



TESTING INTERNATIONAL

Editor: Anita M. Hubley

PRESIDENT

Prof. Bruce Bracken
College of William and Mary
School of Education
P.O. Box 8795
Williamsburg, VA
U.S.A. 23187-8795

PRESIDENT-ELECT

Prof. J. Muñiz
Universidad de Oviedo
Facultad de Psicología
Plaza de Feijoo, s/n
3300 Oviedo
Spain

SECRETARY

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

TREASURER

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

PAST-PRESIDENT

Prof. David Bartram
SHL Group PLC
The Pavilion
1, Atwell Place
Thames Ditton, KT7 0SR
U. K.

COUNCIL MEMBERS

Dr. M. Born
Erasmus University of Rotterdam,
The Netherlands
Dr. Iain Coyne
University of Hull, U.K.
Dr. David Foster
Galton Technologies Inc., U.S.A.
Prof. Cheryl Foxcroft
University of Port Elizabeth,
South Africa
Prof. Thomas Oakland
University of Florida, U.S.A.

REPRESENTATIVES

Prof. J. G. Adair, IUPsyS
University of Manitoba, Canada
Prof. R.K. Hambleton, IAAP
University of Massachusetts, U.S.A.

EDITORS

International Journal of Testing
Prof. Bruno D. Zumbo
Dept. of ECPS, 2125 Main Mall
University of British Columbia
Vancouver, B.C.
Canada V6T 1Z4

Testing International

Prof. Anita M. Hubley
Dept. of ECPS, 2125 Main Mall
University of British Columbia
Vancouver, B.C.
Canada V6T 1Z4

EDITOR'S NOTE

2

PRESIDENT'S LETTER

2

BRIEF ARTICLES

- A Paradigm Proposal for Personality Testing: A French Experience in Industrial Psychology
Daniel Matzkin & Claude Léonard (France) 4
- Using DIF in Test Adaptation for Cross-Lingual Assessment
Avi Allalouf (Israel) 6
- Quality Standards for E-Learning: Cross Validation Study of the Demand Driven Learning Model (DDLm)
Krista Breithaupt (U.S.A.) & Colla J. MacDonald (Canada) 8

REPORTS

- On-line Readings in Testing and Assessment (ORTA) Project: Progress Update
Cheryl Foxcroft (South Africa) & Marise Born (The Netherlands) 12
- ITC's International Guidelines on Computer-Based and Internet-Delivered Testing
Dave Bartram (U.K.) & Iain Coyne (U.K.) 13
- Portuguese News about Psychological Tests
Aura Montenegro (Portugal) 13

THE NOTICEBOARD

- Paul Meehl Passes Away 14

UPCOMING CONFERENCES ON TESTING AND ASSESSMENT

15

EDITOR'S NOTE

Anita M. Hubley
University of British Columbia
CANADA

Dear Friends,

I hope you will enjoy this issue which contains a number of articles from colleagues around the world. In our first article, Daniel Matzkin and Claude Léonard (France) discuss the often contentious issue of personality testing in personnel selection and performance assessment and describe the use of Léonard's (2002) IPSO, a French-Canadian personality test *battery* designed for use in personnel selection. Our second article, by Avi Allalouf (Israel), describes the use of differential item functioning (DIF) in test adaptation. Test translation is a relatively common activity in the testing field but not everyone is familiar with the role of differential item functioning (DIF) in this process. Finally, in our third article, Krista Breithaupt (U.S.A.) and Colla J. MacDonald (Canada) present a validation study that is a follow-up to their 2002 *International Journal of Testing (IJT)* article on an online survey designed to evaluate e-learning programs using the Demand-Driven Learning Model (DDLML).

Also in this issue are two reports in which the ITC Council requests your assistance. We are still seeking individuals willing to author manuscripts for the *On-line Readings in Testing and Assessment (ORTA) Project* and we are requesting feedback on the first draft of the ITC's *International Guidelines on Computer-Based and Internet-Delivered Testing*. We also have a report from Aura Montenegro (Portugal) on a Portuguese adaptation and validation of the Retirement Satisfaction Inventory (RSI) by A.M. Fonseca and C. Paúl (2002) that was published in *Revista Psychologica*. Finally, I present a brief biographical sketch and summary of a few of the many accomplishments of Professor Paul Meehl who passed away in February of this year. Professor Meehl was an influential scholar in the testing and assessment field and, in 2002, was identified as one of the 100 most eminent psychologists of the 20th century.

In his ITC *President's Letter*, Bruce Bracken challenges us to move toward a broader model of testing fairness. If you are working in this area of testing, please consider submitting a piece to *Testing International* for our next issue!

PRESIDENT'S LETTER

**Assessing a Population that Speaks More Than
200 Languages: Issues of Fairness in Testing**

Bruce Bracken
College of William and Mary
U.S.A.

In a fairly recent report, Pasko (1994) reported that more than 200 languages are spoken by students in the Chicago public schools. Chicago also is reputed to be the largest Polish-speaking city in the world after Warsaw and is second only to Athens as a Greek-speaking city. The title of this column intentionally insinuated that the population referred to was the world population, not a single city in the United States. As international cities go, however, Chicago is not unique. International cities around the world host similarly diverse populations, and testing specialists in every country face issues of equitable testing practices for linguistically and culturally diverse populations.

Psychologists have long sought to identify effective procedures for assessing individuals who lack the manifest language to demonstrate their latent abilities. Among the earliest of such attempts was that of the French clinician, Jean Itard, who was presented with the challenge of assessing and modifying the cognitive abilities of Victor, a feral youth who has since acquired the moniker "The Wild Boy of Aveyron" (Carrey, 1995; Itard, 1932). In addition to Itard's well-publicized efforts, other international clinicians are similarly noteworthy. Seguin (1907), for example, is acknowledged for having developed one of the first nonverbal tests of cognitive ability, the *Sequin Form Board*, to address this issue.

In addition to the linguistic diversity that has resulted from historic opportunity- and event-driven emigration and the more recent effects of globalization, it also should be recognized that considerable linguistic diversity has long existed among aboriginal inhabitants of all continents. For example, throughout the United States several surviving Native American nations and tribes, especially those that live in the Western U.S. (e.g., Navaho), Hawaii and Alaska, as well as some long-standing residents (e.g., African Americans) either speak different native languages or nonstandard forms of English (e.g., Cajun, Gullah, Pigeon-English).

What is the common denominator among these disparate linguistic and cultural groups? All of these individuals, for many different root reasons, may be seriously disadvantaged when they are assessed by traditional language-loaded psychological tests, regardless of the construct the tests purport to assess (Bracken & McCallum, 1998). Examinees' disadvantages in these instances lie in the linguistic and/or cultural demands that are placed on them; language and cultural demands that may be nonessential elements of the construct intended to be assessed. Any extraneous variable (e.g., linguistic or cultural demands of a test) that contributes to error variance is by definition "construct irrelevant." Such construct irrelevant variance contributes to test bias and may negatively affect the consequential validity of the test for groups of individuals from the same linguistic or cultural backgrounds (Bracken, 2003).

