

MAPAC News

Mid-Atlantic Personnel Assessment Consortium, Inc.
Fall 1999

President:

Linda Dunn
Borough of Manhattan Community
College
199 Chambers St., NY, NY 10007
(212) 749-2710
DUNNLND@cs.com

Past-President:

James E. Frankart
Pennsylvania State Civil Service
Commission
320 Market Street
Harrisburg, PA 17108-0569
(717) 783-2697, Ext. 3517
(717) 783-0838 (FAX)
JFrankart@scsc.state.pa.us

President-Elect:

Elliot D. Lasson, Ph.D.
MD Dept. of Budget and Mangmt.
301 W. Preston St.
Baltimore, MD 21201
(410) 767-4928
(410) 333-7674 (fax)
ELASSON@dbm.state.md.us

Acting Treasurer:

Will Martin
Principal Personnel Examiner
New York State Dept. of Civil Service
Testing Services Division, Room 345
Albany, NY 12239
(518) 457-2359
(518) 457-4239 (fax)
wjm1@cs.state.ny.us

Incoming President's Message

Holiday Greetings To all MAPAC Members and Friends

The New Year fast approaches with the changes and challenges of every new year magnified as we enter the 21st century. With sadness, I would like to thank Charley Sproule and Betsy Kaido, who will be formally leaving MAPAC in January, for their unparalleled contributions to the organization. Charley, one of the MAPAC founders and currently initiator and chair of the joint projects committee, has had so many leadership roles that I can't quite imagine the organization without him. In word and deed, he represents all that MAPAC strives to achieve. Betsy is currently chair of the newsletter and probably the most prolific MAPAC presenter in history! For MAPAC's membership, these changes could not be greater.

On the positive side, I would like again to congratulate Elliot Lasson on his election as president-elect and look forward to his programs in the coming year. His insight, knowledge and efficiency have been well demonstrated to me as I have worked with him on the program committee in the past year.

Next year's committee chairs and I have spent some time talking about MAPAC in the next year. Top priorities remain those set by Jim Frankart last year—the increased committee participation by all members and

the overall importance of jurisdictions' participation. The many activities of Charley's collaborative projects committee are visible testimony to this effort. His committee will present drafts of the current collaborative projects at the January meeting and these drafts will be the impetus for continued work next year.

In addition, we would like to increase the participation of those jurisdictions that are not now active, expand the current membership and once again to increase outreach with I/O and other relevant student populations. More to follow on these issues at the January meeting!

I wish you all the happiest of holiday seasons! The MAPAC family will be with me in mind in the coming month. I look forward to seeing you all in January and to leading MAPAC in the Year 2K.

Sincerely,

Linda Dunn
President

Outgoing President's Message

Over the past two years, I was fortunate enough to serve as your President- Elect and President. The time has gone by quickly and now I turn the leadership over to Linda Dunn and Elliot Lassen. I am sure they will continue to carry on the MAPAC torch with honor and

President's Message	1
Fall 1999 Business Meeting	2
Fall Meeting Presentations	3
What Does the Literature Say?	14
Upcoming Events...	15

2 MAPAC NEWS

dignity.

I want to thank all the Officers and members who have made my experience a joyful one. I am proud to be a part of a profession whose individuals so generously share their time and talents. Through your continued cooperation we can make MAPAC even bigger and better. Thanks again.

James Frankart
Past President

MAPAC Fall Business Meeting Minutes

September 30, 1999
New York, New York

The meeting was called to order by President Jim Frankart.

Roll Call: 12 member jurisdictions were present.

Reports

Treasurer: We currently have \$12,323.98 in the treasury.

Program: MAPAC will continue to respond to the member's interests and needs in programming by surveying members. The committee was thanked for doing a fine job this year.

Training: The Job Analysis course was offered in September 1999 and will probable be held again in January. The training committee is also looking for volunteers to work the IPPMAC to update the Exam Planning course.

Membership: Jim Frankart will head the membership committee. He reported that Massachusetts

3 MAPAC NEWS

is interested in joining.

Communications: We are looking for members to be part of this committee.

Joint Projects: The group continues to work on the topic of supervisor assessment. The roles of the work groups were summarized and Charley Sproule suggested that any MAPAC jurisdiction that complete the assessment survey and information forms and agree to exchange supervisory test materials, be able to receive copies of the supervisor test items MAPAC compiled earlier.

Old Business

Gifts - Maryland DOT will select a new gift to use to thank the presenters.

Web Master - David Hamill was thanked for working on the new MAPAC web page.

New Business

Elections – Elliot Lasson was unanimously chosen as president-elect.

Social Hour – MAPAC will pay for the food and paper products for the social hour but contributions should be solicited for the beverages.

IPMAAC – Robin Talsnik is soliciting names for speakers or presentations for the June 2000 IPMAAC meeting. It was voted that MAPAC would pay for refreshments for one break at this meeting.

Training – MAPAC unanimously

agreed to cover the cost for room for MAPAC instructors as circumstances warrant.

Submitted by:
Ann Stewart
Delaware State Personnel Office

Tribute to Jerry Moylan Gerard Clifford Moylan Torruella (1949-1999)

We were all saddened last summer at the tragic loss of one of our members, Jerry Moylan. Jerry was killed in a bicycling accident when he was on vacation in Spain. Jerry was a long time member of MAPAC. He served as Treasurer in 1983-84 and was Chair of our Affirmative Action Committee in 1980. To honor his memory, the officers of MAPAC decided that we would present a plaque to his family to thank them for sharing Jerry with us. His sister, Patricia, who lives in New York was invited to attend the presentation. We were pleasantly surprised when his brother Lawrence and daughter Alexandra drove up from Baltimore and joined us for the ceremony.

Those who knew Jerry personally offered remembrances. I offered my recollection of how Jerry had made me feel welcome at MAPAC. Several others also spoke of how Jerry had been a positive influence in their lives. A plaque was then presented to the family on behalf of MAPAC.

