

TESTING INTERNATIONAL

Vol. 25, July, 2011 Editor: Jan Bogg



PRESIDENT
Prof. John Hattie
University of Auckland
New Zealand

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

SECRETARY
Prof. Dave Bartram
SHL Group, UK

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

PAST-PRESIDENT
Prof. Marise Born
Erasmus University Rotterdam
Netherlands

COUNCIL MEMBERS
Elected Members
Prof. Frederick Leong
Michigan State University, USA
Prof. Kurt Geisinger
Buros Center on Testing & University
of Nebraska, USA
Prof. Dragos Iliescu
SNSPA Bucharest, Romania
Aletta Odendaal
University of Johannesburg
South Africa
Dr. Solange Wechsler
Pontificia Universidade Catolica de
Campinas, Sao Paulo, Brazil
Co-Opted Members
Dr. David Foster, Kryterion, USA
Prof. Ron Hambleton, University of
Massachusetts, USA
Observers
Prof. Marise Born, Erasmus University
Rotterdam, Netherlands
Prof. Paula Elosua, Universidad del País
Vasco, San Sebastián, Spain

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

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

Testing International
Dr. Jan Bogg
University of Liverpool, UK

CONTENTS

PRESIDENTS LETTER

2

ARTICLES

ITC International Guidelines Quality Control in Scoring, Test Analysis
and Reporting of Test Scores

Avi Allalouf

2

Past ITC Scholars Update

3

The impact of ITC Guidelines on Test User for test development: the Bra-
zilian experience

Solange Muglia Wechsler & Maria Angela Lourençoni, Brazil

6

Addressing the Issue of Non-normality in SEM: A comparative Overview
of LISREL, EQS, AMOS, and Mplus

Barbara M. Byrne

8

NEWS & CONFERENCES

10

PRESIDENTS LETTER

Dear ITC members and friends,

This past year has been exciting -- after 12 years in New Zealand I moved in April to a new position at the University of Melbourne. This involves becoming Director of the Melbourne Education Research Institute, succeeding Prof Barry McGaw. There are so many exciting new challenges, new colleagues, and new opportunities. We have left our young adult sons in NZ, brought the dogs, in the process of changing cricket allegiances, and starting anew in an exciting vibrant city (and I know many ITC members came to Melbourne last year for the International Congress of Psychology and experienced the city). The measurement issues have a degree of similarity across the two countries but sufficient differences. I oversaw the development of the NZ assessment and reporting engine now available in all schools that places much emphasis on helping teachers interpret who they had success with, about what, and how efficiently. Australia is more concerned with national testing of all students to drive accountability regimes. So there is much to do to turn the Australian thinking on its head.

The team, led by my new colleague Patrick Griffin, is undertaking exciting work in the development of collaborative assessment, and in new ways of thinking and doing testing in the internet space. The title of his work, with so many IJT members throughout the world is the "Assessment and Teaching of 21st. Century Skills". The major outcome, I believe, will be the permission the findings of this project will give us to think differently about assessment. I know you will learn much from this study -- and Patrick and his team hopefully will be presenting at our next year's conference in Amsterdam.

The work of ITC continues. Since its meeting in Hong Kong, the ITC has been undergoing some major internal organizational changes to make our administration more efficient and supportive for our members. By the time we meet in Turkey we will have completely changed the way we manage our records, do invoicing and distribute information, through the development of a web-based membership database. Apologies to those of you for whom this change may have caused some disruption in the first quarter of this year over issuing of invoices for membership dues. We do believe we have all the bugs out of the new system, and it should make life a lot simpler for both you and us in the future.

However, we have not just been fixing the

admin. The new Standards for Quality Control in scoring, test analysis, and reporting of test scores, chaired by Avi Allalouf are about to be released for public consultation, the two next conferences (Amsterdam, 2012, and San Sebastian 2014) are well advanced in planning, the IJT journal and Testing International are going from strength to strength, and the upcoming annual meeting of the ITC Council in Turkey promises a rich and full agenda. In addition, we have been collaborating with the EFPA Standing Committee on Tests and Testing to provide a full stream of papers and symposia on tests and testing at the Istanbul ECP in July.

I will be reporting to you about the outcomes of the forthcoming ITC Council meeting in the next Newsletter.