ITC 'Solutions'

The ITC has been actively involved in issues of testing fairness through a variety of venues and activities. However, as President of the ITC, I would like to see the agenda move even farther beyond a largely deficit-based *psychometric bias* detection approach (e.g., Differential Item Function; Differential Test Function) and toward a broader model of *testing fairness*. The ITC has been oriented in this direction to some degree in its past projects, publications, conferences, and ITC sponsored symposia (e.g., the Test Adaptation Conference held in Georgetown; the ITC Guidelines on Test Use; the Winchester Conference on Computer-Based Testing). I would like to see an even more direct and expansive treatment of the topic at the international level.

In this vein, the ITC Council will discuss plans for a future conference devoted to the topic of *Fairness in Testing* at its annual meeting in Vienna (July, 2003). Hosting a conference that would explore issues related to creating fair tests and employing fair testing practices in clinical, industrial, and educational settings was one of my five primary goals as ITC President. I am pleased to pursue this goal through a possible ITC conference. Consistent with this goal, I am scheduled to present a workshop on equitable assessment practices at the First National Conference on Psychological Assessment / VI International Conference on Psychological Assessment, July 23-26 in Campinas, Sao Paulo,

Brazil. This opportunity will allow me to combine my desire to discuss equitable assessment practices with my goal of increasing ITC membership and involvement throughout the Americas. I am excited about forwarding both of these important goals and activities, and envision the ITC possibly creating an important set of guidelines that comprehensively address issues related to fairness in testing from the perspectives of test authors, test users, and test takers.

References

- Bracken, B. A. (Mar., 2003). Fairness in testing: An investigative model. Presentation at the National Association of School Psychologists annual conference, Toronto, Canada.
- Bracken, B. A., & McCallum, R. S. (1998). Universal Nonverbal Intelligence Test. Itasca, IL: Riverside Publishing.
- Carrey, N. J. (1995). Itard's 1828 memoire on 'mutism caused by a lesion of the intellectual functions': A historical analysis. Journal of the American Academy of Child and Adolescent Psychiatry, 34(1), 655-1661.
- Itard, J. M. G. (1932). The wild boy of Aveyron. New York: Appleton-Century-Crofts.
- Pasko, J. R. (1994). Chicago - - Don't miss it. Communiqué, 23(4), 2.
- Seguin, E. (1907). Idiocy and its treatment by the physiological method. New York: Teachers College, Columbia University.

International Journal of Testing (IJT)

Official Journal of the ITC

IJT publishes original articles addressing empirical research, theoretical issues, and methodological approaches, as well as integrative and interdisciplinary reviews of testing related topics and reports of current testing practices in psychology, education, counseling, human resource management, and related disciplines.

Editor: Bruno D. Zumbo

Publisher: Lawrence Erlbaum Assoc.

<http://www.educ.ubc.ca/faculty/zumbo/ijt/>

BRIEF ARTICLES

A Paradigm Proposal for Personality Testing: A French Experience in Industrial Psychology

Daniel Matzkin

Ecole des Hautes Etudes en Sciences Sociales &
Editions et Applications Psychologiques-France
FRANCE

Claude Léonard

Editions et Applications Psychologiques-France
FRANCE

Introduction

The story of Phineas Gage – a railroad construction supervisor who experienced startling personality changes following a serious brain injury – made famous by Damasio (1994), is one of the multiple fascinating cases that can be seen in personality literature. Certainly, personality is one of psychology's most captivating issues. It has even become a topic encompassing different fields of research and interests: e.g., personality disorders, social and industrial psychology, neuroscience, anthropology, and political science.

Turning to personnel selection concerns, it is no secret that personality assessment has an ambiguous position. Yet, in an European country like France, personality tests are still one of the most widely used by organizations in selection procedures (Courrier Cadres, 2001).

On the one hand, some researchers and industrial psychologists conclude that personality testing has low or no validity for forecasting job performance and job selection (e.g., Bruchon-Schweitzer & Lievens, 1991; Guion & Gottier 1965, see Hogan, 1991 for a theoretical review on this subject). A sample of arguments given by those defending this position include: validity coefficients for personality measures are low; personality is not a stable behavior; label definitions given to personality dimensions resulting from factor analysis tend to vary from one study to another (i.e., labels are therefore meaningless concepts). Other critics stress that a number of personality tests are made of non contextual items which tend to bias personality measures and interpretations that are made out of them (e.g., Bing et al., 2002;

Schmit et al., 1995). Most psychologists agree upon this point.

On the other hand, empirical results seem to point out that personality assessment can be a major and successful component of a personnel selection system (Hogan, Hogan, & Roberts, 1996). For example, the five factor personality model – as well as others – has gained robust empirical considerations in job performance prediction (Schmidt & Hunter, 1998). However, it is clear that such a discrepancy concerning personality research is an uncomfortable situation for personality researchers, recruiters, examinees, and scientific psychology in general. This brings us to the purpose of this short article: since personality is central to psychological understanding, how can this disagreement be overcome? Our aim is to present a “new” paradigm for personality testing which has been experienced in France for some years, giving appreciable results in job forecast. As readers will learn, our proposal is all but revolutionary, it is simply based on psychologists' common sense.

What about Personality Test Batteries?

Everyone knows that when psychologists need to assess subject's abilities they often use test batteries. For example, applicants for a clerical job will be asked to pass standardized tests to evaluate specific skills and abilities: e.g., arithmetic skill, verbal skill, and office skills. Each one of the tests is intended to inform us about the candidate's proficiency. Now, is it possible to imagine a situation where a single proficiency test is capable of predicting correctly a candidate's suitability for any type of job? Surely not! Well, to our opinion this is the core problem with which personality testing is confronted. How can we expect to grasp all facets of personality linked to a particular job with a single personality test, even with the “best one” in hands? Like aptitude tests, we have the opinion that personality tests differ on what they measure or pretend to measure. Some are more job-oriented, others stand at a more general level of description, while others belong to the clinical domain, etc. Thus, each test can only give a partial overview of a person's personality. The point is then, why should we not also adopt personality test batteries, akin to what is currently done in aptitude testing?

For job selection, not only should psychologists make use of an appropriate personality test, but

even use different ones at the same time. As all test batteries, such a procedure is certainly time consuming but is still very profitable: different levels of personality data inform us about the coherence or in(coherence) of the candidate's profile. Moreover, information coming from various personality tests can be confronted with one another to add new information or to detect biases in response modes. In other words, more accurate and complex conclusions can be drawn from a personality test battery procedure than from a single personality test. To check this out, we now leave these theoretical considerations to illustrate a practical case. For that, we need to introduce IPSO, the only personality test battery used in France, and possibly one of the few currently existing worldwide.