4 MAPAC NEWS

Jerry's sister Patricia accepted the plaque on behalf of the family. Patricia said it best when she said that as long as individuals and groups continue to remember and honor his name, then Jerry is still with us. MAPAC can continue to honor Jerry by keeping our standards as high as his were.

Submitted by:
James Frankart
Pennsylvania Civil Service
Commission

Fall Meeting Presentations

MAPAC Training Workshops

Traditional and Not-So- Traditional Approaches to Evaluating Test Items

Will Martin
Principal Personnel Examiner,
NYS Department of Civil Service

The topics covered under this half-day workshop dealt with evaluating the effectiveness of multiple-choice items, short answer essay tests, and written work simulations.

Questions to consider as part of the evaluation include:

- What is the test designed to do and how do we know when it is doing its job?
- What does a good item analysis look like?
- Does using good items always generate good item analysis?

The workshop began with a

discussion of item analysis. This

is the first tool for evaluating written multiple-choice items. It answers the basic question: Is this item making a distinction among the test takers? Although the label may vary, the item analysis usually contains three main pieces of information:

1. Measure of difficulty (How easy or hard is the item?)
2. Measure of relationship (How does this item relate to the other items in the test?)
3. Comparison of high scoring verses low scoring test takers on each choice

Item analysis can be a very helpful tool in analyzing a written multiple-choice test. However, you must not rely solely on the numbers because they can deceive you. If the numbers are the only things you look at and the item is using the wrong key, then you may obtain a "good" item analysis result but the item would still be wrong. You must always look at the test item in conjunction with the item analysis. Test items do not exist in a vacuum.

Item analysis can also be used to evaluate whether or not test items have differences between the level of difficulty among gender and/or ethnic groups. Two methods were discussed:

1. Using differences in item mean item difficulties to compare and evaluate score differences between groups. This method may be thought

5 MAPAC NEWS

of as a “barometer” of

2. group differences. The larger the differences becomes, the greater the indication that you have real differences between the groups.
3. Using Chi Square to evaluate the differences between groups. This is a mathematical procedure that compares the differences between what we would expect if there were no differences and what is actually observed.

Both methods are useful but they operate from different assumptions. As with any technique there is no substitute for the good judgment of the analyst.

The remainder of the workshop was spent looking at item analysis procedures for short answer essay tests and written work simulations. Mr. Martin used a 15-item in-basket exercise to illustrate item analysis for short essays. The test consisted of 15 items requiring test takers to state how they would respond to each of the 15 in-basket stimuli. The responses were rated on a scale of 0 to 5 points. The number of points awarded to each response was based on the number of primary, secondary and inappropriate actions the test taker included in the response. The analysis presented the scores for each item in a tabular format in terms of the item's minimum and maximum score, mean, and standard deviation.

The last type of item analysis discussed was for a written work simulation. Results are presented in tabular or SPSS-type Crosstab analysis form showing the number and percent of test takers who selected each option in each section of the test.

Although simplistic in appearance, item analysis can be a powerful tool to assist us in our evaluation of a test item, whatever its format, and help us determine its relative contribution to making distinctions among test takers.

Submitted by:
James Frankart
Pennsylvania State Civil Service
Commission

Setting Passing Standards

Will Martin

Principal Personnel Examiner,
NYS Department of Civil Service

This half-day workshop dealt with how to decide where to draw the line between those who “pass” the test and those who do not. The setting of this point on the score distribution is based on the assumption that those who score above the passing standard can either perform or learn to perform the duties of the job while those who score below the passing standard cannot perform or learn to perform the job. For purposes of this workshop, passing standard and pass point are considered equivalent.

6 MAPAC NEWS

The first question we dealt with was why the passing standard is so important. From the human-interest side there are several

concerns. Have you ever failed an examination? Have you ever scored low on an examination and as a result not gotten a job or promotion? If you answered "yes," did you immediately jump to the conclusion that the examination fairly and accurately represented your ability to do the job? I think not. Like it or not, we are the gatekeepers. Pass point decisions we make have great impact on individuals. It can mean the difference between getting hired or remaining on the outside of the system.

From the professional point of view, we can look at what the professional guidelines and standards say about setting passing standards. The Uniform Guidelines on Employee Selection Procedures (1978) require that there must be a clear rationale for the selection of the passing standard and a justification of the procedures that were used to select the passing standard. If the passing standard has adverse impact against any race, sex, or ethnic group, the user must justify the procedures used to set the standard and do what it can to minimize the group differences.

There are two kinds of errors that occur when setting a pass point. Type I errors occur when you fail test takers who could successfully perform the job. When these errors occur, those who could do the job lose out. Type II errors occur when you pass persons who cannot perform the job. When Type II errors occur, the employer can lose out. The job of the selection

specialist is to minimize both

7 MAPAC NEWS

Type I and Type II errors. The process of setting pass points requires that you (the selection specialist) and Subject Matter Experts (SMEs) make a series of informed judgments. These judgments should be based on rationales that include such factors as the number of openings, the selection ratio, and the social policies of the organization.

We then looked at three methods that use SME judges to assist selection specialists in making pass point decisions. These are the Nedelsky, Angoff, and Ebel methods. These methods are attributable to the research of Leo Nedelsky in the 1950s and William Angoff and Robert Ebel in the 1970s. All three methods have six steps in common:

1. Select the judges – they must be qualified to make the decision about what constitutes a borderline test taker.
2. Define the borderline knowledge and skills needed for the job.
3. Train the judges.
4. Lead the judges through the judgment process.
5. Collect and evaluate the judgment data.
6. Combine the judgment data to identify the resultant pass point.

Our task was then to act as judges and use the three different methods. First we were given information about the KSAs of a “borderline” supervisor. Then we used each technique to evaluate a set of five items for each method.

In the Nedelsky method, we were asked to identify the answers that a borderline test taker would recognize as wrong. Then, using this information, we identified the expected score for each item. For example, if you expect the borderline test taker to eliminate two of the four answer choices, the expected difficulty (probability that borderline test takers will answer the item correctly) for that item would be .50 or 50%. The pass point is calculated by totaling the difficulties each judge expects for each item and then averaging that total across all the judges. The resulting mean would be the pass point. Care must be exercised in evaluating this number because any outlier scores could greatly affect this mean.