John
John Hattie
ITC President, 2010-2012

~~~

### **ITC International Guidelines Quality Control in Scoring, Test Analysis and Reporting of Test Scores**

**Avi Allalouf  
National Institute for Testing & Evaluation  
Jerusalem**

ITC is dedicated to developing guidelines for use in the various realms of testing and assessment. To date, three guidelines have been prepared: ITC Guidelines on Adapting Tests, ITC Guidelines on Test Use, and ITC Guidelines on Computer-Based and Internet-Delivered Testing. A Guideline for Test Security is in preparation. Tests are used for many purposes and sometimes serve as the basis for important decisions, and yet it cannot be taken for granted that tests -- or the testing process -- are of a high quality. Mistakes, such as the computing or reporting of an incorrect score may have a number of disturbing implications in any context of measurement. There are so many steps involved and so many stages at which mistakes can occur, that the only way to achieve a low rate of error is by maintaining extremely high standards at all times and applying adequate QC procedures. The Guidelines we have developed include general points that should be considered and agreed upon prior to scoring, test analysis and reporting of test scores, and also contain chronological and working guidelines -- i.e., step-by-step detailed QC instructions. The Guidelines relate to a wide variety of techniques and are suitable for use in a number of different applications and assessment situations, namely, educational, clinical,

and occupational testing; the main focus is on large-scale high-stake educational tests.

### Historical background and current status

Since 2008, I have been engaged in the development of the QC Guidelines with the superb assistance of Marise Born. Three preliminary versions of the Guidelines were reviewed by a group of testing and assessment experts in the educational, clinical and work domain: *Alvaro Arce-Ferrer, James Austin, Jo-Anne Baird, Giulia Balboni, Helen Baron, Dave Bartram, James Butcher, Janet Carlson, Iain Coyne, Kurt Geisinger, Ron Hambleton, John Hattie, Jason Lamprinou, Fred Leong, Tom Oakland, Fred Oswald and Christopher Rush*. The reviews were very helpful and valuable suggestions were made which will be incorporated in the next versions of the Guidelines. The concept and purpose of the Guidelines were presented (*Allalouf, Baron, Born, Geisinger and Leong*) at the European Congress of Psychology in Oslo in July, 2009. In addition, a pre-conference workshop on the Guidelines was held at the 7<sup>th</sup> ITC conference in Hong Kong in July 2010. The final version was sent in May 2011 to Dragos Iliescu, chairperson of the ITC Subcommittee on Research and Guidelines. It is expected that the ITC Council will discuss how to release and distribute the QC Guidelines to the public in due course.

### LATEST ISSUE

#### International Journal of Testing, Volume 11, Issue 2, 2011

The Examining Evidence for the Validity of PISA Learning Strategy Scales Based on Student Response  
*Therese N. Hopfenbeck and Andrew Maul*

Teachers' Perceptions of Large-Scale Assessment Programs Within Low-Stakes Accountability Frameworks  
*Don A. Klinger and W. Todd Rogers*

A Cognitive Diagnostic Modeling of Attribute Mastery in Massachusetts, Minnesota, and the U.S. National Sample Using the TIMSS 2007  
*Young-Sun Lee, Yoon Soo Park, and Didem Taylan*

Measurement Invariance of Scores from the Inventory of School Motivation Across Chinese and US College Students  
*Laura L. B. Barnes and Lihua Xu*

## Articles

### Past ITC Scholars Update

#### Tatjana Turilova-Miščenko - Latvia



My name is Tatjana Turilova-Miščenko. I have obtained my master degree in Educational Psychology at the University of Latvia in 2007. During the study, I took a test construction and adaptation course read by M. Raščevska. That course prompted me to choose the development of the "Classmates' Friendship Relationships Questionnaire" as the main task of my master thesis. It became the subject of my first article published in the Baltic Journal of Psychology in 2008. I continued to improve test construction skills during the doctoral studies supervised by prof. M. Raščevska in the University of Latvia.

In 2010 I became the ITC sponsored scholar. The opportunity to participate in the ITC Conference in Hong Kong was very important for my professional development in the field of test construction and adaptation. Besides all, this new experience helped me to improve my doctoral thesis "Bilingual adolescents' verbal comprehension and working memory". Within thesis three new tests were developed: Scientific Concepts Test, Vocabulary Test and Syllable-Word Working Memory Test. Each test has Latvian and Russian version that were simultaneously developed and have similar psychometric properties. An article about psychometric properties of the Scientific Concepts Test Latvian and Russian version is in progress. The same test will be presented on the 12<sup>th</sup> European Congress of Psychology which will be held in Istanbul, Turkey from July 4-8, 2011. In June this test will be presented on the Latvian Clinical Association meeting. The Syllable-Word Working Memory Test Latvian version will be presented on the 11<sup>th</sup> European Conference on Psychological Assessment which will be held in Riga, Latvia from August 31 till September 3, 2011.

During the last year I have been participated in a variety of other activities, i.e. in the Nordic-Baltic Doctoral network activities. I took part in the 1<sup>st</sup> United Latvian Psychologist Congress and the Scientific Conference of University of Latvia.

I still work as a school psychologist and I plan to get a doctoral degree in the nearest future. My research interests are not limited to verbal comprehension and working memory, but cover all cognitive abilities, especially in the context of education. I plan to improve my skills in the area of test construction and adaptation and use them to develop this field in Latvia. I would also like to get an experience in neuropsychology.

## **Natalija Curkovic - Croatia**

I am Natalija Curkovic from Croatia, proud ITC sponsored scholar in 2010. Since 2007 I have worked for the Croatian government agency for educational assessment (National Centre for External Evaluation of Education) in the Research and Development Department. Also, I am an ABD student in Educational Psychology with concentration in Psychometrics at University of Zagreb, Croatia.



ITC conference in Hong Kong was the starting point of an exciting period in my life which has been focused on my education to become the first Croatian expert for application of Item Response Theory (IRT) in developing and analyzing standardized knowledge tests.

My interest in IRT began during the ITC conference held in Liverpool in 2008. The conference was also the beginning of my cooperation and friendship with Professor Ronald K. Hambleton, a member of ITC Council who recommended me for the ITC scholarship. During his visit to Croatia in 2009, Professor Hambleton recognized the quality of my work in the field of test development and psychometric analyses. He invited me to come to the United States to further my knowledge at the prestigious Research and Evaluation Methods Program at the University of Massachusetts Amherst where I have been a visiting scholar since September, 2010. During the conference in Hong Kong, I had a great opportunity to meet many honorable ITC members who spent some period of time in Amherst learning from Professor Hambleton and his colleagues who work in the Center for Educational Assessment. They were very generous in sharing their experiences and giving me professional and personal advice for my stay in Amherst.

This last academic year has been extremely useful to my development as a testing specialist. I was privileged to attend Professor Hambleton's two-semester Item Response Theory course, Professor Stephen Sireci's Scaling class as well as Multivariate Statistics class held by Professor Lisa Keller. Professor Craig Wells taught me about structural equation modeling techniques. All the classes were organized in a manner to equip students with strong theoretical knowledge and practical application of theory. Training for use of relevant software packages was provided as well. Furthermore, I had opportunities to participate in some of the projects conducted by the Center for Educational Assessment.

Besides the knowledge that they taught me, Professor Hambleton and his colleagues have also been dedicated to share the values and professional standards proclaimed by ITC.

Recently, I have been working on my doc-

toral dissertation. The subject of my dissertation is a validation of taxonomies of knowledge used in test development. To get the answers to my research questions, I use mixed methodology approach: sophisticated psychometric tools, and cognitive interviews. Both professors and students from the University of Massachusetts have been helpful for fruitful discussions regarding my research.