IPSO, a Personality Test Battery

IPSO (Léonard, 2002) is a French-Canadian personality test battery. It is employed for personnel selection in different types of organizations (e.g., industry, banking and insurance, military). It is principally used to assess work positions involving decision making qualities (e.g., officers, cosmonauts, managers, truck drivers). By now, more than 30 scientific articles have been published on IPSO, most of which are in French (the publication list is available upon email request), and the battery has been administered to over 20,000 persons in France. The IPSO battery is comprised of four different, but complementary, personality tests:

- IP9 detects possible adaptation difficulties and health risks within organizations,
- TD9 evaluates decision making personality typologies,
- QMA evaluates motivation to action, and
- SGP screens for candidate's management styles.

There are two specific aspects about IPSO. The first one is that not all tests have the same weight in the psychological diagnostic. For example, while it is easy for candidates to bias in some way the management style questionnaire (SGP), for social desirability purposes, experience shows that such biases are more difficult to perform with the IP9 and TD9 tests. Therefore, response biases are, in general, easily detected when comparing IP9 and TD9 to the other tests. The second important aspect of IPSO concerns the score analysis of each test. Interpretation is not a linear juxtaposition of

scores belonging to the different tests, but the interaction of all available information. By this procedure, simplistic personality interpretations vanish. Sometimes results of each test converge towards an unambiguous interpretation, sometimes high or low scores are tempered by comparing them to the other tests, especially to IP9 and TD9. Over time, construct and predictive validity procedures have provided evidence supporting this personality assessment method (Léonard, 2002). Above all, we think this "success" is mainly due to the fact that IPSO is a personality battery test.

Conclusion

For some professions, such as those involving high levels of decision making, personality assessment is crucial. Regardless of this fact, we have pointed out the strong difference existing between psychologists as to whether personality evaluation is a judicious procedure or not in job selection. In presenting IPSO, we intended to show that personality testing can become a very powerful predictive procedure when personality test batteries are used as in aptitude evaluation. This paradigm should be reserved for jobs where personality facets are more crucial than a skills evaluation. The choice of the personality tests comprising the test battery should be wisely chosen in order to give different yet complementary information on the candidate's personality. Since personality is a fascinating domain of research in science and is, without doubt, crucial for organizational psychology, our intention was to present to psychologists that personality testing can become even more powerful and interesting using the personality test battery procedure we have presented. We would enjoy having feedback commentaries from readers about this procedure. To our view, this "new" paradigm should contribute to increasing predictive validity coefficients in personality testing.

References

- Bing, M.R, Whanger, J.C., Davison, H.K., & Vanhook, J.B. (2002). Incremental validity of the frame-of-reference effect in personality scale scores: A replication and extension. IPMAAC's 26th Annual Conference, New Orleans, LA.
- Bruchon-Schweitzer, M., & Lievens, S. (1991). Le recrutement en Europe. Psychologie et Psychométrie (numéro spécial), 12 (2).
- Courrier Cadres n°1384-16 février 2001. Paris, France.

- Damasio, A. (1994). Descartes' error: Emotion, reason, and the human brain. Putnam, New York.
- Guion, R.M., & Gottier, R.F. (1965). Validity of personality measures in personnel selection. Personnel Psychology, 18, 135-164.
- Hogan, R.T. (1991). Personality and personality measurement. In M.D. Dunnette & L. Hough (Eds.), Handbook of industrial & organizational psychology, Vol. 2. Consulting Psychologists Press, Inc.
- Hogan, R., Hogan, J., & Roberts, B. (1996). Personality measurement and employment decisions: Questions and answers. American Psychologist, 51, 469-477.
- Léonard, C. (2002). IPSO, manuel du praticien. Paris, France: Editions et Applications Psychologiques.
- Schmidt, F.L., & Hunter, J.E. (1998). The validity utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124, 262-274.
- Schmit, M.J., Ryan, A.M., Stierwalt, S.L., & Powell, S. (1995). Frame-of-reference effects on personality scores and criterion-related validity. Journal of Applied Psychology, 80, 607-620.

Using DIF in Test Adaptation for Cross-Lingual Assessment

Avi Allalouf

National Institute for Testing & Evaluation
ISRAEL

Test adaptations have become very common in recent years, as cross-cultural and cross-lingual comparisons of traits, achievements and abilities have spread. If a valid and reliable test exists in one language, and there is a need for a test on the same topic in other languages, translating and adapting is the rational way to proceed. As a rule, tests should be administered in the examinee's native language, unless the aim is to test foreign language ability. Translating is one step in the process of adapting a test from one language (the source) to another language (the target). It is a long process, aimed at generating a translated test form that is equivalent, or at least very similar, to the original, source language test form. The adaptation process should follow a detailed guide (e.g., Van de Vijver & Hambleton, 1996), which usually prescribes a series of consecutive steps such as independent reviews, double translation and back-translation. Since a high degree of similarity in the psychometric characteristics of each of the translated items is essential, several steps in the

adaptation process involve the use of differential item functioning (DIF) analysis.

Definition and Methods

Definition: An item functions differently across groups if examinees of equal ability, but from different groups (here, source and target language groups), do not have an equal probability of responding correctly to that item. In other words, DIF exists if the psychometric characteristics of the source items and translated items are not the same.

Methods: DIF detection can be carried out statistically, after the test forms have been administered, using empirical data. Among the main methods employed for this purpose are: Delta Plot, Mantel-Haenszel, Logistic Regression, IRT Based, and SIBTEST; each method has its own advantages. By examining the item's content, trained experts can, to a certain degree, anticipate DIF without empirical data. Hambleton and Jones (1995) compared judgmental and empirical procedures, and found them to be quite similar.

After statistical DIF analysis, each item is usually assigned one of the following statuses: (a) no or negligible DIF, (b) moderate DIF, or (c) large DIF (*A-*, *B-*, and *C-level* DIF respectively, according to Educational Testing Service (ETS) terminology).

In translation and cross-lingual assessment, DIF detection methods assist in making crucial decisions before and after adapting a test: (a) determining the translatability of tests and items (before), and (b) scoring, equating and maintaining a cross-lingual item bank (after).

Determining the Translatability of Tests and Items

Not every test, test type, item, or item type can be adapted from one language to the other. The decision whether to translate a test or an item should be taken based on findings from cross-lingual DIF studies.