When using this method there are two major issues to consider:

1. Do you give the judges the key answers? Some say the judges should know the answers without having to see the key; others say the goal is to set a pass point, not test the judges’ knowledge of the field.
2. Should the judges perform this evaluation individually or as a group? Secondly, should you require the judges to reach consensus?

In making decisions on these issues you should consider how many judges you have and how many questions you have for the judges to review. Also, you must be concerned about managing one or more individuals who may

try to dominate the group of judges.

8 MAPAC NEWS

In the Angoff Method, the judges' task is to determine the probability that a borderline test taker will get the question correct. For each question, you ask each judge to directly estimate the probability that a borderline test taker will answer the question correctly, and then you sum the probabilities to determine the pass point. Many of the same issues, such as whether to give the judges the key answers and whether to force consensus, also apply to this method of setting a pass point.

The last method the workshop covered is the Ebel Method. This method requires the judges to make two decisions about each question:

1. What is the relevance of the question to the job? Judges indicate their decision by placing the question into one of four categories (i.e., essential, important, acceptable, or questionable).
2. What is the level of difficulty for the borderline test taker? Judges indicate their decision by placing the question into one of three categories (i.e., easy, medium or hard).

The selection specialist places the questions in a 3 by 4 table according to the categories the judges assigned. The judges then estimate the percentage of questions in each of the twelve cells that borderline test takers would answer correctly. The pass point is then arrived at by multiplying the number of questions in each cell by the estimated percentage correct, and summing these products.

Again, the same concerns that apply to the other methods also apply to the Ebel method.

In conclusion, as a selection specialist, you should keep the following items in mind when setting a passing point:

1. Be aware of the cognitive processes that judges engage in.
2. Provide judges with as much information and instructions as possible.
3. Try to make the judges comfortable with the process.
4. Evaluate the product as you move through the rating process.
5. Be willing to modify as the needs of the examination dictate.
6. Above all else, be rational and document your decisions.

Submitted by:
James Frankart
Pennsylvania State Civil Service
Commission

Meeting Presentations

Strategic Planning Session
Charley Sproule
Pennsylvania State Civil Service
Commission

Charley Sproule summarized the current MAPAC project on "Improving Supervisory and Management Assessment Procedures." The purpose of the project is to improve the variety and quality of supervisory

assessment procedures available to MAPAC members.

9 MAPAC NEWS

The project includes five initial activities:

1. collecting descriptive information on supervisory assessment procedures currently used by MAPAC members, and determining the supervisory assessment needs of members
2. identifying or developing taxonomies of supervisory work activities and job requirements
3. summarizing the literature on the relative validity and fairness of alternative assessment procedures for supervisory jobs
4. updating MAPAC procedures for test exchange
5. developing job analysis tools for use in studying supervisory jobs

Work groups have been formed to carry out this project, and work is underway on each of the five activities. The strategic planning session included a brief oral summary of the work done to-date in each of the five project activity areas. Future project activities will build on the results of the initial efforts for supervisory jobs and expand the scope of the project.

As of the date of the strategic planning session, one-half of MAPAC member agencies had responded to the "Supervisory Assessment Survey" and completed "Test Information Forms." These survey tools are being used to collect data on the current supervisory assessment practices and needs of MAPAC member agencies. To encourage

participation in the project, jurisdictions that respond to the

survey and meet other conditions can obtain the products of the previously completed MAPAC Test Item Resources Improvement Project.

For further information contact Charley Sproule of the Pennsylvania State Civil Service Commission (telephone 717 787-5974 extension 3533, email csroule@scsc.state.pa.us).

Submitted by:
Charles F. Sproule
Pennsylvania State Civil Service Commission

Implications of the Revised Testing Standards for Personnel Testing Practice
Wayne Camara, Ph.D.
Research and Development Director, The College Board

Dr. Camara reviewed major additions and changes in the revised Joint Standards for Educational and Psychological Tests and discussed the implications of the new Standards for personnel testing and assessment. At the time of this presentation to MAPAC (9/30/99), the final version of the new Standards had been endorsed by the three sponsoring organizations (AERA, APA, & NCME), but had not yet been published. The new Standards are expected to be available for purchase from the American Educational Research Association in late October 1999 (see www.aera.net) and (<http://www.apa.org/science/standards.html>). Following are some highlights of Dr. Camara's

10 MAPAC NEWS

presentation to MAPAC. More details are available in a handout that is available from Dr. Camara (email address: wcamara@collegeboard.org).

Previous Standards were published in 1955, 1966, 1974, and 1986. One major change in the 1999 Standards compared to most previous versions is that there is no categorization of standards by level (e.g., primary, secondary, etc.) in the 1999 version. Some 1999 standards have conditional statements attached to them, but most apply in all instances.

Structurally, the 1999 Standards are much more extensive than the previous versions. For example, there is a 47% increase in the number of standards (264 vs. 180) and a 33% increase in the number of paragraphs of text (240 vs. 180) in the 1999 Standards compared to the 1985 version. A new chapter was added to the 1999 Standards on the topic of test fairness, and policy issues are emphasized. The fairness chapter represents an area of large, substantive change. The new Standards have substantial sections on bilingual testing accommodation, testing of persons with disabilities, and other hot public policy topics. There is also more advocacy of positions on public policy issues.

There were 103 definitions added to the glossary of the 1999 Standards. The definitions added and deleted reflect trends in assessment during the past 15 years. Some definitions added include scoring rubrics, holistic scoring, defining

proficiency levels, matrix sampling, bias, meta-analysis, construct (defined very broadly), and differential item functioning (DIF).

The 1999 Standards provide criteria for the evaluation of tests, testing practices and the effects of test use, and they promote sound and ethical use of tests. They encourage adoption of the Standards by test developers, sponsors, and users. A test is defined broadly as an evaluation device or procedure in which a sample of behavior is evaluated and scored using a standardized process.

