

MEASUREMENT FORUM

Volume 4 Number 1
Winter/Spring, 1999

A Newsletter for the Promotion of Excellence
in Measurement Instruction

www.mayfieldpub.com/psychtesting

Prepared for use with *Psychological Testing and Assessment*, Fourth Edition,
by Ronald Jay Cohen, and Mark E. Swerdlik.

Welcome to the fourth year of the publication of *Measurement Forum*. This issue (Volume 4 Number 1) of *Measurement Forum* coincides with the publication of the Fourth Edition, of *Psychological Testing and Assessment*. In addition to the Fourth Edition of the text, a revised edition of the *Instructor's Manual and Resource Guide*, Test Bank (available in a computerized or hard copy format), and our student workbook, *Exercises in Psychological Testing*, have also been published and are available for your review. If you have not received a review copy of these revised publications, please contact Mayfield at 800/433-1279 or jbauer@mayfieldpub.com. Although this newsletter is designed for use with the text *Psychological Testing and Assessment*, our main objective is to provide information that will promote excellence in measurement instruction.

Also with the issue, we announce the availability of online support through our Web page at www.mayfieldpub.com/psychtesting. The Web site is available to users of our textbook as well as anyone else interested in accessing

the information that can only be found there. Be sure to visit our Web site frequently for updates and additions to this unique teaching and learning resource.

In addition to current and past issues of *Measurement Forum*, the full text of our Test Developer Profiles (TDPs) is currently available at this site. We asked a number of test developers to prepare candid essays about the various aspects of the test development process, including the joys and challenges. By using these TDPs you will provide your students with a glimpse into the real world of testing and assessment. Many contemporary test developers, as well as influential figures of the past, are highlighted.

This issue of *Measurement Forum* contains a number of classroom demonstrations/activities submitted by our readers as well as other items which we hope will assist you in reaching your instructional objectives for your measurement courses. Please consider contributing **your** ideas for classroom demonstrations for our next issue. We look forward to hearing from you.

Best wishes in 1999!

Mark E. Swerdlik
Ronald Jay Cohen
Editors

EDITORS:

Mark E. Swerdlik, Illinois State University
Ronald Jay Cohen, St. John's University

Please address all correspondence to:

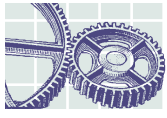
Mark E. Swerdlik
Department of Psychology
Campus Box 4620
Illinois State University
Normal, IL 61790-4620
(309) 438-5720
fax: (309) 438-5789
e-mail: meswerd@ilstu.edu

For additional copies of the *Measurement Forum* or for more information about the new fourth edition of *Psychological Testing and Assessment*, contact Mayfield Publishing Company at (800) 433-1279 or jbauer@mayfieldpub.com. Current and past issues of *Measurement Forum* are also available at www.mayfieldpub.com/psychtesting.



MAYFIELD PUBLISHING COMPANY

1280 Villa Street
Mountain View, CA 94041
800-433-1279



CONNECTIONS

In response to our last issue of *Measurement Forum*, we received the following communications:

Department of Educational and School Psychology
Indiana University of Pennsylvania
246 Stouffer Hall
1175 Maple Street
Indiana, Pennsylvania 15705-1087
(412) 357-2316
(412) 357-6946 (FAX)

March 22, 1998

Dr. Mark Swerdlik
Department of Psychology
Campus Box 4620
Illinois State University
Normal, IL 61790-4620

Dear Dr. Swerdlik:

I saw your note about the newsletter related to introductory courses in educational tests and measurement which appeared in the Trainers Forum. I have been teaching such a course for the last three years and would like to receive the newsletter. With one of my classes meeting for three hours on a Wednesday night, I need all the new ideas I can get! I have also enclosed a brief piece about activities I have used in class. Perhaps you would like to use it in the newsletter.

Sincerely,

Victoria B. Damiani, Ed.D., NCSP
Assistant Professor of Educational and
School Psychology

Please let us hear from you. Your correspondence expressing an opinion, a teaching tip, and/or an article would be most appreciated.



CLASSROOM DEMONSTRATIONS

*Integration with
Text: Chapter 7*

TEACHING THROUGH TEST DEVELOPMENT

Suzanne M. Phillips Gordon College
(sPhillips@gordon.edu)

An effective way to teach students about test development is to develop a measure with them. This article provides an overview of what a test development project might involve, structured around the five stages of test development provided in Chapter 7, with two tryout-analysis-revision cycles. I am eager to hear the comments and reflections of professors who use test development projects in their courses, and of those interested in doing so.