Findings from DIF Studies: Generally, translated items do vary in the amount of DIF found. Angoff and Cook (1988) analyzed the equivalence between the *SAT* and its Spanish-language counterpart, the *Prueba de Aptitud Academica (PAA)*. They found that verbal items that contain more text have lower DIF than items containing less text, where every word is critical and every translation problem has a

crucial effect on item performance. For example, reading comprehension items have lower DIF than verbal analogies. On the other hand, no DIF is expected in non-verbal items such as math or figural items, as noted by Gafni and Melamed (1991). Some studies list the possible causes of DIF between test forms in different languages. For example, Allalouf, Hambleton and Sireci (1999) studied the translation of the verbal reasoning domain of the *PET (Psychometric Entrance Test, which is used in selecting candidates for universities in Israel) from Hebrew to Russian*. They found that DIF is likely to occur if there are differences between source and target language in: (a) word difficulty, (b) item format, (c) cultural relevance, and (d) content. In another study, Gierl and Khaliq (2001) identified four similar sources of DIF in Canadian achievement tests administered in English and French. The sources were: (a) omission or addition of words or phrases that affect meaning, (b/c) differences in words or expressions inherent/not inherent to the language or culture, and (d) format differences between the test forms in different languages.

Implications of DIF Studies: One of the implications of DIF studies is that test constructors and translators should, if made aware of the results of these studies, exercise more care in adapting a test from one language to another. This care should prevent them from translating items suspected of having cultural bias, seek target language words of comparable difficulty, preserve the original item format, and avoid including content that gives an advantage to one of the groups.

Scoring, Equating and Maintaining a Cross-Lingual Item Bank

After the test has been adapted from one language to another, statistical DIF analysis methods are used to locate the items that function differentially in the two languages. Such items may cause bias in scoring in favor of one of the language groups.

Scoring: After the DIF analysis reveals the items whose psychometric characteristics were altered by the translation, these items (depending on the degree of DIF) should not be included in computing the scores. Items with DIF (usually the large DIF items) are to be removed from the scoring process of both forms (source and target), making the test forms more similar and the scores more comparable. Before the DIF items are omitted, they should be examined carefully by subject

matter experts because, in very rare cases, as was pointed out by Ellis (1989), real differences emerging from experience, knowledge or culture may occur in language groups' performance on specific items. These differences may be considered relevant and in such cases, the DIF items are not to be omitted from the scoring process.

Equating/Linking: A translated test can be equated to the source language test even if some items in the target language function differently from the source items. This can be accomplished by using only the non-DIF translated items as common items, or an anchor, for (linear or IRT) equating. Deciding which items are suitable for inclusion in the anchor may be somewhat iterative, since deleting too many DIF items may shorten the anchor too much.

Maintaining cross-lingual item banks: Cross-lingual item banks consist of translated items for future use. DIF items are to be removed from item banks to avoid their use in future tests. This has unfortunate economic implications as translated items are costly to produce. Substantiated findings of DIF studies can guide the revision of a DIF item by suggesting possible causes of DIF. Revision of DIF items can produce non-DIF (or lower-DIF) items that can be saved for future use.

References

- Allalouf, A., Hambleton, R.K., & Sireci, S.G. (1999). Identifying the causes of DIF in translated verbal items. *Journal of Educational Measurement*, *36*, 185-198.
- Angoff, W.H., & Cook, L.L. (1988). *Equating the scores of the Prueba de Aptitud Academica and the Scholastic Aptitude Test* (College Board Report No. 88-2). New York: College Entrance Examination Board.
- Ellis, B.B. (1989). Differential item functioning: Implication for test translation. *Journal of Applied Psychology*, *74*, 912-921.
- Gafni, N. & Melamed, E. (1991). *Equating different language versions of a psychometric test*. Research Report No.148, Jerusalem: NITE.
- Gierl, M.J., & Khaliq, S.N. (2001). Identifying sources of differential item and bundle functioning on translated achievement tests: A confirmatory analysis. *Journal of Educational Measurement*, *38*, 164-187.
- Hambleton, R.K., & Jones, R.W. (1995). Comparison of empirical and judgmental procedures for detecting differential item functioning. *Educational Research Quarterly*, *18*, 21-36.

van de Vijver, F.J.R., & Hambleton, R.K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, *1*, 89-99.

Quality Standards for E-Learning: Cross Validation Study of the Demand Driven Learning Model (DDL M)

Krista Breithaupt
American Institute for CPAs
U.S.A.

Colla J. MacDonald
University of Ottawa
CANADA

Introduction

E-learning is the latest form of educational technology that lends potentially high levels of interactivity, information access, and communication economies to assist in the design and management of learning programs needed by busy adults (ASTD, 2002; Khan, 1997; Mann, 2000). Business and industry employers, as well as an increasing number of educators, recognise the need for informed application of learning technologies to enhance the quality and flexibility of educational and human resource development programs (e.g., Conference Board of Canada, 2000, 2001). E-learning technologies hold special appeal; these programs allow employees to pursue advanced credentials without interrupting their service to employers (MacDonald & Gabriel, 1998). A review of literature from practitioners and researchers in this area has revealed a distressing gap between the use of this technology and sound pedagogical models (Khan, 1997; Salmon, 2000; Willis, 2000). Serious consequences result when e-learning organizations are criticised for poor overall program quality.

Quality has been defined in terms of the design of the e-learning experience, the contextualized experience of learners, and evidence of learning outcomes (Carr & Carr, 2000; Jung, 2000; Salmon, 2000). However, e-learning course design and overall program quality are sometimes compromised in an "...effort to simply get something up and running" in response to pressing demands of the consumer (Dick, 1996, p. 59). Educators and researchers have voiced concern over the lack of appropriately rigorous evaluation studies of e-learning programs (e.g., Cheung, 1998; Lockyer, Patterson, & Harper, 1999; Reeves &

Reeves, 1997). Funding of the development and deployment of novel programs may be emphasized, while resources are not tagged to support expertise for evaluation (Wills & Alexander, 2000). In addition, evaluation methods used in more conventional programs may not be appropriate for e-learning. New evaluation tools, methodologies, and solutions need to be devised (Zuniga & Pease, 1998).

Some researchers have made steps towards developing evaluation instruments to assess e-learning, particularly distance education programs (e.g., Biner, 1993; Cheung, 1998; Stanford, 1997). These are most often developed to assess only specific program content and are not suitable for wider application. Perhaps this lack of generic utility is the reason why these measures are usually not subjected to rigorous psychometric study. There appears to be a need in many fields of education for validated e-learning measures with desirable psychometric properties.

Collaboration between the consumers (learners), administrators, educators, and researchers is required to ensure effective design, management, and evaluation of e-learning programs (ASTD, 2002; Horton, 2001). Quality assurance and accountability in e-learning programs have become critical issues in our competitive economic environment where educational institutions compete with one another for efficient high-quality programs. This study extends the theoretical work and pilot study that led to the development of a new model for e-learning. This model is the Demand-Driven Learning Model (DDL M; MacDonald et al., 2001) which was created to fill the widening gap between sound pedagogy and economic pressures for accountability. The DDL M is offered as one strategy for evaluating e-learning programs via application of a quality assurance framework or model.