Some key provisions of selected chapters follows.

Chapter 1 – Validity -- describes validity as a unitary concept and refers to validation as an on-going process of accumulating evidence. Validity is the interpretation of scores by users, not the test itself. Validation is a joint responsibility of the test developer and user. The chapter emphasizes validity evidence based on the response process, validity generalization, and the consequences of testing. New requirements include requiring evidence that practice and coaching do not affect test scores, requiring evidence to support claims of the benefits of testing, and investigation of unintended consequences of test use. Some standards on criterion-related validation studies were deleted from this chapter compared to the 1985 Standards.

Chapter 2 -- Reliability and Measurement Error -- Dr. Camara did not discuss this chapter. But

11 MAPAC NEWS

he said it is the chapter that is most difficult to understand, and many I/O psychologists will not be familiar with the concepts it contains.

Chapter 3 -- Test Development -- provides a step-by-step description of test development and scoring. Considerable attention is given to performance assessment. It includes a requirement for sensitivity reviews of tests by a diverse panel. Test content should assure that inferences for all groups are valid. More documentation is required. Test users must train raters and ensure adequate reliability.

Chapter 4 -- Scales, Norms, and Score Comparability -- contains increased emphasis on linkages, adaptive testing, cut scores, and standard setting. One new requirement is to examine item context effects when changing item order in multiple forms of the same test. Regular re-examination of norms is required. Criterion referenced cut scores are endorsed.

Chapter 5 -- Test Administration, Scoring and Reporting -- states that persons of different backgrounds, ages or familiarity with testing may need nonstandard modes of test administration. This led to a discussion of a possible need to give diagnostic tests to determine if candidates are prepared to take the test. When test data are stored, administrators/users must preserve the test protocol.

Chapter 6 -- Supporting Documentation -- emphasizes clearly written and understandable materials for test users. It contains the "consequences" concept. That is, the higher the consequences of the test, the more the required analysis and documentation. Considerable detail is provided on what to document in test manuals.

Chapter 7 -- Fairness -- emphasizes the importance of fairness across all aspects of testing and provides a context for the standards. Nearly all of the standards in this chapter are new or substantially revised. Fairness is to be judged in the context of feasible test and non-test alternatives. The chapter contains four definitions of fairness; however, no one definition is endorsed. The chapter emphasizes sensitivity review panels, balanced content/rubrics, employing multiple measures, and recognition of the strengths and limits of tests. Studies of fairness are required when research shows DIF or effects of construct irrelevant variance. Studies are required to ensure that mean score differences do not result from construct underrepresentation or irrelevant variance. The chapter emphasizes the responsibility of informing policymakers of the likely consequences of using a test.

Chapter 8 -- Rights and Responsibilities of Test Takers -- Test takers are to be provided, in advance of testing, where

appropriate, information about the nature of the test, use of scores, and confidentiality. Test taker responsibilities are described.

Chapter 9 -- Testing Individuals of Diverse Linguistic Backgrounds -- Dr. Camara described this as the "worst chapter." The implication of the provisions of this chapter is that language problems are so enormous and prevalent that testing should sometimes not be conducted, and individual accommodations are needed depending on language dominance. Testing in a person's primary language is recommended, but the chapter identifies many problems and issues with translating and equating tests. The chapter requires collecting validity data for each linguistic subgroup and determining language proficiency before testing.

Chapter 10 -- Testing Individuals with Disabilities -- is much more instructional and descriptive than the 1985 Standards. It includes a provision on use of professional judgement in test modification, and a provision on use of multiple sources of information.

Chapter 14 -- Testing for Employment/Credentialing -- Dr. Camara described this chapter as "a good primer on employment testing" and the best chapter in the new Standards. The chapter puts the burden for conducting research on sample size rather than organizational size. It contains an informative graphic and

12 MAPAC NEWS

narrative model of the validation process. Some additions to these standards compared to the 1985 version include: stating the objective of testing, stating that cut scores should not be regulated by a desire to reduce the number of persons passing, and defining the role of each test when a composite predictor is used. Local validation studies are dependent on their feasibility, and existing validation evidence should be considered when interpreting the results of local studies.

The 1999 Standards include 264 separate standards. Dr. Camara advised that, for documentation purposes, anyone involved in “high stakes testing” should list each standard and comment on how their test development and research meets each standard, or why meeting the standard is not relevant or feasible for the particular assessment process.

Submitted by:
Charles F. Sproule
Pennsylvania State Civil Service
Commission

[Editor’s Note: Lance Seberhagen contributed the following update on how to order the new Standards: price is \$31.95 (for nonmembers of AERA/NCME) or \$25.95 (for members), plus shipping & handling. Call (800) 628-4094 to order your copy. Also, congratulations to Dr. Camara, who was elected as president-elect of APA Div. 5 (Evaluation, Measurement, and Statistics) for 1999-2000.]

Planning and Correcting for Artifacts in Criterion Related Validity Studies

Richard R. Reilly, Ph.D.
Howe School of Technology
Management, Stevens Institute
of Technology

Validity

Validity is the extent to which a selection procedure can be used to make accurate predictions of future job-related employee behaviors. Criterion-related validity is a means to quantify this accuracy by correlating selection procedure scores with job performance measure scores.

The goal of criterion-related validity is to estimate the true validity of the predictor (selection device) for the applicant population. However, the goal may be difficult to achieve because of practical, or methodological constraints.

In an ideal validation study:

All applicants who are tested are hired and placed into the job. Job performance is measured with perfect reliability. The study sample is infinitely large.

However, a typical validation study has these shortcomings:

Only the highest scoring applicants are hired and placed into the job (range restriction). Job performance is measured with some unreliability (criterion unreliability). Samples are small.

Dr. Reilly presented procedures for correcting obtained validity coefficients with respect to these three shortcomings. He began by reviewing three predictive validity designs:

Full-Range Predictive Validation – All applicants are tested and then placed on the job. This design is seldom used because organizations are unwilling to put all applicants on the job, and turnover can change the sample in non-random ways before job performance is measured.