I have met many important and successful people in the field of educational testing. Every time I introduced myself as an ITC scholar, they expressed respect and emphasized the importance of ITC's goals and work. When I go back home, I will try to implement my knowledge and professional standards that I acquired working with ITC people in order to reinforce the growing system of educational assessment in Croatia.

At the end, I want to express my gratefulness to the whole ITC community and especially to the ITC Council who recognized my work and supported me in my intention to learn and help my country to meet ITC's standards in testing.

## **Happy Zulu – Zambia**

From the time I attended the International Test Commission (ITC) Conference in Hong Kong last year July, I had been basically working on my Master's dissertation which I submitted two months to the board of examiners. The programme I am pursuing is Master of Science in Clinical Neuropsychology at the University of Zambia and when my paper has passed through both the internal and external examiners, I will graduate this year in August. The topic for this study is 'Performance of urban and rural adult populations on neuropsychological Tests in Zambia.'



Other than research and the writing up of my dissertation, I have been actively involved as a part-time lecturer at the University of Zambia, School of Education, Department of Adult Education and Extension Studies. I teach Psychology of Social Work to students under the University of Zambia pursuing Diploma in Social Work.

I have had an opportunity to make both the oral and poster presentations at the University of Zambia on the topic I presented at last year's ITC. I had an opportunity to share with my colleagues here in Zambia about ITC especially with regards to the ITC guidelines with regards to educational and psychological tests. I have exchanged also a few emails with my fellow past sponsored ITC scholars.

My interest in research is to explore more in the field of tests especially in the area of Neuropsychology. I would also like to study some challenges that this field of testing is faced with especially here in Zambia. This is more so regarding the cultural,

application and ethical issues which most researchers face in this field.

### **Fraide A. Ganotice Philippines**

My attendance to the 7th Conference of the International Test Commission on July 19-21, 2010 in the Chinese University of Hong Kong has provided me with the necessary exposure which truly enriched my research interests as neophyte researcher. Specifically, I would like to believe that because of this exposure which ITC has accorded me, I now included cross-cultural translation and validation in my current research interests. This line of research has been emphasized by Robert Roe in his presentation entitled "Testing for travelers: Past and future in one of the special symposia in the 7<sup>th</sup> ITC Conference.

Specifically, my study entitled "We're classmates, can we be friends?": Translation and Validation of the Filipino Version of Classmates' Friendship Questionnaire (CFQ) in the Philippines" was an outcome of my collaboration with Ms Tatyana Turilova-Miscenko, a co-recipient of young scholarship travel award of ITC, who also attended the 7<sup>th</sup> ITC Conference. Primarily, the primary objective of the said study was to describe the translation process done in the original English version of Classroom Friendship Questionnaire (Miscenko & Rascevska, 2008) and determine the psychometric acceptability of the newly developed Filipino version (CFQ-F) of the scale. The CFQ-F with four dimensions: social contacts, trust, support and cooperation, and lack of hostility, and trust, was completed by a sample of 466 high school students. Both within-network and between-network approaches to construct validation were adopted in the study. Results of statistical analyses performed suggest that the instrument has a good internal consistency and support is provided for its construct validity in terms of its factorial structure and correlations with other external variables. Using confirmatory factor analysis, I found strong evidence for the four-factor structure of CFQ-F. The psychometric characteristics of this scale justify its usefulness in future research involving Filipino participants. Implications for cross-cultural research are discussed. This paper has undergone the process of a paper review and will be published in the Educational Measurement and Evaluation Review (EMER), Second Edition.

Furthermore, together with Dennis McInerney and Ronnel B. King, I completed a study entitled "Exploring the Cross-Cultural Validity of the Sense of Self Scale (SoS) in Hong Kong and the Philippines" which will be presented by Dennis McInerney in the 6<sup>th</sup> Self Biennial International Conference in Quebec Canada. Earlier, we also worked on the cross-cultural validation of Inventory



of School Motivation which was already submitted for publication in an international journal.

In the future, I would like to be exposed more on the current trends in psychometrics and testing in general. I hope to attend the 8<sup>th</sup> Conference of the ITC in the University of Amsterdam in the Netherlands on July 3-5, 2012. I believe that this gathering of scholars will be another opportunity for advancement.

### **Zhongquan Li - China**

I am now working as an assistant professor at Department of Psychology, School of Social and Behavioral Sciences, Nanjing University, P. R. China. It is my great honor to be chosen as a sponsored young scholar to the 7<sup>th</sup> Conference of the International Test Commission 2010 held in Hong Kong. I benefit a lot from the opportunity of meeting prominent scholars in psychological and educational testing around the world, especially the chance of talking with members of the ITC Council. I really appreciate for their time and insightful suggestions. Here I am very glad to share my work and current research with you. Hope to possess the possibility of having research collaboration of joint publication with our ITC colleagues.

I am teaching three courses this semester for undergraduate students, Psychological Testing, Personnel Assessment, and Statistics and SPSS for Psychology. Within the classroom, I utilize the Powerpoint presentation, video materials and some interaction exercises to enhance my teaching. I encourage students to read classical and recent published articles relevant to course topics, and to make presentation in front of the class. I also like discussing on all sorts of topics with the students. I enjoy teaching at Nanjing University, and students also give high rating to my teaching.

My primary research interests focus on test security and privacy concerns aiming to make psychological tests fairer to test-takers. This line of studies is a continuance on my previous dissertation work. The research questions I am exploring are how test exposure can be partly dealt with by automatic item generation, how cheaters can be captured with identification of unusual response, how tests and assessment procedure are better designed to reduce distortions in employment context. Several papers on these topics have been published in Chinese peer reviewed Journals.

I am also interested in the role of emotion in decision making, especially financial and social decision making. Decision making plays a key role in our daily life. People with various cultural backgrounds and self concepts may have different expectations, norms and values under different emotion states, which in turn have the potential to influence their decision making as well as their subse-



quent behaviors. Recently, We have conducted a serial study on how subliminal primed emotions influence performances on various risky decision making tasks, and find some interesting interaction between emotion and decision making task types. Part of the results in this study will be presented on the 9th Biennial Conference of the Asian Association of Social Psychology this summer.

In addition, I am involved in two cooperative research projects, which fulfils my interest in having a better understanding of individual and organizational behaviors through more sophisticated and more appropriate research methods and statistical techniques (SEM, HLM, LCA et al.). One project focuses on investigating the structure, the dimensionality, and the developmental courses of Post-traumatic stress disorder (PTSD) among larges samples suffered from various trauma events, such as earthquake, violent attack and debris flow. Several articles have been published in *Personality and Individual Differences*, *Depression and Anxiety*, *Journal of Abnormal Child Psychology* and so on. The other attempts to study the career developmental trajectories of college students and which factors have impacts on the trajectories with a longitudinal design. Now we are at the stage of constructing a measure for career development.

In a word, thanks Jan and ITC newsletter for giving me the chance to express myself. All comments and suggestions on my work and research are welcome. I am looking forward to communicating and collaborating with you.