The DDL M as a Quality Standard

The DDL M was developed through a collaborative process between industry experts and academics, the latter having strong foundations in curriculum design, evaluation methods, and psychopedagogy. After an exhaustive review of the literature on constructs related to e-learning, members of the research team constructed a conceptual framework to address high quality standards within e-learning. Early and continual involvement of end users (i.e., industry experts) in the design kept the model

relevant for learners (working adults), secondary beneficiaries (employers), and educators (including program designers or providers). A graphic representation of DDLM constructs is offered in Figure 1.

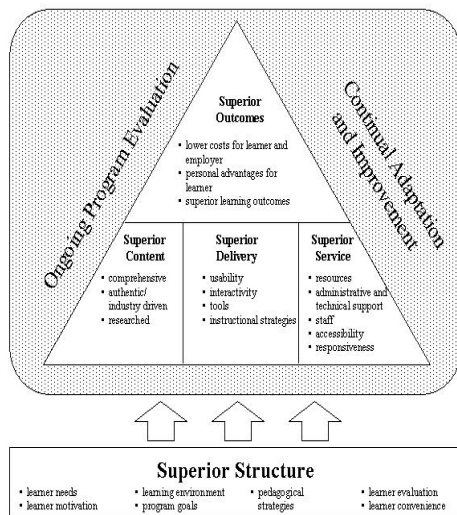


Figure 1: DDLM Dynamic Structure

Throughout the development process, industry experts stipulated that outcomes should become a vital component of the model. Practitioners wanted a model that would evaluate the sustained acquisition of workforce-specific task skills as distinct from a mere 'satisfaction' survey. So our inclusion of the outcome construct in the model represented a new focus compared to constructivist instructional design (ID) models in the extant literature. The corollary of this practical focus was the companion online DDLM survey, developed for evaluating e-learning programs with consistency and precision. Aligning the evaluation tool with an appropriate theoretical framework also represented a useful and innovative contribution to online education and training.

The Online DDLM Survey

The DDLM survey questions were targeted to elicit the learner's appraisal of the efficiency and ease of participating in the program, and the quality and relevance of curriculum content. An exhaustive collection of 196 possible survey items were reduced to 59 items and five subscales in the pilot study. The final number of items retained was: 24 items for Structure, 10 items for Delivery, 8 items for Service, 8 items for Content and 9 items for Outcomes. All subscales had high reliabilities (alpha values ranged from .93 to .97). Questions in the first four subscales were accompanied by five response options: never, rarely, sometimes, often, or always. Questions in

the Outcomes subscale were presented with four response options; strongly disagree, disagree, agree, or strongly agree. Higher scores were associated with positive responses. A sample question for each section is presented in Figure 2.

[Content]
<i>The content includes</i> ...realistically complex learning tasks which are similar to those faced in the workplace.
[Delivery]
<i>Presentation of material on the site features</i> ...easy to find and use screen elements.
[Service]
<i>Administrative and technical support</i> ...complaints are quickly handled by professors / learning facilitators.
[Superior Structure]
<i>The course</i> ...respects my experience and current knowledge.
[Outcomes]
<i>As a result of my participation in this course</i> ...I have applied new knowledge in the workplace.

Figure 2: Examples of DDLM Survey Questions

The DDLM is not limited to any content area and may be used to evaluate e-learning instruction of any topic. Currently, the DDLM is being used as a quality standard to design, develop, deliver, and evaluate teacher education and graduate education courses at the University of Ottawa in Canada. It has been used as a quality standard to develop and evaluate MBA courses at Royal Roads University in Canada. The DDLM is also being used as a quality standard to design, develop, deliver and evaluate training to health care teams in three Canadian provinces and to evaluate in-service education for teachers in the state of New York, U.S.A.

Research Questions

The focus of the present cross-validation study is on generalizations from our initial pilot test of the DDLM and the online survey developed for formative curriculum evaluation. Our primary motivating questions were the following:

1. Does new data using the online survey retain desirable psychometric properties?
2. Are the expected dynamic relationships among DDLM constructs supported in the current study?

Results and Discussion

A total of 21 learners enrolled in e-learning programs responded to the DDLM online survey for the cross-validation study. This represents about 30% of learners registered in a graduate-level university e-learning program during January and February of 2002. Participants were similar to the pilot sample described in MacDonald et al. (2002; see Table 1). That is, they tended to be age 30 and older, married with children and the majority was male.

Table 1
Cross-Validation and Pilot Sample Participant Characteristics

<i>Characteristics</i>	<i>Cross Validation Sample N (%)^a</i>	<i>Pilot Sample N (%)^a</i>
<i>Gender</i>		
Male	18 (86)	54 (58)
Female	3 (14)	37 (40)
<i>Education</i>		
University	14 (67)	39 (42)
College	6 (29)	52 (56)
High school	1 (5)	2 (2)
<i>Age</i>		
<30 years	1 (5)	3 (3)
31 to 40	10 (48)	88 (95)
41 to 50	7 (33)	2 (2)
over 50	3 (14)	
<i>Marital Status</i>		
Single	1 (5)	13 (14)
Single parent	4 (19)	4 (4)
Married, with children	14 (67)	52 (56)
Married, no children	2 (10)	22 (24)
<i>Total Participants</i>	21	93

^a Some learners did not provide responses to some demographic questions in either study, some percentages are rounded and will not sum to 100%.

Alpha values for the subscales in the current study ranged from .88 to .96 and were comparable to those in the pilot study. Mean scores for each subscale were: Delivery M = 4.1, Service M = 3.9, Content M = 3.6, Outcome M = 3.5 and Structure

M = 4.0, and were similar to the means obtained in the pilot study. Strong positive correlations were expected among the scores on the five subscales and are presented in Table 2.

Table 2
Pearson Correlations Among DDLM Sub Scale Scores^a

<i>Sub Scale</i>	<i>Structure</i>	<i>Delivery</i>	<i>Service</i>	<i>Content</i>
<i>Structure</i>	1.0			
<i>Delivery</i>	.659*	1.0		
<i>Service</i>	.678*	.713*	1.0	
<i>Content</i>	.559*	.419*	.117	1.0
<i>Outcomes</i>	.563*	.523*	.359*	.473*

^aN ranged from 18 to 21, * p < .01.

Superior Structure was expected to be a holistic measure of desirable e-learning program features or a high quality standard. By contrast, the Content, Delivery, Service and learner Outcomes can be viewed as indicators that have a logical predictive relationship to this goal. In the pilot study, a regression model based on this hypothesis was tested and found to be statistically meaningful ($R^2 = .48$; $df = 5,87$; $F = 16.3$, $p < .01$). Delivery and Service scores were the strongest (and statistically significant) indicators of Superior Structure. The sample size of the present study was inadequate for a similar analysis, but correlations between Superior Structure and each subscale ranged from .559 to .678 and were similar to those found in the pilot study. As with the pilot study, the weakest relationship was found between Content and Service ($r = .117$).