Predictive Validation with Explicit Selection – Applicants are hired based on the selection procedure (applicants must score above the cut off point, and there is top-down selection). This explicit selection results in a sample whose standard deviation that is smaller than the applicant population’s standard deviation. This “restriction-in-range” produces a biased (lower) estimate of validity.

Predictive Validation With Incidental Selection – Applicants are hired based on some other variable (e.g., an interview) that is correlated with the selection device being studied (e.g., a test). This approach results in *incidental* restriction-in-range, and also produces a biased (lower) estimate of validity.

Dr. Reilly then described concurrent validity designs. In these designs, both the selection device scores and the job performance data are collected from employees at a single point in time. These designs may use a mix of explicit and incidental

13 MAPAC NEWS

selection, and like the second two predictive designs described above, result in estimates of validity that are biased (lower).

The Artifact Corrections

Range Restriction

Correction for range restriction is based on these assumptions:

The population relationship between x (selection device) and y (job performance criterion measure) can be described by a linear (straight line) function. The linear function has the same slope for both the restricted (obtained sample) and the unrestricted (theoretical "everyone-is-hired") groups.

The population distribution of x and y is bivariate normal.

The standard deviation of y (job performance criterion measure) around x (selection device) for any point on x is the same. This is the concept of homoscedasticity. In other words, the standard error of estimate (S_{est}), which is the standard deviation of y around x, is constant across the range of x.

The computational formula for the S_{est} is:

$$S_{est} = s_y \sqrt{1 - r_{xy}^2}$$

S_{est} is same throughout the distribution of x and does not change with selection on x. If we know the population (unselected) S_y , we can estimate a corrected correlation coefficient (R_{xy}).

Homoscedasticity allows us to say that the S_{est} for the selected sample is the same as that for the unselected population.

Expressed as a formula, we have:

$$s_y \sqrt{1 - r_{xy}^2} = S_y \sqrt{1 - R_{xy}^2}$$

where the lower case letters refer to the selected sample values and the capital letters refer to population values.

The linearity assumption means that the same regression line will be the correct line regardless of the degree of selection. Using the same convention as above for distinguishing between sample and population values, we have, for the regression coefficient:

$$r_{xy} s_y / s_x = R_{xy} S_y / S_x$$

In addition, for simplicity in the final formula, we set:

$u = s_x / S_x$ (Note: The denominator uses the capital letter.)

Combining the equations for S_{est} and the regression coefficient, substituting u for s_x / S_x and solving for R^2 results in the correction formula:

$$R_{xy} = r_{xy} / (\sqrt{u^2 + (1 - u^2)r_{xy}})$$

Correcting for Unreliability in the Criterion

Dr. Reilly pointed out that:

All criteria have some degree of unreliability.

Unreliability means that some of

the variation in criterion is error variance.

Error variance cannot, by

14 MAPAC NEWS

definition, be predicted and therefore, unreliability will lower the validity.

Unreliability also exists in the predictor but is not usually corrected for.

Dr. Reilly suggested using coefficient alpha to compute the criterion's reliability. Coefficient alpha is appropriate to use for measures that are made up of independent subparts, and it can be computed from a single administration of the measure. The subparts of the measure could be ratings supplied by multiple job performance raters or multiple items on a job knowledge or work sample test. Coefficient alpha is also convenient to use because software packages (e.g., SPSS) will compute it routinely. The formula for coefficient alpha is:

$$\alpha = 1 - \sum s_i^2 / S_t^2$$

where s_i is the subpart standard deviation and S_t is the standard deviation for the total measure. In cases where you do not have the data necessary to compute criterion reliability, you may use an estimate from the literature. Rothstein (Rothstein, H.R. (1990). Inter-rater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. Journal of Applied Psychology, 75, 322-327.) is one source for an estimate. Rothstein found an asymptotic reliability for supervisory ratings of .55. Using .6 is a conservative approach (i.e., it will undercorrect).

The formula for the correction is:

$$r_{xyc} = r_{xyo} / \sqrt{R_{yy}}$$

where r_{xyc} is the corrected validity coefficient, r_{xyo} is the obtained validity coefficient and R_{yy} is the reliability of the criterion.

The proper order in which to do the corrections is to correct for criterion unreliability, then correct for range restriction.

Sample Size

In addition, Dr. Reilly reviewed the implications of small sample size:

Sampling error is likely to be greater for small samples. Power is reduced (i.e., the probability of finding significant results when the results are truly significant is reduced).

When the true validity is relatively small, a larger sample size will be needed in order to detect any validity at all.

Conclusion

Dr. Reilly completed his presentation by discussing the implications for planning criterion-related validity studies. He pointed out that range restriction, criterion reliability and sample size can be addressed in various ways:

Range Restriction

Keep all applicant test data. This allows for the calculation of statistics necessary to correct for range-restriction.

Use range restriction procedures

15 MAPAC NEWS

to correct for incidental or explicit selection.

Criterion Reliability

Take steps to improve reliability of criterion by carefully training raters and building sound work sample or job knowledge tests to measure performance.

If possible, estimate the reliability of the criterion empirically.

If the data to calculate criterion reliability are not available, use an assumed value of .6 (when the criterion is supervisors' ratings).

Sample size

Review research (especially meta-analytic) to get an estimate of likely true validity.

Plan for a sample large enough to have sufficient power.

Use tables (e.g., Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York, New York: Academic Press.) to estimate power.

Submitted by:

Will Martin

NYS Department of Civil Service

Title VII Cases: Adverse Impact and the Shifting Burden of Proof

Dr. Joel Lefkowitz, The City University of New York

After a morning of discussing the new joint Standards for Educational and Psychological Testing, the possibility of Title VII proceedings may have weighed heavily on the minds of MAPAC

members. Dr. Joel Lefkowitz, an expert with considerable experience testifying in court, was on hand to provide guidance for preparing to defend a Title VII charge alleging adverse impact.