I have worked on test development projects with undergraduates in measurement courses for five years. These courses involve 12 to 25 students. Prior to the measurement course, students have taken statistics and research methods; most have conducted research projects and understand research ethics. Students work on this project in groups with 3 to 12 members. Ideally, groups have 5 or 6 members: small enough that people feel accountable and large enough for members to complete tedious work efficiently. Group members keep individual logs of the tasks they complete for the project and the time they spend.

Groups form during the first three weeks of the semester around interest areas that students develop through brainstorming. Students brainstorm individually for five minutes, writing down their ideas without editing or being critical. Then they share ideas with the group, free associating to others' ideas as they are heard. This process produces many more good ideas that we can use in a single class. (Some favorites include tests of spatial orientation, "perfect pitch," memory for names, extraversion, the ability to do several tasks at once, self-esteem, personal organization, coordination, and a sense of time.) Students then indicate their top three choices anonymously, and a list of the most popular options is created.

Students form groups around particular items on that list, select a group organizer, and move into the test development process. I provide them with an empty binder to keep in an accessible place (in the laboratory, outside my office), which they use as an archive throughout the semester and as a resource when writing the test manual at the end of the semester.

Test conceptualization: Students who have never developed their own test often describe this stage as "getting an idea for a test." Once they form groups, they are ready to move on to test construction. Their first surprise in this project is the amount of work that remains at this initial stage. I provide students with a worksheet based on the questions in the "Test Conceptualization" section of Chapter 7, including items about the intended audience, the general format of the items,

the desired breadth (or narrowness) of the content area, and the need for the test. Students can learn about existing measures on the Buros' Web site (www.unl.edu/buros) and through the *Directory of Unpublished Experimental Mental Measures*, now published by American Psychological Association. Students are often disappointed (and ready to change topics) when they learn that their ideas are not unique; they benefit from hearing that no existing measure is perfect, and that new measures are commonly developed even when other measures exist—the intelligence tests discussed in Chapter 9 are a good example. With the information they gather and from their own conversations, students complete the test conceptualization worksheet and make it the first entry in their binder, along with printouts from Web searches and photocopies of other materials.

Test construction: Students should move into this stage by the 5th week of the semester. Again, students use a worksheet highlighting the issues covered in Chapter 7, including scaling of items, item format, and scoring procedures. The difficulty of this stage varies with the item format. Students interested in developing a group-administered measure of face and name recognition may spend a couple of weeks selecting photographs and creating a videotape or slide-show that presents test stimuli appropriately. Those developing a measure of extraversion may spend just a few hours writing a list of adjectives for Likert scaling. In any case, students consistently struggle with three issues at this stage: (1) they are reluctant to develop many extra items, because creating items is difficult work; (2) they are reluctant to think about a process for establishing validity, wanting to put that off until after the first tryout, not realizing how frustrating it is to revise a test without some validity measure; and (3) they are not critical of their work: items seem perfect when they are first created. This last issue is addressed if students do an informal “pre-tryout,” in which friends outside the class (or classmates in other groups if it is a large class) take the test and provide feedback. A most helpful form of feedback comes from the “think-aloud” procedure described in Chapter 7, in which subjects articulate their thought process as they take the test. Throughout the test construction stage, the professor serves in the unpopular but necessary role of providing foresight, telling students that extra work will benefit them later.

Tryout: Students should do their first “trial run” with naive subjects by the 8th week of the semester. Somehow, this seems early to students. I have found it helpful around this time of the semester to ask students to create a schedule for the remainder of the project, with target dates for completion of the various steps, which helps them see the need to move ahead.

In addition to the measure that they have created, students need consent and debriefing forms and a validity measure. Clean copies of all of these forms, clearly marked as “Tryout #1,” are entered into the binder. The IRB review can be expedited if participating in the tryout places subjects

at a minimal risk, which is often so. Generally, students gather data through the Introduction to Psychology subject pool. Students persuaded of the value of factor analysis may want to run enough subjects to complete one, but I have never worked at schools with large enough subject pools to permit this. Instead, students run thirty to fifty subjects and “make do” with coefficient alpha and inter-item correlations in place of factor analysis.

Groups wishing to gather data outside the subject pool might consider these options: if they can identify two groups that should differ on the measure (athletes and non-athletes on a coordination measure, or first year students and residence staff on a measure of social problem solving), they can assess the construct validity of the measure. If the measure is short, is not socially sensitive, and does not require controlled conditions for administration, students in a large group may gather a lot of data by each administering the measure to several people they know, though special care must be taken to preserve anonymity and address other ethics issues.

Analysis and Revision. For students who do not yet understand reliability and validity coefficients, the analysis stage is a great concrete learning experience. Because of the many analyses that need to be done, students use a computerized statistical program, like SPSS. This frees students to explore ideas (sex differences, creating a subscale out of particular items) without worrying about the effort it would take to execute those ideas by hand. Analyses are printed out for the binder and labelled clearly as coming from “Tryout #1.”