Some evidence of discriminant validity was found. Specifically, DDLM scores were not strongly related to demographic variables. Correlations between educational level and DDLM variables were low, non-significant and, in some cases, negative. This is similar to findings from the pilot study.

Conclusion and Implications

The similarity of findings from the pilot and validation samples provides some confidence that the underlying relationships among DDLM constructs are generalizable. Scale score properties in both studies suggest these items measure consistent underlying traits and scores have adequate precision. The DDLM survey tool, and the quality standard defined as Superior Structure, are offered as one means of evaluating and improving e-learning programs. Desirable qualities that comprise Superior Structure are consistent with the need for improved access to lifelong education opportunities for many adult learners who require a combination of part-time or full-time study for retraining, advancement, career change, or response to downsizing.

Any study of the validity of inferences from a generic measure of the quality of e-learning must be referenced to an underlying theoretical framework. To this end, the DDLM was used as the basis for an evaluation measure that could be administered online to examine the effectiveness of an e-learning program from any discipline. Our intention is to allow this measure to be applied easily for immediate feedback to program providers, and we encourage others to apply and evaluate programs using the DDLM survey. Possible benefits include rapid identification of strengths and weaknesses of various programs, timely intervention and resource allocation.

Note: This research is part of a three-year project funded by the Office of Learning Technologies (Department of Human Resources Development Canada, Government of Canada). The DDLM survey is available from the authors: kbreithaupt@aicpa.org, cjmacdon@uottawa.ca or in MacDonald et al. (2002).

References

- ASTD: American Society for Training and Development. (2002). The ASTD e-learning handbook: Best practices, strategies, and case studies for an emerging field. Edited by A. Rossett. New York: McGraw-Hill
- Biner, P.M. (1993). The development of an instrument to measure student attitudes toward televised courses. The American Journal of Distance Education, 7(1), 62-73.
- Carr, C.S., & Carr, A.M. (2000). Instructional design in distance education (IDDE): A web-based performance support system for educators and designers. The Quarterly Review of Distance Education, 1 (4), 317-325.
- Cheung, D. (1998). Developing a student evaluation instrument for distance teaching. Distance Education, 19(1), 23-42.
- Conference Board of Canada. (2001) E-Learning for the workplace: Creating Canada's lifelong learners. Online at http://www.conferenceboard.ca/education/pdf/e-learning_for_the_workplace.pdf Retrieved September 27, 2002.
- Conference Board of Canada (2000). Solutions for employers: Effective strategies for using learning technologies in the workplace. Online at <http://www.conferenceboard.ca/education/pdf/solutions1.pdf> Retrieved September 27, 2002.
- Dick, W. (1996). The Dick and Carey model: Will it survive the decade? Educational Technology Research and Development, 44(3), 55-63.
- Institutional concerns. The Quarterly Review of Distance Education, 1(4), 281-291.
- Horton, W. (2001). Evaluating e-learning. Alexandria, VA: American Society for Training and Development.
- Jung, I. (2000). Technology innovations and the development of distance education: Korean experience. Open Learning, 15(3), 217-231.
- Khan, B. H. (1997). Web-based instruction (WBL): What is it and why is it? In B.H. Khan (Ed.), Web-based instruction (pp. 5-18). Englewood Cliffs, NJ: Educational Technology Publications.
- Lockyer, L., Patterson, J., & Harper, B. (1999). Measuring effectiveness of health education in a web-based learning environment: A preliminary report. Higher Education Research and Development, 18(2), 233-246.
- MacDonald, C.J., Breithaupt, K., Stodel, E., Farres, L., & Gabriel, M.A. (2002) Evaluation of web-based educational programs: A pilot study of the Demand-Driven Learning Model. International Journal of Testing, 2(1), 35-61.
- MacDonald, C.J., & Gabriel, M.A. (1998). Toward a partnership model for web-based learning. The Internet and Higher Education, 1(3), 203-216.
- MacDonald, C.J., Stodel, E., Farres, L., Breithaupt, K., & Gabriel, M.A. (2001). The demand driven learning model: A framework for web-based learning. The Internet and Higher Education, 1(4), 9-30.
- Mann, B.L. (2000). Perspectives in web course management. Toronto, ON: Canadian Scholar's Press.
- Reeves, T.C., & Reeves, P.M. (1997). Effective dimensions of interactive learning on the World Wide Web. In B.H. Khan (Ed.), Web-based instruction (pp. 59-66). Englewood Cliffs, NJ: Educational Technology Publications.
- Salmon, G. (2000). E-moderating: The key to teaching and learning online. London: Kogan Page.
- Stanford, S.W. (1997). Evaluating ATM technology for distance education in library and information

science. *Journal of Education for Library and Information Science*, 38(3), 180-190.

Willis, J. (2000). The maturing of constructivist instructional design: Some basic principles that can guide practice. *Educational Technology*, 40(1), 5-16.

Wills, S., & Alexander, S. (2000). Managing the introduction of technology in teaching and learning. In T. Evans & D. Nation (Eds.), *Changing university teaching: Reflections on creating educational technologies* (pp. 56-72). London, England: Kogan Page.

REPORTS

On-line Readings in Testing and Assessment (ORTA) Project: Progress Update

Cheryl Foxcroft
University of Port Elizabeth
SOUTH AFRICA

Marise Born
Erasmus University Rotterdam
THE NETHERLANDS

In the December 2002 edition of *Testing International*, readers were introduced to the broad parameters of the On-line Readings in Testing and Assessment (ORTA) project. To refresh readers' memories, the vision of ORTA is for the ITC to provide a number of independent, yet integrated readings on aspects related to testing and assessment on-line and free of charge for the purposes of:

- Making a meaningful contribution to the teaching of psychometrics, testing, and assessment internationally so as to better prepare psychometricians, assessment practitioners, and researchers for modern day testing and assessment, and
- Making quality assessment and testing readings available to instructors and students in developing countries in particular as it is often difficult for scholars and researchers in these countries to gain access to American or European textbooks or journal articles.

The readings will be provided in a number of units. However, rather than adopt too rigid an approach to the units to be included and how they will be grouped, experience has already shown, based on input received on the article that appeared in the last edition of *Testing International*, that it may well be wise to be flexible regarding the topics for the readings as people in the field have already made some useful suggestions. Thanks to those who have come forward with suggestions to date.

Identification of Possible Contributors

Some potential contributors have contacted the editors as a result of the article in the last edition of *Testing International*. ITC Council members have also suggested potential names. Letters of invitation are to be sent to these potential contributors during June. However, we are still on the lookout for more contributors. So, if you work in the field of psychometrics, testing and assessment and would like more information about ORTA and/or would like to prepare a reading for ORTA please feel free to contact either Cheryl Foxcroft (cheryl.foxcroft@upe.ac.za) or Marise Born (born@fsw.eur.nl).