~~~

The impact of ITC Guidelines on Test User for test development: the Brazilian experience

Solange Muglia Wechsler
Pontifical Catholic University of Campinas

Maria Angela Lourençoni
Faculdade Max Planck

Introduction

Psychological tests in Brazil remained almost unchanged during the prior 30-year period. Tests typically were acquired from other countries, mainly from the United States, and translated into Portuguese, often with no studies related to their adaptation, validation, and norming in Brazil. As a result, the tests were criticized and discredited for being irrelevant for use in Brazil, including for the diagnosis of psychological disorders. This situation changed drastically through the efforts of Brazilian scholars from various universities throughout the country.

A number of scholars created laboratories for test development at their local universities. Furthermore, they saw the need to create a national association dedicated to the promotion of sound methods for developing, evaluating, and using

tests. This vision led to the creation of the Brazilian Institute of Psychological Assessment (IBAP). Its mission to promote research and development that promotes sound methods for adapting, developing, evaluating, and using test in Brazil.

The promotion of better practices on test development increased considerably following the first IBAP national convention on test development and use in Campinas (São Paulo), held in 2003 and attended by more than 1,000 psychologists, scholars, other professionals, and students. At this time, the Federal Council of Psychologists assumed leadership for addressing severe problems that characterized psychological assessment.

The *International Test Commission's Guidelines on Test Use* was essential to form the basis of the new regulation designed to promote higher standards. For example, Brazilian psychologists were restricted to using tests that were evaluated and found to demonstrate suitable validity and reliability and in other ways to advance assessment practices. An expert commission composed by scholars in assessment was organized to carry out this evaluation process. A commission was organized to carry out this evaluation process. The first test evaluations reported by the expert's commission indicated 50% of the available published tests had to be removed from the market because they did not provide sufficient information on the test's validity, reliability, and currency of norms.

