around the world made their pilgrimage to Bronx, New York to seek a personal audience with Dr. Anastasi. In one case in 1988, B.F. Lomov headed a team of 8 psychologists from the Psychological Institute of the Academy of Sciences of Russia, who adoringly surrounded the diminutive Anastasi at Fordham for four hours, tapping her extensive knowledge of psychometrics (picture below). Besides her prolific writing of over 200 publications, of course there was also Anne’s work as a gifted teacher, deft administrator, award-winning consultant, and passionate architect of scientifically-based, culturally-sensitive educational testing policy.

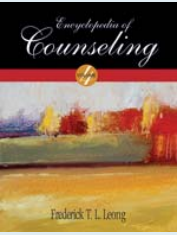


References

- Anastasi, A. (1937). *Differential psychology*. New York: Macmillan. [2/e 1949 with JP Foley, 3/e 1958]
- Anastasi, A. (1954). *Psychological testing*. New York: Macmillan. [7/e 1996 with SP Urbina]
- Anastasi, A. (1964). *Fields of applied psychology*. New York: McGraw Hill. [2/e 1979]
- Hogan, J.D. (2000, Winter). The founding of Psi Chi at the Ninth International Congress of Psychology. *Eye on Psi Chi*, 4, 11-13.
- Hogan, J.D. (2003, Winter). Anne Anastasi (1908-2001). *American Journal of Psychology*, 116, 649-653.
- Merenda, P.F. (2005, Dec 1). Psychometrics in the 21st Century. Invited presentation to the 2005 Anne Anastasi Forum, Fordham University, Bronx NY.

~ ~ ~

New for 2008 Encyclopedia



Sage Publications has recently released the 4-volume Encyclopedia of Counseling. Frederick Leong, who is Professor of Psychology and Director of the Centre for Multicultural Psychology Research at Michigan State University in East Lansing, Michigan, USA is the Editor-in-Chief of the Encyclopaedia. Leong is also a member of the Executive Council of the International Test Commission and a recent recipient of the American Psychological Association's Award for Distinguished Contributions to the International Advancement of Psychology. The *Encyclopedia of Counseling* provides a comprehensive overview of the theories, models, techniques, and challenges involved in professional counseling. The *Encyclopedia of Counseling* has approximately 600 entries and over 1,800 pages. This definitive resource covers all of the major theories, approaches, and contemporary issues in counseling. The four volumes of this Encyclopedia are flexibly designed so they can be used together as a set or separately by volume, depending on the need of the user. Each volume covers a major focus of counseling: (a) Volume 1 on the Changes and Challenges Facing Counseling in the 21st Century, (b) Volume 2 covers Personal and Emotional Counseling, (c) Volume 3 deals with Cross-Cultural Counseling, and (d) Volume 4 covers Career Counseling. Key themes covered in the Encyclopedia include Assessment, Testing, and Research Methods; Physical and Mental Health; Human Development and Life Transitions; and Therapies, Techniques and Interventions.

7th ITC Conference, July 19-21, 2010 Hong Kong

The next ITC Conference, will be held in Hong Kong from 19-21 July 2010, with pre-conference workshops on 18 July 2010. The conference will be the 7th ITC conference in a line of very successful conventions, all of which have been at the cutting-edge of the field of psychological and educational testing. The 7th ITC conference is a historic event. For the very first time an ITC conference will be held in a non-Western country, evidencing the global significance of the field of psychological and educational testing. The conference will provide opportunity for a variety of themes, among which themes such as Testing across borders, Testing and policy issues, Professionalization and training in testing, and Testing standards. The conference will contain eminent keynote speakers and invited symposia organizers, an interesting scientific program, and a range of workshops.

The conference will be hosted by the Chinese University of Hong Kong in the English language, and will take place right after the 27th ICAP conference in Melbourne, Australia. Hong Kong is one of the safest cities in the world to visit, English is widely spoken, and travelling and accommodation are easy and comfortable. We invite you to attend ITC's 7th conference in this very dynamic part of the world, where the field of psychological and educational testing and assessment is moving forward rapidly.



Fanny Cheung
Chair 7th Organising
Committee
Marise Born
President-Elect ITC

**Would like to provide a
review of test development or issues in YOUR country?**

Submit for consideration to the Editor of Testing International

Jan Bogg
jbogg@liverpool.ac.uk