With the appropriate data, students can explore item difficulty (or variability), item validity, item reliability, and item discrimination indices. This works well when the entire group is involved (with one or two people “specializing” in each statistical index), evaluating each item with all available information. This allows students to share expertise and keeps them from feeling overwhelmed. Often students create a chart to record observations. Hearing students argue over whether a highly reliable item (which isn't valid) or a valid item (with modest reliability) is most valuable to the test is one of the professor's rewards in this project.

The revised measure (a clean copy of which is entered in the binder) is administered to a new group of subjects, preferably by the 11th week of the semester. Materials are marked for the binder as “Tryout #2.” The resulting data are analyzed as described above; students are eager to compare the new test and the original to observe improvements, providing opportunities to learn about validity shrinkage (Chapter 6) and about the relationship between test length and reliability (Chapter 5). Students may or may not complete a second revision.

At the end of the semester, students submit their work logs, the binder, and a test manual. The test manual may be ten to thirty pages long, depending upon the size of the group and the complexity of the project. Students examine existing

test manuals for ideas and organizational strategies. They may create norms from the data gathered in the second tryout.

Students are never satisfied with the most recent version of their test, and often include suggestions for additional revisions as a last chapter in the manual. Despite this dissatisfaction, students feel a great sense of accomplishment as they complete this project. They learn a lot about test development, as well as about working together on a large project. There is usually a discussion of additional work on the project after the end of the semester, reflecting students' investment, and students toy with the idea of developing the measure for publication.

CLASSROOM DEMONSTRATIONS *Integration with Text: Chapters 6, 9 & 17*
C.O.M.I.T. DEMONSTRATION AVAILABLE

The Computer-Optimized Multimedia Intelligence Test (C.O.M.I.T.) is now available for demonstration. C.O.M.I.T., developed by TechMicro, Inc., is a computer based intelligence test which represents one of the first tests specifically optimized for computer based administration. It can be used with individuals ages 6-18 and includes both verbal and nonverbal subtests, many of which are similar to those found on well-known individually administered intelligence tests such as the Wechsler Scales and Stanford-Binet, and many group intelligence scales. To obtain a copy to demonstrate this innovative test for your class call 1-888-88-COMIT. In addition to issues related to intelligence tests, computer-assisted psychological assessment, test reliability and validity, the class demonstration can lead to interesting discussions related to test bias and fairness (the test can be translated from English into 6 languages at the click of a button).

AN EXERCISE IN CONSTRUCT VALIDITY *Integration with Text: Chapter 6*
Janet V. Smith
Pittsburg State University
jsmith@pittstate.edu

To help students fully understand the complexities of construct validity, the following class exercise can be quite valuable. Once I have introduced the different types of validity, I give students a self-report questionnaire to complete. This can be any non-threatening measure that involves a construct familiar to students but difficult to define. A measure that I have found to work well is Rubin's love scale (Rubin, 1970). Once students have scored their questionnaires, they typically begin to question what the scale is really measuring and

whether it is in fact a good measure of love, or whatever construct has been selected for the exercise. This usually quickly leads to a lively discussion of how to even define the construct in question. I then ask students to work in small groups to decide how they might go about establishing the validity of the questionnaire. This helps students recognize the difficulty in identifying suitable criterion variables and gets them thinking about different types of evidence that can be gathered in support of construct validity. The groups then share their ideas with the rest of the class and the exercise is processed together. Finally, I report to the evidence for construct validity given by the test developers and the class is always quite eager to evaluate this.

References

Rubin, Z. (1970). Measurement of romantic love. *Journal of Personality and Social Psychology*, 16, 25-273.

CLASSROOM DEMONSTRATION KAPPA EXAMPLE

Jeffrey Kahn *Integration with Text: Chapter 5*
Illinois State University
jkhahn@ilstu.edu

In our graduate course titled Test Theory, we spend some time talking about inter-rater reliability, particularly kappa. It is often hard for students to fully understand the difficulties and nuances of inter-rater reliability by looking at examples from the literature. Therefore, with the help of all students involved, we had a live demonstration. Before continuing, it should be noted that we had a small class, and this seemed to help. However, I am confident that the demonstration could easily be adapted for larger classes.

The demonstration involved seeking two volunteers from the class of 12. Of course, students were slow to volunteer (as I had not yet informed them for what they would be volunteering), but finally two brave students agreed to help. I told these two students that they would be rating ten targets on three different dimensions. Their job was to work independently and generate their own categorical ratings. The ten targets were the remaining ten students. I asked the remaining students to parade down in front of the classroom and line up, sort of like a police line-up. The targets were informed to be silent as the two judges made their ratings.