Psychologists now see the value of these regulations. Many new and better tests have been developed. Their manuals provide more complete information on each test's constructions/adaptation process, validity, reliability, and norms. This has led to greater respect for tests by those who administer and use test results. Graduate programs focus more attention on test development and use. Four national conventions sponsored by the Institute of Psychological Assessment (IBAP) have been held so far. The fifth conference is to occur at the end of this month.

Method

In order to evaluate the impact of the new regulations on test movement, an investigation was conducted to analyze the abstracts presented at the third (2007) and fourth (2009) national conferences of the Brazilian Institute of Psychological Assessment (IBAP). A total of 351 abstracts from 2007 and 537 from 2009 were available from IBAP's webpage (www.ipabnet.org.br). The abstracts came from research presented at round tables, symposia, oral presentation or posters. The following categories were defined for this analysis: a) populations; b) themes; c) objectives ; d) areas.

Results

Considering the amount of studies presented and the specific populations investigated, it can be observed there was a significant increase on tests designed for teenagers and adults (12-60

years old). Young people (19-30) continue to be the preferred sample (Figure 1).

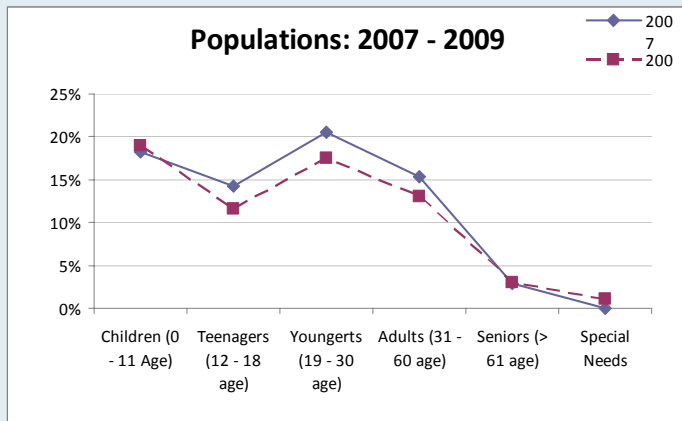


Figure 1- Abstracts comparison from 2007-2009 national conference on tests

Regarding themes, test validation was the highest in both periods, cognition, mental health and social skills were also popular at both conventions (Figure 2).

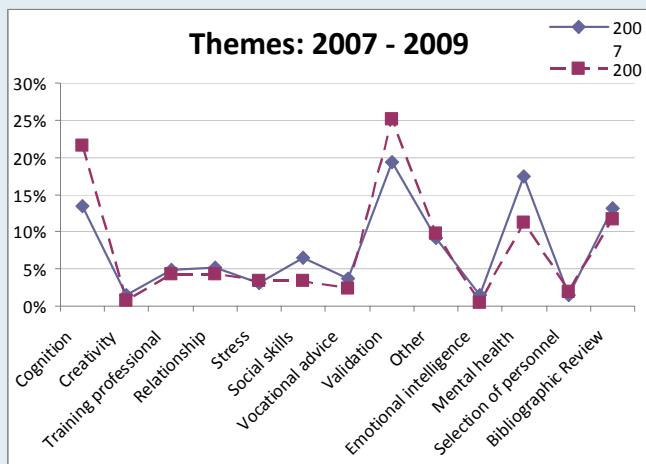


Figure 2- Most researched themes in 2007-2009 national conferences on tests

The study objectives most mentioned on the abstracts had high similarity between test validation and practical use. Validation concerns were much higher at the 2009 convention. Practical use and diagnosis also was higher in 2009. On the other hand, there was a decay of interest on test adaptation, reliability and norms (Figure 3).

When comparing areas of study, clinical psychology is the one with the highest interest, followed by educational and organizational psychology. A variety of areas will probably be developed for the next convention (Figure 4).

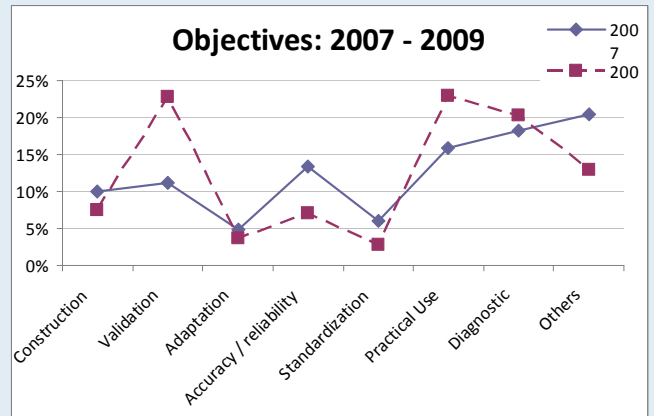


Figure 3- Research objectives presented in 2007-2009 conference abstracts

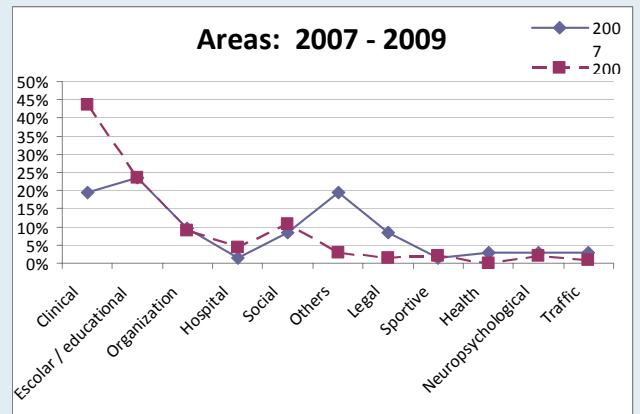


Figure 4- Areas comparison between 2007-2009 conference abstracts

Conclusion

The growth of test movement in Brazil can be observed from the last two test conferences sponsored by the Brazilian Institute on Test Development (IBAP). Greater interest was detected regarding test validation, more so than test adaptation or norming. These results indicate that more tests are being constructed in the country, than just adapting them from foreign nations.