A few people raised hands and smiled knowingly when Dr. Lefkowitz asked if anyone had been involved in a Title VII proceeding, and many laughed when he asked if the experience was a valuable personal and professional growth opportunity. Everyone agrees that Title VII proceedings are best avoided in the first place. Toward this end, Dr. Lefkowitz had a few recommendations.

Work collaboratively with others on test development. Working in isolation is dangerous. Have the benefit of a devil's advocate on your planning team.

Use an educated appeal to supervisors or other management when they are less than enthusiastic about careful test development. Experts in assessment have the responsibility and right to educate others on professional standards for testing.

Assure that there is justification for what needs to be justified.

Develop quality tests.

When asked about the usefulness of criterion related validation studies, Dr. Lefkowitz admitted that in real life, much of the outcome of court cases depends on the individual judge, expert witness testimony, and the lawyers involved in the case.

The bad news is that these factors are

16 MAPAC NEWS

generally uncontrollable. The good news is that the quality of the test also has an effect on the outcome.

If a test is challenged, Dr. Lefkowitz recommends preparing testimony that is easily understood by a jury and having good representation. The communication skills of lawyers and expert witnesses are also critical.

Dr. Lefkowitz discussed briefly history and trends of Title VII cases. During the Reagan era, Title VII proceedings focused around sex and age discrimination cases. He sees future challenges focussing on how much of the job domain is tested and how non-cognitive factors are tested. Preliminary research has shown that adding non-cognitive factors does not reduce adverse impact.

The group discussed the development of alternate kinds of tests that might achieve the same validity with less adverse impact. (See below reference to Nassau County police examination.) The manipulation of statistical results is reason to scrutinize validation studies carefully.

Consent decrees resulting from court cases were also discussed. For some purposes the decrees are problematic. Early decrees appear to have been attorney driven -- instead of built upon the needs of the agencies and test taker populations that must live by them. And because change to the decrees may require action by both sides, it is difficult to amend them.

Also, some of the decrees that require differential scoring for race or other subgroups are contrary to standards which prohibit using cut-off scores as a mechanism for controlling the number of candidates reachable for employment.

On the positive side, the decrees suggest that some agreement had occurred among the parties.

Dr. Lefkowitz also acknowledged that decrees sometimes allow innovative proposals, not normally acceptable, to be implemented.

During the session, members shared and discussed current problems involving difficult testing situations. For example, some agencies are questioning the legality, practicability, and desirability for testing English language comprehension and speaking capability in a candidate pool with English as a second language, especially when the candidates are otherwise well qualified for the test. Unfortunately, new Standards provide little practical guidance in this area. The Standards also fall short in practical guidance for standardizing testing of populations with other special needs.

No one knows for sure how the new Standards will effect Title VII actions in the future; however, it will be interesting to see how the changes affect the jobs of assessment professionals. Continuing research, ongoing discussion and close attention to court cases will guide MAPAC

members.

17 MAPAC NEWS

Read about the Nassau County Police Study on the IPMAAC website-- www.ipmaac.org.

Submitted by:
Christine Whiteley
NYS Department of Civil Service

Selection Instruments: Which Predict Best?

Harold W. Goldstein
Baruch College, The City University of New York

The short answer to the question is that cognitive ability measures work very well. But before getting into the details of predictor quality, Goldstein asked and answered a more fundamental question: Why care about selection and selection instruments? The answer to that question includes a list of reasons:

- Organizations are competing in the war to gain a competitive advantage in the global economy.
- As with any war, the combatants follow a plan. One of the key components of a winning plan is a strategy for staffing the organization.
- Organizations are competing for the best talent. They need good selection tools to win that competition.
- An organization's employees make (or break) the place. Employees use both their technical skills and their personalities as they operate in their organizations.
- Good selection instruments impact the utility of the entire employee selection process. Organizations are looking at the cost/benefits of the

selection systems, including measures of cognitive and non-cognitive factors.

Goldstein used Schmidt and Hunter's validity generalization data (Schmidt, F., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262-274.) to show how well various kinds of measures fare as performance predictors:

Predictor	Validity (r)
Work Sample	.54
General Cognitive Ability	.51
Interview (Structured)	.51
Peer Ratings	.49
Job Knowledge	.48
T&E Behavioral Consistency	.45
Integrity Test	.41
Interview (Unstructured)	.38
Assessment Center	.37
Biographical Data	.35
Conscientiousness	.31

Selection systems using more than one measure can do better, as more Schmidt and Hunter (1998) data illustrate. When a measure of general cognitive ability is combined with another measure, the resulting validity coefficient (Multiple R in the table below) is greater.

Predictor Combined with Cognitive Ability Measure	Multiple R	% Gain in Validity
Integrity Test	.65	27
Work Sample	.63	24
Interview (Structured)	.63	24
Conscientiousness	.60	18
Assessment Center	.53	4
Biographical Data	.52	2

There has been increasing investigation of the utility of personality and interest measures as research focuses on assessing prospective employees' ability to contribute to work-group and organizational performance. Another reason for increased interest in measures of the non-cognitive side of performance is the continuing search for measures that have less adverse impact.

18 MAPAC NEWS

In general, biographical data, work samples and structured interviews have less adverse impact and they have high validity and utility, especially when combined with other measures.

Submitted by:
Susan M. Reilly
U.S. Immigration and
Naturalization Service

Employment Law Update

Maureen E. Reilly, Esq., Ph.D.
Morgan, Lewis & Bockius LLP

Dr. Reilly's presentation outlined various cases in Labor and Employment Law, updates in Employment Testing, recent activity by the Department of Labor and other related issues of interest over the past year. More specifically the cases outlined primarily pertained to recent developments related to the Americans with Disabilities Act and cases involving discrimination.