The instructions to the judges were to first rate the targets on whether or not each has dark hair or light hair. I suspected that inter-rater agreement would be high but not perfect. The raters wrote down their judgments. Second, I asked the raters to indicate whether they believed the target

is a Chicago Cubs or a St. Louis Cardinals fan (this is an important issue to many baseball fans in Central Illinois). I suspected that rating targets on this dimension would yield agreement fairly close to chance. Finally, the raters were instructed to rate whether or not each target was well educated. Being a graduate class, I had no doubt that all targets would be rated “yes” for this dimension.

After the ratings were complete, the student targets were allowed to return to their seats. We then used the actual ratings as examples of how to compute kappa. The first dimension, dark hair or light, yielded very good agreement and proved to be a good example of higher-than-chance agreement. The second dimension, Cubs versus Cardinals, yielded agreement that was exactly equal to chance. For this example, students were able to see that 50% agreement is nothing to boast about. Finally, the third dimension yielded 100% agreement—each target was rated as being well educated. This example was useful in describing the effects of very high (or very low) base rates on kappa.

This demonstration worked because of the different dimensions on which targets were rated. Selecting dimensions more relevant to the content of a course or the zeitgeist of a particular university could make the exercise even more enjoyable. This activity could easily be adapted for larger classes, perhaps by having all non-targets rate the targets but only analyze data from two judges at a time.

Students in the Test Theory class seemed to enjoy the exercise. The use of actual examples and real-life ratings helped the students gain a more clear understanding of what kappa means. In the larger scope, I believe that students were able to appreciate the challenges of attaining high levels of inter-rater reliability. This real-life example of principles of inter-rater reliability was an enjoyable change from other classroom activities, and it seemed to foster a more active learning on the part of the students

CLASSROOM DEMONSTRATIONS APPLICATION EXERCISES ENLIVEN COURSES IN EDUCATIONAL TESTS AND MEASUREMENT

Victoria B. Damiani, Ed.D., NCSP
Indiana University of Pennsylvania (412) 357-2316

*A*sk your undergraduate students in education why they would like to be teachers and you are likely to hear statements about liking children, enjoying the opportunity to help others learn, believing in the value of education, and appreciating the chance to engage in a hands-on career. It is understandable that a course which has the words tests and measurement in the title, graphs and mathematical formulas

in the text, and a note in the catalog that says “required,” may not pique the interest of these students. At our university, Educational Tests and Measurements is taken by students in their junior year. Therefore, the students are often in public schools all or part of the day and take the course at night. The mix of content that does not appear to be interesting, the late hour, and involvement of MATH (!) can really challenge an instructor to keep students engaged.

Of course, those of us who have worked in schools most of our lives understand the importance of being knowledgeable about assessment, not only for the sake of our students, but for the protection of our own professional well-being. I usually begin my course by making a clear connection between sophistication regarding assessment and success in teaching. I bring newspaper articles that show ratings of school districts with regard to standardized achievement testing. I discuss the potential role of standardized assessment in performance evaluation of teachers and point out the value of being able to discuss current assessment procedures during a job interview. My approach to the course is to treat the students as teachers in training, that is, as learners with a clear professional goal in mind. I am progressing toward my goal of improving my instruction in this course, but I am not there yet. Following are some instructional techniques I have utilized that have been well-received by the students. They incorporate four basic themes.

Teamwork—Teamwork is utilized for two reasons. First, and truthfully foremost, is that it breaks the monotony of instructor presentation. Secondly, it mimics the emerging real-life environment of the teacher. The days when a teacher could just close the door and operate alone with the class are gone, if they ever truly existed. Teachers today often team teach as well as work on instructional support teams, curriculum committees, and crisis intervention teams, to name just a few. Staying on task as a group, assigning roles, and being responsible for group decision making are some of the skills students build as they address problems to be solved.

Problem-based Learning—This technique originated in the field of medicine. The primary goal is learning through solution of an ill-structured, life-like problem. The process is intended to include data gathering, integration of information, and reasoning, in an effort to attain an original solution to the problem.

Oral Presentation—Presenting is, of course, common in classrooms at all levels. It is intended to foster communication skills, to involve peers in student learning, and to represent a break from exposure to only the instructor’s voice and point of view.

Authentic Activity—Authentic activities require students to apply what they have learned to real world situations. Clearly, the further along in a training program a student is, the more important life-like learning opportunities become.