ITC Guidelines had a powerful impact on this movement, indicating its importance at the international level, in ensuring the quality of psychological instruments provided in different countries.

More information on the coming V National Conference on Psychological Assessment can be obtained from www.congressoibap.com.br

~ ~ ~

Addressing the Issue of Non-normality in SEM: A comparative Overview of LISREL, EQS, AMOS, and Mplus

**Barbara M. Byrne
University of Ottawa**

The growth of structural equation modeling (SEM) over the past 35 years or so has been nothing short of amazing! An interesting offshoot of this escalation has been the equally expanding, yet somewhat more subtle growth of computer software capable of handling the statistical rigors demanded by the SEM methodology. In combination, each component of this symbiotic pair serves as a stimulant for the other in advancing the practice of SEM methodology. As substantive researchers increasingly seek more comprehensive answers to their research questions, statistical researchers are challenged to further advance the capabilities of SEM methodology, which in turn, necessitates further development of existing SEM software programs. Since development of the first SEM program in 1974 (LISREL), the ensuing years have witnessed a steady increase in the advancement of alternative SEM computer software such that there are now several programs from which to choose. Of these, the following four programs are clearly the most widely used: AMOS (Arbuckle, 2009), EQS (Bentler, 2005), LISREL (Jöreskog & Sörbom, 2006), and Mplus (Muthén & Muthén, 2010).

By default, all SEM programs are based on maximum likelihood (ML) estimation. However, a critically important assumption in these analyses is that the data have a multivariate normal distribution in the population. Violation of this assumption can seriously invalidate statistical hypothesis-testing with the result that the normal theory test statistic (c^2) may not reflect an adequate evaluation of the model under study (Hu, Bentler, & Kano, 1992) thereby leading to results that may be seriously misleading. In practice, however, data are most often *not* multivariate normal. Thus, one of the most critical challenges faced by SEM program developers has been how to estimate parameters most appropriately when data are non-normally distributed. The purpose of this article is to provide a comparative overview of the approach taken by LISREL, EQS, AMOS, and Mplus in addressing this issue of non-normality. As restrictions of space necessarily curtail the length of this review, readers interested in a step-by-step walk-through of how these various approaches are applied in practice, are referred to Byrne (1998, 2006, 2009, in press, respectively). We turn now to a synopsis of options available in each of these programs for addressing the issue of non-normality; programs are presented here in the order of their original development.

LISREL (Version 8.8)

One approach to working with data that are non-normally distributed is use of asymptotic (large

sample) distribution-free (ADF) estimators for which normality assumptions are not required (see Browne, 1984). This is the approach embraced by the LISREL program. Implementation of this strategy, however, involves a two-step process. Whereas Step 1 uses the PRELIS companion package to recast the data into asymptotic matrix form, Step 2 focuses on analysis of this matrix based on the use of LISREL with weighted least squares (WLS) estimation.

One major limitation of this approach, however, has been its excessively demanding sample-size requirement. It is now well known that unless sample sizes are extremely large (1000 to 5,000 cases), the ADF estimator performs very poorly and can yield severely distorted estimated values and standard errors (Curran, West, & Finch, 1996; Hu et al., 1992; West, Finch, & Curran, 1995). More recently (see Raykov & Marcoulides, 2000), statistical research has suggested that unless sample sizes are greater than 10 times the number of estimated parameters, ADF estimates cannot be trusted.

EQS (Version 6.1)

In contrast to LISREL, the EQS approach in addressing non-normality is based on the contention that it may be more appropriate to correct the test statistic, rather than use a different method of estimation (Chou, Bentler, & Satorra, 1991; Hu et al., 1992). Satorra and Bentler (1988) developed such a statistic that incorporates a scaling correction for the c^2 statistic when distributional assumptions are violated; its computation takes into account the model, the estimation method, and the sample kurtosis values. The resulting (corrected) chi-square value, called the Satorra-Bentler chi-square ($S-Bc^2$) and standard errors are said to be "robust," meaning that their computed values are valid, despite violation of the normality assumption underlying the estimation method. The $S-Bc^2$ has been shown to be the most reliable test statistic for evaluating SEM models under various distributions and sample sizes (Hu et al., 1992; Curran et al., 1996). In addition, EQS computes robust versions of the key ad hoc indices of fit (CFI, RMSEA, and RMSEA 90% C.I.).

In addition to its usual application with continuous data, the $S-Bc^2$ can be used with non-normal categorical data. Although it basically treats the ordered data as if they are continuous, DiStefano (2002) has reported this scaled approach to be beneficial in yielding standard errors that are more precise than ML estimates for non-normally distributed data having as few as three ordered categories.