The case of Sutton v. United Airlines, Inc., No. 97-1943 involved twin sisters with severe myopia. They filed a claim against United Airlines claiming that they were being discriminated against because they did not have 20/20 vision. The pilot's examination they applied for required 20/20 vision. There was controversy among the courts with respect to mitigating circumstances. The Court rejected the EEOC's interpretation and held that the use of corrective or mitigating measures should be

considered in determining whether an individual

is disabled under the ADA. It was also determined that it is acceptable for employers to make preferences in selective criteria. Other cases discussed were Murphy v. United Parcel Service, Inc., No. 97-1992, and Albertsons, Inc. v. Kirkingburg, No. 98-591.

Cases involving learning disabilities/inabilities were also discussed. In Bartlett v. New York State Board of Law Examiners, 156 F.3d 321 (2d Cir. 1998), vacated and remanded, 119 S.Ct. 2388 (June 24, 1999), a woman who has a Ph.D. and attended law school failed the Bar five times. She asked for an accommodation and was denied. The Board's expert concluded that Bartlett was not learning disabled. The District Court found that Bartlett is substantially limited in major life activity of working, but not reading or learning. The Second Circuit concluded that Bartlett is disabled, but her substantial limitation is in the major life activity of reading or learning (not working). Also discussed was Mondzelewski v. Pathmark Stores, Inc., 162 F.3d 778 (3d Cir. 1998), where it was concluded that an individual's training, skills and abilities must be considered in determining whether the individual is substantially limited in the major life activity of working.

The following guidelines issued by EEOC in March 1999 were highlighted:

- Reasonable Accommodation and Job Applicants
- No pre-offer inquiries

19 MAPAC NEWS

- regarding need for job accommodations, UNLESS
 - Disability is obvious or applicant has voluntarily disclosed disability, AND
 - Employer reasonably believes applicant will need accommodation
- Employer must provide accommodation in the selection process, even if it believes that no accommodation is possible on the job.

Some cases discussed in employment testing include:

- Hayden v. County of Nassau, 180 F.3d 42 (2d Cir. June 9, 1999)
- Lanning v. Southeastern Pennsylvania Transportation Authority (SEPTA), __ F.3d __, 1999 WL 432595 (3d Cir. June 29, 1999)
- Jordan v. New London, Conn., No. 3:97CV1012 (PCD) (D.Conn. Sept. 2, 1999)
- Mitchell v. Utah State Tax Comm'n, 26 F. Supp. 1321 (D.Utah 1998)
- Allen v. Entergy Corp., 181 F.3d 902 (8th Cir. June 22, 1999)

These highlights of a Memorandum of Understanding issued by EEOC April, 1999 were discussed:

- OFCCP is authorized to act as EEOC's agent for Title VII component of dual-filed complaints.
- OFCCP may obtain compensatory and punitive damages in resolution of the

- charge (in addition to back pay).
- EEOC may retain claims and OFCCP will give EEOC information obtained in its investigation.
- Information obtained independently by the OFCCP is not subject to confidentiality provisions that govern EEOC investigations.
- OFCCP has significantly greater power to obtain employers' documents.

Submitted by:
Miriam M. Stokes
Maryland Department of
Transportation
Motor Vehicle Administration

What Does the Literature Say?

One of the standard setting methods that Will Martin covered in his afternoon workshop was the Angoff procedure. Laura Goodwin has recently reported on some results from the use of the Angoff procedure where judges' ratings were used to set the passing standard for a nation-wide professional certification examination.

Goodwin, L. D. (1999). Relations between observed item difficulty levels and Angoff minimum passing levels for a group of borderline examinees. Applied Measurement in Education, 12 (1), 13-28.

The Angoff method for setting passing standards uses judges' ratings of the proportion of minimally adequate test takers who should get each item correct. These ratings are then averaged across the judges and summed. The passing standard is set at this sum.

One of the criticisms of this method is that it depends on the ability of the judges to accurately conceptualize a minimally competent person and then, for each question separately, infer the probability that such a person will be able to answer the question correctly. To provide some hard data on this issue, Goodwin quantified judges' accuracy in a study of the Angoff standard setting procedure used for a high-stakes certification examination. Goodwin compared the judges' minimal passing level ratings (i.e., the proportion of minimally adequate test takers who will answer correctly) on the items to subsequent actual item difficulties (% answering the item correctly). These minimal passing levels (MPLs) were compared to item difficulties computed on all examinees and on the subgroup of examinees who were minimally competent (operationally defined as scoring between one standard error of measurement (SEM) below and one SEM above the passing score). Actual item difficulties and MPLs correlated .55 for both groups (all examinees and the minimally competent). The average difference between actual difficulties and MPLs for the 140 test items was -.03 for the minimally competent group.

20 MAPAC NEWS

Goodwin also categorized the items into three groups based on the difference between MPLs and actual item difficulties:

1. over-predicted -- MPL was more than .10 above the actual difficulty
2. accurate -- MPL was within .10 of the actual difficulty
3. under-predicted -- MPL was more than .10 below the actual difficulty

For the minimally competent group, 27.9% of the item difficulties were over-predicted, 61.4% were accurately predicted, and 10.7% were under-predicted.

Goodwin's work confirms results from previous studies – that judges tend to over-estimate the abilities of minimally adequate performers. This is understandable because the judges are chosen because they are experts in the field, and experts generally have high expectations for performance. These findings do not support the contentions of some that the Angoff standard setting procedure asks judges to do the impossible. After all, they were pretty accurate almost two-thirds of the time. Goodwin ends by pointing out what other studies of standard setting procedures have also said – that the accuracy of MPLs can be improved through careful training of the judges.

Upcoming Events...

**By Lance W. Seberhagen
Seberhagen and Associates
(sebe@erols.com)**

1999

Dec 8-10
Institute for Professional Education.
Seminar. "Linear and Nonlinear
Regression with Applications."
Washington, DC. Contact: IPE, (703)
527-8700.

Dec 9-10
University of Maryland, Department of
Measurement, Statistics, and
Evaluation. Short course. "Logistic
Regression Analysis." College Park,
MD. Contact: Dr. Greg Hancock, (301)
405-3624.