These themes are incorporated into the following activities during the semester:

A brief plan for teaching and assessing in a nonacademic activity with which the student is familiar. This activity takes place at the beginning of the semester and is intended to help the student see the importance of deciding what to teach before deciding on an assessment. Each student makes a list of important things for one to know about skiing, bread baking, golf, or any other nonacademic skill in which the student is competent. Then the students choose some methods of assessing achievement in the area. Ideas are shared with another student and then listed on the board. The importance of a teaching goal and assessment match emerges from the lesson. **(Authentic Activity)**

Review of test catalogs. Early in the semester students review test catalogs. They share their impressions with the class and list 4 terms with which they are not familiar. Issues concerning the cost and business end of assessment, the wide variety of characteristics that can be assessed, and methods for assessing things like depression and self-esteem usually emerge from these discussions and set the stage for the rest of the course. Terms frequently listed for clarification are validity, reliability, standardization, and correlation. **(Authentic Activity)**

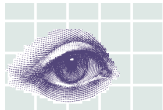
Ethical dilemmas. This activity follows the students' reading of the Code of Fair Testing Practices, a publication of the National Council on Measurement in Education. Student teams are presented with life-like ethical dilemmas involving assessment. One involves a parent request for a copy of a standardized achievement test, another a teacher's question about teaching to the test. One student presents the problem. The others act as an ethics board and delineate the issues involved and potential solutions. The team then presents the issues to the class. **(Problem-based Learning, Teamwork, Oral Presentation)**

Teacher-made test. As a semester-long performance assessment, each student develops a test in an academic area for any grade, 3-12. The test must include true/false, multiple choice, and short answer items, as well as an essay and a performance assessment. The student must also describe an appropriate modification of the assessment for a child with either a visual impairment, hearing impairment, or specific learning disability. In preparation for the project, student teams develop practice tests on a chapter of their book and evaluate each other's tests. **(Authentic Activity, Teamwork**-for practice session only)

Selection of assessment techniques. Toward the end of the semester, the importance of matching assessment techniques with information need is emphasized. Student teams are presented with real-life scenarios and asked to design an assessment procedure. One scenario involves a high school student who is trying to decide whether to apply to college. A second concerns a child with a suspected learning disability

and a third, a student with a problem in written language. After their assessments are designed, the teams are presented with test results and asked to interpret them for their client. Skills in interpretation of percentile ranks, standard scores, and grade equivalents are utilized here. **(Teamwork, Problem-based Learning, Authentic Activity, Oral Presentation)**

Pre and Post test results of students in the Educational Tests and Measurement course based on the Standards for Teacher Competence in Educational Assessment of Students (published by the American Federation of Teachers, National Council on Measurement in Education, and the National Education Association) have shown considerable improvement in skills, based on student reports. However, these classroom activities are subject to the common criticisms of active and collaborative learning procedures. Some students perform better on their own. Some have found the activities boring and some teams finish their work long before others. I vary team composition so that chronic team interaction problems do not develop. I also make teamwork eligible only for extra points, not a large part of the grade for the course. For the most part, though, the activities have been well received, especially in the long evening class. I am currently looking into the use of case studies in this course and would welcome any ideas along those lines.



MEASUREMENT IN THE NEWS

IMPROVING STUDENTS' PERFORMANCE ON ACHIEVEMENT TESTS

Gloria C. Maccow
Guilford County Schools
MACCOWG@aol.com

If we review 1998 issues of Newsweek magazine, we find several articles addressing an important measurement issue: What to do to improve the test scores of students in public schools. Home schooling, tutoring, and reducing class size were options discussed by this national medium. Newsweek also discussed an important related theoretical issue, i.e., how to improve memory. The link between memory and student achievement is an important one when we consider that where and what we teach must be determined by how children acquire, remember, and use information.

During the early grades, the curriculum focuses on teaching basic skills in reading, writing, and mathematics. Students store information in their memory and are asked later to recall words for reading and writing and basic facts to solve math problems. In reading, recall is effective when children recognize on sight words to which they previously were exposed. When students have a large number of words in their sight vocabulary, they can read fluently and at a fast pace. This will have a positive impact on their understanding of what they are reading.

The importance of memory for learning is discussed in a 1998 *Newsweek* article (June 15). The process of remembering involves three features: For reading, students first must take note of the visual features of each word. If they are distracted when the teacher is discussing the words, students will not get the necessary information into memory. Indeed, there are many students who struggle at this level. These students, who have difficulty focusing and sustaining their attention, often are treated with stimulant medication. In a recent article (November 23, 1998), Susan Brink indicated that approximately one million children in America's schools take medication to focus and sustain their attention. For many students, medication intervention is effective. However, it is also important for students to learn strategies to improve their ability to get information into memory. These strategies may include being seated in an area free of distral Government (*Newsweek*, August 31, 1998). From Kindergarten through third grade, California's classes are reduced from an average of 29 students to 20 students. Some teachers believe this will improve test scores, but critics wonder whether tutoring,

increased teacher training, and extending school through the summer might yield similar results and be more cost effective. Indeed, to staff the additional classrooms, California has had to hire teachers with "emergency credentials" which may mean inadequate teacher training.