We are particularly interested to hear from academics, researchers and practitioners in developing countries. This is because we believe that for ORTA to achieve its vision, we need contributions related to topics such as the history of psychometrics and psychological testing in various developing countries, the impact of certain laws or political ideologies on the practice of testing and assessment in a country, or the context-specific challenges to fair assessment practices in a country. This will enable readers from developing countries to gain an overview of the testing and assessment trends and issues in developing countries, which should make the readings more relevant and "real" for them.

ITC Website

<http://www.intestcom.org>

Check out the ITC Guidelines on: (a) **Adapting Tests**, (b) **Test Use**, and (c) **Computer-Based and Internet Delivered Testing** *NEW!*

ITC's International Guidelines on Computer-Based and Internet-Delivered Testing

Dave Bartram
SHL Group PLC
U.K.

Iain Coyne
University of Hull
U.K.

We are pleased to invite ITC members to comment on the first draft version of the ITC's International Guidelines on Computer-Based and Internet-Delivered Testing.

The process used to develop these guidelines is described in the document, as is the role the Guidelines are intended to serve. We are sending copies of the current draft to a wide range of international experts in the area of testing and test use, seeking their comments on the form, structure, and content of the Guidelines. We are also interested in receiving comments on the uses to which the Guidelines might be put and advice on how best to disseminate the final version. However, we must emphasise that this is only a consultation document. It should not be regarded as having any formal status at this time.

This is the first of a number of stages of consultation that are planned, with the goal of producing the final document for launch at the Beijing IUPsyS Congress in August 2004.

A copy of the International Guidelines on Computer-Based and Internet-Delivered Testing document and a feedback form can be found at: http://www.intestcom.org/itc_projects.htm. Once at the above website, simply click on the guidelines – which are referred to in the very first paragraph under ITC Projects – or scroll down to the bottom of the page. If you do not wish to use the feedback form, please feel free to provide comments in any other format you prefer. If you have detailed comments to make on specific guidelines, you may find it easiest to annotate a copy of the original document and return that to us.

We look forward to receiving your comments as soon as possible. Please send your comments to the following address:

Dr. Iain Coyne (ITC Guidelines Project)
Department of Psychology
University of Hull
HULL, HU6 7RX, England
Tel/fax +44 (1482) 465932
email: I.J.Coyne@psy.hull.ac.uk

Portuguese News about Psychological Tests

Aura Montenegro
University of Coimbra
PORTUGAL

This piece focuses on the important theme of retirement. The impact of this life event depends on many variables; however, it corresponds to a new aspect of development between adulthood and old age. Therefore, psychologically and socially significant changes must be studied by means of instruments that offer reliable and valid inferences. Accordingly, two Portuguese researchers, A.M. Fonseca and C. Paúl of the Institut of Biomedical Sciences Abel Salazar at Oporto University conducted a study adapting and validating the Retirement Satisfaction Inventory (Floyd, et al., 1992). In doing so, they analyzed the three scales and their respective subscales: (1) Reasons for Retirement: professional stress, external pressures, personal interests, unexpected circumstances; (2) Life Satisfaction: services and community resources, health and physical activity, family life and marriage; and (3) Pleasure Motives: stress absence, social activities, freedom and personal control.

The Portuguese adaptation and validation of the Retirement Satisfaction Inventory (RSI) respects the essential characteristics of the original version and a French version. It is composed of 55 closed answer items and the meaning fits the structure of the scales.

Methodology

Participants and Procedure: The heterogeneous sample consisted of 253 retired persons (52% female, 48% male) ages 50 to 88 years ($M = 65.2$ yrs) living in the north of Portugal. School levels were 41% basic school, 31% secondary school, and 28% superior formation. As to their origin, 81% were from urban or suburban areas and 19% from rural areas. Administration of the RSI was

done individually or in small groups. No participant resided in an institution.

Analyses: This study used factor analysis with varimax rotation as conducted in Floyd et al. (1992). Four factors were identified in the Reasons for Retirement and Life Satisfaction scales whereas three factors were identified in the Sources of Enjoyment scale. The 43 items included in the definition and characterization of the three scales showed loadings of greater than .25 on their respective factors. Items did not load on more than one factor. Internal consistency coefficients, as determined using Cronbach's coefficient alpha, were considered acceptable for the identified factors.

Conclusions

Essentially, the authors verified the complexity of the task of evaluating life satisfaction during retirement. Several factors, such as family life and marriage, physical health, services and community resources, performed an important role. All of these factors were present in the samples from the U.S., France, and Portugal. This study shows the psychometric properties of the RSI when used with a Portuguese sample.

The RSI is particularly useful for the following goals:

1. Understanding the adjusting standards at the beginning of retirement and the factors that influence life satisfaction at this time.
2. Planning counselling and psychological interventions by allowing for the identification of relevant problems that may affect psychological wellbeing and self-adaptation.

For further information, readers are directed to the following journal: *Psychologica / Revista da Faculdade de Psicologia e Ciências da Educação de Coimbra*. 2002, 29, 169-180.

Become a Member of the ITC

ITC membership is open to organizations as well as individuals. Information and forms are available at:

<http://www.intestcom.org/membership.htm>

THE NOTICEBOARD

Paul Meehl Passes Away

Paul Everett Meehl, Professor Emeritus of Psychology at the University of Minnesota, died on February 14, 2003 at the age of 83. Professor Meehl is perhaps best known internationally for his work on the MMPI, his consideration of the issue of clinical versus statistical prediction, his exploration of the role of construct validity, and his arguments favouring the role of genetics in schizophrenia. As a graduate student in the 1940s, he worked on what ultimately became the K scale of the Minnesota Multiphasic Personality Inventory (MMPI) for his doctoral dissertation. In 1954, he published what he has referred to as his "disturbing little book" entitled "Clinical and Statistical Prediction: A Theoretical Analysis and Review of the Evidence", which showed that clinicians' predictions were either inferior, or at best equal, to those obtained from statistical formulas. A year later, his classic paper with Lee Cronbach, entitled "Construct validity in psychological tests" was published in the *Psychological Bulletin*. In 1962, when he was president of the American Psychological Association (APA), Meehl argued against the commonly held belief that 'bad mothers' or other environmental conditions caused schizophrenia in favour of a genetic explanation. His university web site lists over 180 papers, books and other publications.