In SEM, data are often incomplete, as well as non-normally distributed. When this condition holds, correction based on the $S-Bc^2$ is not appropriate. Rather, analyses should be based on the Yuan-Bentler (2000) scaled statistic ($Y-Bc^2$) which corrects both the test statistics and standard errors

when the input file specifies the use of robust statistics and indicates the presence of missing data.

In addition to these scaled statistics, EQS has three distribution-free statistics based on the distribution of residuals; robust versions of these test statistics are automatically computed when this option is specified. The first of these is of a type developed by Browne (1984; but see earlier comments regarding sample size requirements). In contrast, the Yuan-Bentler residual-based statistic (Yuan & Bentler, 1998) can be used with smaller samples than the original residual-based statistic and does so without any loss of its large-sample properties (Bentler, 2005). Finally, the Yuan-Bentler residual-based F-statistic (Yuan & Bentler, 1998), represents a more extensive modification of Browne's (1984) statistic and is considered by Bentler (2005) to be the best available residual-based test at this time.

AMOS (Version 18.0)

One approach to the analysis of non-normal data in AMOS is to base analyses on ADF estimation (Browne 1984), which can be selected from the estimators offered on the *Estimation* tab of the *Analysis Properties* icon or drop-down *View* menu of AMOS Graphics. However, given the restrictions of sample size noted earlier, this option is typically of little use to most practical researchers.

Unfortunately, the EQS robust test statistics described earlier are not available in the AMOS program. Nonetheless, there remains at least one other viable approach to analyses of non-normal data based on a procedure known as "the bootstrap" (Efron, 1979; West et al., 1995). This approach is the one most commonly taken by AMOS users in addressing the issue of non-normality. The bootstrapping technique enables the researcher to compare the extent to which the ML estimates deviate across the total number bootstrapped samples. More specifically, the AMOS output from a bootstrapping analysis yields five columns of results as follows: (a) the bootstrap estimate of the standard error for each selected model parameter (e.g., factor loading), (b) the approximate standard error of the bootstrap standard error itself, (c) the mean parameter estimate computed across the bootstrapped samples, (d) the difference between the bootstrap mean estimate and the original estimate (i.e., bias), and (e) approximate standard error of the bias estimate. In addition, the program reports 90% bias-corrected confidence intervals. (For an example application of this bootstrapping approach, see Byrne, 2009.)

Mplus (Version 6.1)

As with the EQS program, treatment of non-normal data in *Mplus* is addressed via estimators that can yield corrected test statistics and standard errors. Likewise, these robust estimators vary according to measurement scale of the data as well as to whether the data are complete or incomplete.

For data with outcome variables that are continuous, non-normally distributed, and complete, the MLM estimator, which is reportedly equivalent to the S-Bc² statistic (Muthén & Muthén, 2010), is most appropriately used. Likewise, the MLMV estimator, although computationally more intensive than MLM, is similarly robust with the exception that the chi-square statistic is both mean and variance-adjusted. In the event that continuous data are both non-normal and incomplete, it is most appropriate to base analyses on the MLR estimator, which Muthén and Muthén (2010) posit is asymptotically equivalent to the Y-B c² used in EQS.

Finally, when data are both categorical and non-normally distributed, *Mplus* provides for the use of two robust weighted least squares (WLS) estimators, the WLSM and the WLSMV. Whereas the chi-square test statistic for the WLSM estimator is mean-adjusted, it is mean- and variance-adjusted for the WLSMV estimator. Importantly, these estimators are not appropriate for use with data that are incomplete and involve censored, unordered, or count dependent variables.

In summary, when SEM analyses are based on data that are *not* multivariate normal, the issue of non-normality must be addressed. Clearly, use of robust statistics that can take such non-normality into account provides the simplest approach in dealing with this difficulty. The EQS and *Mplus* programs are both designed to enable this approach. Although readers familiar with SEM methodology will likely be comfortable in using either of these programs, I believe the EQS program is best suited to those having little to no knowledge of SEM.

References

- Arbuckle, J.L. (2009). *AMOS18 User's Guide*. Chicago, IL: SPSS.
- Bentler, P.M. (2005). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.
- Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.
- Byrne, B.M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.
- Byrne, B.M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Erlbaum.
- Byrne, B.M. (2009). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York, NY: Taylor & Francis/Routledge.
- Byrne, B.M. (in press). *Structural equation modeling with Mplus: Basic concepts, applications and programming*. New York, NY: Taylor & Francis/Routledge.
- Chou, C.-P, Bentler, P.M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, 44, 347-357.
- Curran, P.J., West, S.G., & Finch, J.F. (1996). The robustness

of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 7, 16-29.

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis.

Structural Equation Modeling, 9, 327-346.

Efron, B. (1979). Bootstrap methods: Another look at the jack-knife. *Annals of Statistics*, 7, 1-26.

Hu, L-T., Bentler, P.M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112, 351-362.

Jöreskog, K.G. & Sörbom, D. (2006). *LISREL 8.8*. for Windows. Lincolnwood, IL: Scientific Software International.

Muthén, L.K. & Muthén, B.O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Authors.

Raykov, T. & Marcoulides, G.A. (2000). A first course in structural equation modeling. Mahwah, NJ: Erlbaum.