Homework is another technique educators use to improve achievement. However, homework may have little impact on the achievement of elementary-age students (see *Newsweek*, March 30, 1998) and may have a negative impact on students' attitudes toward school. At middle and high school levels, there is a significant positive relationship between homework and achievement. Based on these findings, researchers have made several suggestions for homework at the elementary level. First, it should consist of short assignments. Second, homework should be different from classwork and should require creativity and exploratory activity. Finally, assignments should relate to the next day's classwork and should be focused.

Like educators, many parents are concerned that the public schools are not meeting the needs of children. In an effort to improve their children's achievement, parents often engage the services of tutors (*Newsweek*, March 30, 1998). Some parents move in this direction because they are dissatisfied with their child's scores on standardized achievement tests. An example may be when the scores are not high enough for a child to be admitted to the school's Gifted and Talented Program. However, most parents who seek tutoring services, are trying to remediate their children's academic deficits that are due to learning disabilities or other handicapping conditions.

For parents who believe they can do a better job teaching their children than trained teachers, there is the option of home-schooling. A large number of parents identify safety as their primary reason for choosing to home-school. They want to protect their children from influences such as drugs, violence, and sex. In 1993, all 50 states legalized home-schooling from Kindergarten to college. The Home Education Research Institute estimates that, at this time, approximately 1.5 million children are being educated primarily by their mothers or fathers. This represents a dramatic increase from 1990 when approximately 300,000 students were educated at home.

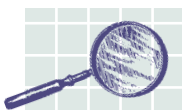
From these articles, it appears that more research is needed to establish the efficacy of programs such as tutoring, class size reduction, and home-schooling. Certainly, it is important to establish how effective these programs are in increasing scores on achievement tests. However, what emerges repeatedly is that both educators and parents are concerned about the quality of education provided by public schools. Furthermore, educational achievement is important to so many that articles on public schools appear frequently in the popular press. It is hoped that such scrutiny will stimulate research on the effectiveness of current programs and lead to proposals that could improve educational achievement for all students.

(Continued on page 8)

MEASUREMENT IN THE NEWS *(Continued from page 7)*

References

- Adler, J. (1998, March 30). The Tutor Age. *Newsweek*, 47-50.
- Begley, S. (1998, March 30). Homework doesn't help. *Newsweek*, 50-51.
- Brink, S. (1998, November 23). Doing Ritalin right: Sure, it works - but there are big flaws in the way it's being given. *U.S. News & World Report*, 76-81.
- Cowley, J., & Underwood, A. (1998, June 15). Memory. *Newsweek*, 49-54.
- Foote, D. (1998, August 31). California's Space Race: Will small classes save the state's failing schools? *Newsweek*, 57.
- Kantrowitz, B., & Wingert, P. (1998, October 5). Learning at home: Does it pass the test? *Newsweek*, 64-70.



RESEARCH SAMPLER

The psychometric base of the Child Abuse Potential (CAP) Inventory, a screening scale for physical child abuse, is provided in an article by Joel S. Milner, the author of the scale. The article, "Assessing Physical Child Abuse Risk: The Child Abuse Potential Inventory", a sample of the scale, and a detailed bibliography is available from: Dr. Joel S. Milner, Director of Family Violence Research Program, Department of Psychology, Northern Illinois University, DeKalb, IL 60115-2892.

Integration with Text: Chapters 6 & 13

William E. Addision and Kristine Hillman of Eastern Illinois University presented an APA paper, "A Three Dimensional Model of the Normal Curve". The paper includes a description of how to construct a 6 piece wooden model of the Normal Curve that is approximately 20 cm high and 32 cm long. The authors also describe a number of classroom activities utilizing the model. Reprints can be requested from Dr. William E. Addision, Department of Psychology, Eastern Illinois University, Charleston, IL 1920.

Integration with Text: Chapter 3

Kirk Heilbrun (HEILBRUN@WPO.AUHS.EDU) of Villanova College of Law presented a paper at the 1998 APA Conference related to the Daubert Standard. The paper, "Principles of Forensic Mental Health Assessment: Guidelines for the Judicial Application of Daubert", presents guidelines for the application of Daubert criteria to cases involving forensic assessment for the courts.

Integration with Text: Chapters 13 & 14.