Meehl received his Bachelor's (1941) and Doctoral (1945) degrees in Psychology from the University of Minnesota. He began teaching there as an instructor in 1944 and was a Full Professor by 1952. At various times in his career, he held faculty appointments in Psychology, Law, Psychiatry, Neurology and Philosophy. He was a former Director of the Clinical Psychology Training Program and former chair of the Department of Psychology. He also held the University of Minnesota's highest faculty honor, Regents Professor, from 1968 until 1990 when he retired. He was a former President of the American Psychological Association and held numerous awards and honors, including (to name but a few): the American Academy of Arts and Sciences (1965), Bruno Klopfer Distinguished

Contribution Award from the Society for Personality Assessment, (1979), the National Academy of Sciences (1987), Gold Medal Award for Life Achievement in the Application of Psychology from the American Psychological Foundation (1989), Educational Testing Service Award for Distinguished Service to Measurement (1994), Award for Outstanding Lifetime Contribution to Psychology from the American Psychological Association (1996), and Distinguished Lifetime Contribution to Evaluation, Measurement, and Statistics from APA Division 5 (1997). In a 2002 article in *Review of General Psychology*, Professor Meehl was identified as one of the 100 most eminent psychologists of the 20th century.

A PDF version of Meehl's fascinating autobiography from Vol. 3 (pp. 337-389) of G. Lindzey's (Ed.) (1989). *A History of Psychology in Autobiography* can be found on Meehl's webpage at: <http://www.tc.umn.edu/~pemeehl/>.

UPCOMING CONFERENCES ON TESTING & ASSESSMENT

American Association for Higher Education (AAHE) Assessment Conference

June 22-24, 2003

Preconference Workshops: June 21 & 22, 2003
Seattle, WA, United States

Website: <http://www.aahe.org/assessment/2003/>

The conference will explore the theme "A Richer and More Coherent Set of Assessment Practices". Specifically, the focus will be on exploring relationships among teaching, learning, and the design of assessment methods.

1st National Conference on Psychological Assessment / VI International Conference on Psychological Assessment

July 23-26, 2003

Campinas, São Paulo, Brazil

Website: www.ibapnet.org.br

The National conference will be hosted by the Brazilian Institute on Psychological Assessment, in coordination with the Iberian-Latin American group, to celebrate the VI International Conference. The conference will focus

on the theme of Science and Social Responsibility with presentations on: educational and psychological test development, neuropsychological testing, politics of test development, testing in different cultural contexts, standards for test construction, ethics on test use, assessment in forensic psychology, testing children with disabilities, state-of-the-art testing development in Latin-American countries.

VII Psychodiagnostic National Congress

October 9-11, 2003

Mendoza, Argentina

Website: www.adeip.org

Contact e-mail: adeip@arnet.com.ar

Contact tel/fax: (0341) 4240013

The theme this year will be "Psychodiagnosis: Issues and contexts".

4th Association for Educational Assessment (AEA) - Europe Conference

November 6-8, 2003

Lyon, France

Website: <http://www.aea-europe.net/page-125.html>

Contact e-mail: aeaeuro@attglobal.net

This conference will be in conjunction with the Agence Nationale de Lutte Contre l'Illettrisme (ANLCI) on the theme "Assessment Challenges for Democratic Societies". Topics will include: literacy assessment, numeracy assessment, international surveys, assessment within universities, and computerised testing. Previous annual conferences have been held in: Prague, Czech Republic; Krakow, Poland; and Frascati, Italy. The 2004 conference is planned to be held in Budapest, Hungary.

CASMA-ACT Invitational Conference on Current Challenges in Educational Testing

November 8, 2003

Iowa City, Iowa, U.S.A.

Website: www.act.org/casma

Contact e-mail: coe-casma@uiowa.edu

This one-day conference, sponsored by the Center for Advanced Studies in Measurement and Assessment (CASMA) and not-for-profit test publisher ACT, is geared toward measurement specialists, practitioners, and policymakers. Attendance is limited to 200. The registration deadline is Oct. 15, but early registration (with fee reduction) is Sept. 30.

2nd International Conference on Measurement in Health, Education, Psychology and Marketing: Developments with Rasch and Unfolding Models

January 20 - 22, 2004

Perth and Fremantle, Western Australia

Pre-conference courses on Rasch measurement are available from Jan. 5-9 and 12-16 and a one-day introductory workshop will be offered on Jan. 19.

Website:

http://www.education.murdoch.edu.au/educ_Rasch_January2004.html

E-mail: g.luo@murdoch.edu.au

Call for Abstracts: Deadline is Aug. 31, 2003. Topics of interest include: epistemology, fundamental measurement and Rasch models; cumulative models for attitude and trait measurement - dichotomous and ordered category models; Rasch model applications in education (e.g., large scale test equating, benchmarking), psychology (e.g., intelligence testing, linking quantitative and stage developmental data), marketing (e.g., pairwise designs for preference and choice studies), and health care (e.g., cross-cultural validity); item banking; computer adaptive testing; using simulation studies for clarifying methodological issues (e.g., tests of fit); developments in Rasch modelling (e.g., differential item functioning); and history and philosophy of measurement and Rasch models.

VII European Association of Psychological Assessment

April 1-4, 2004

Málaga, Spain

Website: <http://www.uma.es/petra/eapa2004>

Contact e-mail: eapa2004@uma.es

Call for Proposals: Deadline is Sept. 30, 2003.

No details about this conference are available yet.

National Council on Measurement in Education

April 13-15, 2004

San Diego, CA, U.S.A.

Website: <http://www.ncme.org>

Call for Proposals: Deadline is Aug. 11, 2003.

The NCME conference is held in conjunction with the American Educational Research Association (AERA) conference. Topics of interest include: policy, legal, and ethical issues; classroom assessment; performance or alternative assessment; large-scale assessment; licensure and certification testing; computer-based

testing; standard setting; technical and statistical issues in test development; equating and other test score transformations; statistical characteristics of items (including DIF); statistical measures of test quality, including reliability and validity; and test use with specific populations. Note: Nonmembers must become members or be sponsored by a member to submit a proposal to the conference.

The Japan Association for Language Teaching (JALT) Testing & Evaluation SIG

2004 Conference

E-mail: yvonne@sc4.so-net.ne.jp

In May, 2003, the JALT Testing & Evaluation SIG sponsored a conference on communicative language testing at the Kyoto Institute of Technology as part of a JALT Pan-SIG Conference. There were about 20 different presentations on testing and assessment. The JALT Testing and Evaluation SIG is also considering events for a 2004 Conference. If you are interested in presenting at this conference, please contact Yvonne Ishida (Ritsumeikan University).

Submission deadline for the December 2003 issue of *Testing International* is November 1, 2003.

Please submit all articles and reports (preferably as IBM PC-compatible Word or WordPerfect files) to:

**Dr. Anita Hubley, Editor
Testing International
Dept. of ECPS, 2125 Main Mall
University of British Columbia
Vancouver, BC
CANADA, V6T 1Z4**

Or via e-mail at anita.hubley@ubc.ca

Please feel free to duplicate copies of *Testing International* and distribute them to your colleagues. Address any correspondence regarding *Testing International* to the Editor at the address shown above.