Satorra, A., & Bentler, P.M. (1988). Scaling corrections for chi square statistics in covariance structure analysis. *American Statistical Association 1988 Proceedings of the Business and Economic Sections* (pp. 308-313). Alexandria, VA: American Statistical Association.

Yuan, K.-H. & Bentler, P.M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, 51, 289-309.

Yuan, K.-H. & Bentler, P.M. (2000). Three likelihood-based methods for mean and covariance structure analysis with non-normal missing data. In M.P. Becker (Ed.), *Sociological Methodology 2000* (pp. 165-200). New York: Wiley-Blackwell.

Tests and Testing Track at the European Congress in Istanbul, July 2011

Dave Bartram

ITC Secretary and Convener of the EFPA Standing Committee on Tests and Testing

Following the success of the joint EFPA/ITC Tests and Testing track at the Oslo European congress in 2009, we have been organizing a repeat of this for the Istanbul Congress. Working with Prof Canan Summer, who is the Turkish Psychological Association's representative on the EFPA Standing Committee on Tests and Testing, I have been promoting and organizing this stream on behalf of both the EFPA Standing Committee and the ITC Council.

This year we have five symposia, six oral papers and a roundtable discussion session. The Roundtable is on *Test User Qualification in Europe*, and will provide a briefing on the developments being carried out by EFPA and a chance for audience discussion with those leading the EFPA project (Dave Bartram, Pat Lindley, Sverre Nielsen, Ana Hernandez, Lars Michaelsen, and Eva Bergvall). Individual papers are being presented by: Cigdem Is Guzel, Herrera Rojas, Katharina Lochner, Krunoslav Matesic and Olga Mitina. The symposia bring together some of the leading experts in the field from around the world. The symposia include:

The EFPA Test Review Model: Time for an update?

Convened by Dave Bartram, with contributions from Arne Evers, Carmen Hagemester, Jose Muniz, Sverre Nielsen and Patricia Lindley.

International Testing Practices.

Convened by Jose Muniz, with contributions from him, Ronald Hambleton, Arne Evers, Dragos Iliescu and Jens Egeland.

Validating Educational and Psychological Tests: Theory, Application and Future Directions.

Convened by Stephen Sireci, with contributions from Steve Sireci, Paula Elosua, David Foster, Giray Berberoglu and Kurt Geisinger.

From indigenous to cross-cultural personality assessment.

Convened by Fanny Cheung, with contributions from her, Marise Born, Frederick Leong and Dragos Iliescu.

Internationalization of Test Reviewing.

Convened by Dave Bartram, with contributions from Sverre Nielsen, Patricia Lindley, Kurt Geisinger and Arne Evers.

We hope this stream will become a permanent feature of future European Congresses and look forward to seeing you in Istanbul and welcoming you to the Tests & Testing sessions.

NEWS

Award for Council Member



At the recent annual meeting of the American Educational Research Association, Professor Ronald Hambleton received the Robert L. Linn Award for his outstanding contributions to measurement and educational policy, including his work in the areas of test adaptation research, international assessment, item response theory, score reporting, and setting performance standards. Professor Hambleton has been a member of the ITC for 29 years serving the ITC in a number of different roles.

Journal Call for Papers

The Brazilian National Institute of Psychological Assessment is calling for papers (in English) to publish in the Brazilian Journal of Psychological Assessment. Information from: wechsler@lexxa.com.br

EAPA Book Series Call for Proposals

The EAPA Book Series is an initiative of the European Association of Psychological Assessment (EAPA) and Hogrefe Publishing GmbH that aims at producing a series of volumes on state-of-the-art theory and research in psychological assessment in various fields of psychology. Each volume will focus on a topic and will host theory, research and applications. The book series is expected to be of interest to researchers, teachers and students of psychology as well as practitioners.

Experts in psychological assessment with a good publication record in the field and prior experience in editing at an international level are invited to submit proposals to the Editorial Board. Volume editors can come from Europe or other parts of the world. Each proposal is evaluated based on its merits and whether it fits to the overall planning of the series. For more information on the book series and guidelines to the authors, visit the EAPA website (www.eapa-homepage.org/bookseries). Book proposals should be directed to Tuulia Ortner, e-mail: tuulia.ortner@fu-berlin.de

2011

11th European Conference on Psychological Assessment (ECPA2011) of the European Association for Psychological Assessment (EAPA)

Riga, Latvia, August 31st to September 3rd, 2011
Conference website: <http://www.ecpa11.lu.lv/>

2012

International Conference on Health, Wellness and Society

Chicago, IL, USA, 10th-11th March 2012
Conference website: <http://health-conference.com>

The 8th Conference of the ITC: Modern Advances in Assessment: Testing and Digital Technology, Policies and Guidelines

Amsterdam, July 3rd-5th, 2012
Conference website: www.itc2012ams.com

30th International Congress of Psychology: Psychology Serving Humanity

Cape Town, SA, 22nd-27th July 2012
Conference website: www.icp2012.com

Contact Us

We now have the following email addresses:

ITC President President@intestcom.org
ITC Secretary Secretary@intestcom.org
ITC Treasurer treasurer@intestcom.org

Editors
International Journal of Testing
ijteditor@intestcom.org

Testing International tieditor@intestcom.org

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