Dec 16
Metro NY Association for Applied
Psychology. Dinner Meeting. Dr. Ken
Pearlman, Lucent Technologies. "The
Light and Dark Sides of Competency
Modeling." New York, NY. Contact:
Metro, (212) 539-7593.

2000

Jan 11
Metro NY Association for Applied
Psychology. Dinner Meeting. Dr.
Richard Hackman, Harvard University.
"Leading Teams." New York, NY.
Contact: Metro, (212) 539-7593.

Jan 12
PTC/MW. LUNCHEON MEETING.
Speaker to be announced. Pier 7
Restaurant, Washington, DC. Contact:
Dr.
Lance Anderson, Caliber Associates,
Fairfax, VA, (703) 385-3200.

Jan 24-25
University of Michigan, Joint Program in
Survey Methodology. Short Course.
"Introduction to Survey Estimation."
College Park, MD. Contact: JPMS,
(800) 937-9320.

Jan 27-28
University of Maryland, Department of
Measurement, Statistics, and
Evaluation. Short course. "Introduction
to Multilevel Analysis Methods,
Hierarchical Linear Models." College
Park, MD. Contact: Dr. Greg Hancock,
(301) 405-3624.

Feb 9
PTC/MW. LUNCHEON MEETING.
Speaker to be announced. Pier 7
Restaurant, Washington, DC. Contact:
Dr. Lance Anderson, Caliber
Associates, Fairfax, VA, (703) 385-
3200.

Feb 17
Metro NY Association for Applied
Psychology. Dinner Meeting. Dr. Bill
Byham, Development Dimensions
International. "The Millennium
Elephant and Other Challenges Facing
the Industrial-Organizational
Psychologist in the Next Century." New
York, NY. Contact: Metro, (212) 539-
7593.

Feb 22-23
University of Michigan, Joint Program in
Survey Methodology. Short Course.
"Survey Management." Bethesda, MD.
Contact: JPMS, (800) 937-9320.
Mar 3-5
I/O & OB Graduate Student
Conference. Chattanooga, TN.
Contact: James LeBreton,
jlebreto@utk.edu.

Mar 6-8
Society for HR Management.
Employment Law & Legislative
Conference. Washington, DC.
Contact: SHRM, (703) 548-3440.

Mar 8-12
Academy of HR Development.
Conference. Raleigh-Durham, NC.
Contact: AHRD, (504) 334-1874.

Mar 9
Metro NY Association for Applied
Psychology. Dinner Meeting. Dr.
Virginia Schein, Gettysburg, PA.
"Understanding Poor Women in Work:
Opportunities for Industrial and
Organizational Psychologists." New
York, NY. Contact: Metro, (212) 539-
7593.

Mar 13-15

21 MAPAC NEWS

Institute for Professional Education.
Seminar. "Applied Multivariate Methods
Using Popular Statistical Computing

Packages." Washington, DC. Contact:
IPE, (703) 527-8700.

Mar 13-15
Institute for Professional Education.
Seminar. "Linear and Nonlinear
Regression with Applications."
Washington, DC. Contact: IPE, (703)
527-8700.

Mar 14-15
University of Michigan, Joint Program in
Survey Methodology. Short Course.
"Tailored Design of Mail, Internet, &
Other Self-Administered Surveys."
Bethesda, MD. Contact: JPSM, (800)
937-9320.

Mar 27-29
Institute for Professional Education.
Seminar. "Applied Statistics."
Washington, DC. Contact: IPE, (703)
527-8700.

Winter Meeting 2000

The Winter MAPAC meeting will
held in Baltimore on January 19-
21, 2000. The meeting location
is the Hilton Baltimore & Towers,
20 West Baltimore Street,
Baltimore, MD 21201. Nancy
Abrams will present the training
workshop (Wednesday, January
19) on developing training and
experience tests. Thursday and
Friday presentations will cover:

- Minimum qualifications
- Personality assessment
- The Department of Labor's
new national work
description and
classifications system
- Migration of technical personnel
into management
- The meaning of "business
necessity" with respect to the
Civil Rights Act of 1991

MAPAC Committee Chairs

Program:
Elliot D. Lasson, Ph.D.
MD Dept. of Budget and Mangmt.
301 W. Preston St.
Baltimore, MD 21201
(410) 767-4928
(410) 333-7674 (fax)
ELASSON@dbm.state.md.us

Training:
Robyn Talesnik
State Highway Administration
Recruitment and Examination
Division Room 602
Baltimore, MD 21203-0717
(410) 545-5630
rtalesnik@sha.state.md.us

Membership:
James E. Frankart
Pennsylvania State Civil Service
Commission
320 Market Street
Harrisburg, PA 17108-0569
(717) 783-2697, Ext. 3517
(717) 783-0838 (FAX)
JFrankart@scsc.state.pa.us

Note from the Editor

This is my last issue of the
NEWS. My time as one of New
York State's representatives to
MAPAC has come to a close, so
I am stepping down as NEWS
editor. I heartily thank all the
contributors who have provided
the summaries of the meeting
presentations over the last three
years. I could not have put the
NEWS together without you.
Thank you also to Jill Guice,
who, as my fellow editor for two
years, polished the NEWS,
formatted it and arranged for its
distribution.

I was overwhelmed by the
thanks that MAPAC gave me at

22 MAPAC NEWS

the September meeting. I appreciate

having had the opportunity to actively participate in MAPAC and to meet so many great people.

Betsy Kaido

MAPAC Members

Cities

Baltimore, MD
New York, NY
Philadelphia, PA
Rochester, NY

Counties

Baltimore, MD
Prince George's, MD
Fairfax, VA

States

Delaware
Maryland
New Jersey
New York
Ohio
Pennsylvania

Other

City University of New York
Immigration and Naturalization Service
Maryland Department of Transportation
Metropolitan Washington Airports Authority
New York State Unified Court System
New York City Transit Authority
Port Authority of N.Y. & N.J.
School District of Philadelphia
U.S. Postal Service

If you are interested in becoming a member of MAPAC, please contact our Membership Committee Chair, Jim Frankart, for more information!