Robert Perloff of the Katz Graduate School of Business at the University of Pittsburgh presented a paper at the 1998 APA Conference entitled, "Test Taker Rights and Responsibilities: Practical, Substantive, and Stylistic Issues". The paper presents his thoughts on the efforts of the Joint Committee on Testing Practices to develop Test Takers Rights and Responsibilities protocols. In his paper, Dr. Perloff makes reference to a cartoon in the April 6, 1998 issue of *The New Yorker* which depicts a gravestone with the inscription, "Here lies Frederick Jones Verbal 680-Math 720". Reprints of the paper can be requested from Dr. Robert Perloff, Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, PA 15232, 412/648-1554.

Integration with Text: Chapter 2

Stephen G. Sireci (Sireci@acad.umass.edu) and B. Bastari of the University of Massachusetts-Amherst presented a paper at the 1998 APA Conference dealing with the comparability or equivalence of constructs measured by different language versions of assessment instruments. Their paper, "Evaluating Construct Equivalence Across Adapted Tests" presents a critique of several methods for evaluating structural equivalence across different language versions of tests or questionnaires. They concluded that weighted multidimensional scaling and confirmatory factor analysis were most effective.

Integration with Text: Chapters 2 & 6.

A symposium was conducted at the 1998 annual conference of the National Association of School Psychologists dealing with the most recent revision of the Standards for Psychological and Educational Tests. Contact Dr. David Goh (goh@qcvaxa.acc.qc.edu) for an update on status of the newest edition of the Standards scheduled to be released shortly.

Integration with Text: Chapter 1

Two recent papers have been published related to the validity of the Graduate Record Examination. "Does the Graduate Record Examination Predict Meaningful Success in the Graduate Training of Psychologists" is available from Dr. Robert Sternberg, Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205. "Comprehensive Meta-Analysis of the Predictive Validity of the Graduate Record Examinations: Implications for Graduate School Selection and Performance" is available from Dr. Nathan R. Kuncel, Department of Psychology, University of Minnesota, N218 Elliott Hall, 75 East River Road, Minneapolis, MN 55455-0344. Results of the Kuncel study indicated that the GRE are valid predictors of graduate grade point average, first year graduate grade point average, comprehensive exam scores, and faculty ratings with subject tests being better predictors than Verbal, Quantitative, and Analytical tests. The GRE did not predict time to complete the degree program and number of publications. The Sternberg study also found the test to be useful in predicting first year grades but not other kinds of performance.

Integration with Text: Chapters 6 & 10

The final reports on Evaluating National Tests and Assessments of the National Academy of Sciences studies are available at the National Academy's Web site at <http://www.nas.edu/morenews>.

Integration with Text: 10

The symposium, "Testing Diverse Populations with Three Nonverbal Intelligence Tests", was presented at the 1998 APA conference. Three papers were presented related to the use of the General Ability Measure for Adults, The Universal Nonverbal Intelligence Test, and the Naglieri Nonverbal Ability Test, with diverse populations. Summaries of each of the three papers and additional information are available from: Dr. Jack A. Naglieri, 356 Arps Hall, 1945 North High Street, Ohio State University, Columbus, OH 43210 or e-mail at: naglieri1@osu.edu.

Integration with Text: Chapters 8 & 9.

GRE ADDS WRITING ASSESSMENT

Integration with Text: Chapter 10

The following press release was developed by the staff of the Educational Testing Service and describes the new GRE Writing Assessment that will be offered beginning in October of 1999. Additional information can be obtained from ETS representative Kevin Gonzalez at (609) 734-1617, or e-mail him at kgonzalez@ets.org. Additional information about the nature of the test with examples of writing samples are provided on the GRE Web site located at www.gre.org/writing.html.

New GRE Writing Assessment Score Highlights Key Competency

Starting in October 1999, the new Graduate Record Examinations (GRE) Writing Assessment will give prospective graduate students the opportunity to demonstrate the critical reasoning and analytical writing skills recognized as essential for success in school and beyond.

The new test was created by Educational Testing Service (ETS) at the request of and in cooperation with the GRE Board. The assessment will significantly expand the range of skills assessed by the GRE General Test and the GRE Subject Tests, including the candidate's ability to:

- articulate complex ideas clearly and effectively
- examine claims and accompanying evidence
- support ideas with relevant reasons and examples
- sustain a well-focused, coherent discussion
- control the elements of standard written English

The GRE Writing Assessment will be offered independently of the GRE General Test and GRE Subject Tests, and will be available year-round at all ETS-authorized computer-based testing centers worldwide. Additional

information will be available on the GRE Web site at www.gre.org/writing.html in early November 1998 and will be forthcoming in GRE publications.

The assessment consists of two analytical writing tasks: a 45-minute "Present Your Perspective on an Issue" task and a 30-minute "Analyze an Argument" task.

The "Issue" task states an opinion on an issue of broad interest and asks test takers to address the issue from any perspective they wish, so long as they provide relevant reasons and examples to explain and support their views. The "Argument" task presents a different challenge: it requires test takers to critique an argument by discussing how well reasoned they find it. Test takers are asked to consider the logical soundness of the argument rather than agree or disagree with the position it presents.

For the "Issue" task, the test taker will be able to choose one of two essay topics randomly selected by computer from the "Issues" pool. The "Argument" task does not present a choice of topics; instead, the test taker will be presented with a single topic randomly selected by computer from the "Argument" pool. More than 100 topics have been developed for each writing task, and both topic pools will be widely published. All topics used in the GRE Writing Assessment have passed rigorous reviews for fairness, and, in national field test trials, proved accessible and appropriate for entry-level graduate students from many disciplines and from various countries and cultural groups.

Test takers may either word process or hand write their essays. GRE reader training and scoring processes are designed to ensure that the same standards are applied to all essays, no matter which response mode is selected.

Essay scoring will be performed by college and university faculty experienced in teaching writing or writing-intensive courses. All readers will have undergone careful training and will have passed stringent GRE qualifying tests.

Each essay will be scored by two readers, on a 6-point holistic scale according to criteria published in GRE scoring guides. A single analytical writing score, representing the average of a test taker's scores for the two essays, will be reported on the same 6-point scale. Scores will be sent to institutions and test takers within 10 to 15 days.

Test familiarization materials explaining the goals, content, format and other characteristics of the new GRE Writing Assessment will appear in several places—on the Web at www.gre.org/writing.html, in student brochures, and in materials sent directly to colleges and universities. Information will include:

- the complete set of directions for both tasks
- the entire pool of "Issue" and "Argument" topics, which test takers can review prior to the test
- samples of scored essays with readers' commentary
- the word-processing tutorial specific to this assessment

ETHICAL CONFLICTS RELATED TO PSYCHOLOGICAL TESTING AND ASSESSMENT

Integration with Text: Chapters 2, 6, 16 and 17

For those of you who include discussion of ethical issues/conflicts related to the use of psychological tests in your measurement course, you should review Don Bersoff's 1995 book, *Ethical Conflicts In Psychology*. The book includes a chapter on psychological assessment and provides a treatment of ethical issues related to all areas of psychology. In particular, topics such as test validity (Cohen & Swerdlik Chapter 6); industrial/personnel psychology and integrity tests (Cohen & Swerdlik Chapter 16); computerized psychological testing (Cohen & Swerdlik Chapter 17); as well as ethical issues related to the general process of psychological assessment (Cohen & Swerdlik Chapter 2). Chapters are authored by Bersoff and others with expertise in the various areas of psychological testing and assessment including Joseph D. Matarazzo and Irving Weiner. These chapters make excellent supplementary readings for your students or provide useful information to supplement your lectures on this topic. The book is available from APA (800) 374-2721.

References

Bersoff, D. N. (1995). *Ethical conflicts in Psychology*. Washington, D.C.: American Psychological Association

PROJECT SYLLABUS

Condensed from the Fall, 1998

Newsletter of the Society for the Teaching of Psychology

Looking for some new ideas for your classes? The Project Syllabus Task Force of Division 2, Teaching of Psychology, has established a syllabus section of the OTRP (Office of Teaching Resources in Psychology) Web page (<http://www.lemoyne.edu/OTRP/projectsyllabus>) and faculty can now download documents using Adobe Acrobat Reader. In addition, the task force is beginning to add links to Web-based syllabi. If you have syllabi that you believe would make unique contributions to their collection and/or if you have accessible Web syllabi for your courses, please submit a digital copy (or URL) for the Task Force to review (Mallen@csubak.edu; Mary Allen, Department of Psychology, CA State University, Bakersfield, CA 93311-1099).

TEACHING MATERIALS AVAILABLE FROM TEST PUBLISHERS

A number of the large test publishers, including The Psychological Corporation (800/211-8378), Riverside Publishing (800/323-9540), and American Guidance Service (800/328-2560), have developed supplementary teaching materials (overheads/Powerpoint presentations, handouts etc.) for a number of their recently published tests. A number of these materials can be downloaded from their Web sites. As types of materials available and methods of distribution differ, it is best to contact each publisher directly to inquire about availability. These publishers also distribute periodic newsletters (e.g., *Assessment Information Exchange* published by AGS) with information that can serve as useful supplements to lecture/discussion in your classes.