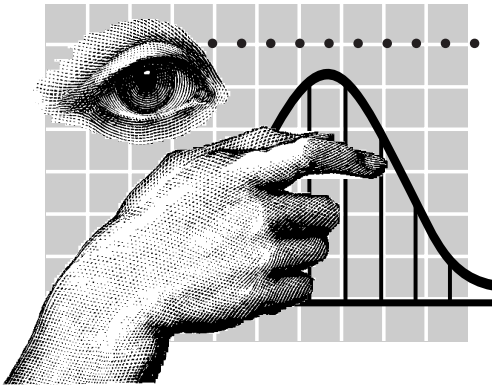



# MEASUREMENT FORUM

Volume 3 Number 1
Fall, 1997

A Newsletter for the Promotion of Excellence in Measurement Instruction

www.cas.ilstu.edu/psychology/measurementnewsletter.html

Prepared for use with ***Psychological Testing and Assessment,*** Third Edition, by Ronald Jay Cohen, Mark E. Swerdlik, and Suzanne M. Phillips.


Welcome to Issue Number 1 of Volume 3 of the *Measurement Forum*. With this current issue, we join Mayfield Publishing Company in continuing to meet our primary objective of providing information that will promote excellence in measurement instruction.

We trust you had a relaxing summer and your fall semester has begun smoothly. We hope this issue of the *Measurement Forum* will contain ideas for classroom demonstrations and supplementary materials for your classes that will contribute to making your semester a rewarding one for you and your students. In this issue, we have also included a number of items related to teaching more generally. These items include, for example, active learning techniques and discussion of dual relationships with students. We hope these items will stimulate your thinking of issues that pertain not only to teaching psychological measurement but any course in higher education.

This issue contains a number of classroom demonstrations/activities submitted by *Forum* readers. We encourage you to take some time, as did *Forum* readers such as Patricia L. Bromley of University of Wisconsin at Platteville, George Fago of Ursinus College, and Bruce Bracken of the University of Memphis, to share an idea or two with us for our next issue. In addition, at the request of Professor Jack Flynn from Connecticut State University to develop a set of slides/overhead transparencies for use in the teaching of psychological measurement (see Professor Flynn's letter in the Winter/Spring 1997 issue on page 6 and the item in this issue, "Call for Art), we are continuing to compile materials. Please consider contributing to this effort that could benefit all instructors.

Best wishes,

Mark E. Swerdlik
Ronald Jay Cohen
Editors



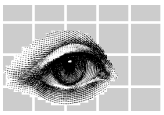

Please address all correspondence to:

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@rs6000.cmp.ilstu.edu


For additional copies of the *Measurement Forum* or for more information about the new third edition of ***Psychological Testing and Assessment***, contact Mayfield Publishing Company at (800) 433-1279 or jbauer@mayfieldpub.com

## BIBLIOGRAPHY AND EXAMPLES OF ACTIVE LEARNING TECHNIQUES AVAILABLE

Active learning refers to techniques where students do more than simply listen to a lecture. Students are doing something including discovering, processing, and applying information. The elements of active learning include talking and listening, writing, reading, and reflecting. A complete bibliography and examples of active learning techniques were developed by Dr. Kathleen McKinney of the Center for the Advancement of Teaching at Illinois State University. Copies are available from Mark Swerdlik, Department of Psychology, Campus Box 4620, Illinois State University, Normal, IL 61790-4620.



## MEASUREMENT IN THE NEWS

*Integration with Text: Chapter 10*

### THE EFFECTS OF USING A NORMATIVE APPROACH TO DETERMINE ACHIEVEMENT OF EDUCATIONAL OBJECTIVES

Gloria C. Maccow
Guilford County Schools
North Carolina

The news media provide reasons for us to wonder about how the achievement scores of American students compare to the scores of students in other countries. In December, 1996 *(Newsweek),* Pat Wingert reported that American eighth grade students scored below the international average on math tests. The explanation for the comparatively lower scores was not that American students spend less time in class than students in Japan and Germany, for example, but in the way math is taught in the United States. According to the report, math teachers in the United States tend to teach skills without developing students' understanding of concepts. Furthermore, the eighth grade math curriculum in the United States covers topics mastered by seventh graders in other countries. Wingert concluded what others have said before: because educators set low standards for students, students rise to mediocrity.

In contrast to the dismal findings for eighth grade students, nine-year-olds scored above the international average in science and math according to a more recent report (Glick, *Newsweek,* June 23, 1997). The author hypothesized that the differences in scores between elementary-age students and eighth graders in the United States is that educational reforms have not reached middle school. Although the explanations advanced are meritorious, there are two other factors that must be considered in interpreting the results of the Third International Mathematics and Science Study. These factors are both related to statistics and measurement.

The first has to do with the statistics used to describe student achievement. The information provided in the *Newsweek* article used mean scores. Because the statistical average is sensitive to extremely large or extremely small scores, information on variability is necessary for readers to make sense of the findings. For example, it would be important to know how many students scored above or below the average. Were there a number of very low scores that were counterbalanced by extremely high scores? Recently, one of my colleagues described the paradox of average scores. By using mean scores, two people, theoretically, could be admitted to graduate study based on their average score of 1100 on the Graduate Record Examination. Never mind that one student earned a score of 1600 and the other a score of 600. Fortunately, for university professors and for society as a whole, admissions decisions are not based on average scores. In discussing comparative results, the popular press must understand the shortcomings of mean scores. Data about the variability of the scores would improve the quality of the information presented.

Second, there *should* be concern about the way we measure student achievement and the standards students must meet to earn specific grades. Kurt Wiesenfeld (*Newsweek,* June 17, 1996), a professor at Georgia Institute of Technology in Atlanta, described how students "wheedle for a degree as if it were a freebie T shirt" (p. 16). He lamented the fact that students who did poorly on exams and activities throughout the semester expected to bargain for a passing grade after final grades were posted. I concur with Professor Wiesenfeld that students do not establish the connection between their grades and their own effort. Indeed, in arguing for a higher grade, students often use reasons unrelated to academic performance. Some students argue they need a better grade to avoid losing their scholarships. Other students do not have such (valid?) reasons for their requests. One of my own students, a senior who would be a teacher in a few months, received a final grade of B in my class. After she had received her grade report in the mail, she came to my office to ask for extra credit so she could get an A. Her reasoning was that she did not complete the extra credit assignments during the regular semester but that she wanted some consideration now. The irony was that this student was highly intelligent and could understand novel concepts with minimal effort. She had become used to a system in which her intellectual ability allowed her to get As because her grades were generally higher than the majority of the students in her class. Naturally, she disliked my class where a grade of A was based on mastery of a specified body of knowledge.

Several of the students in my classes have disliked the fact that I did not grade on a "curve." Others were happy that anybody could earn a grade of A by mastering at least 90% of the curricular objectives. This dichotomy is one I remember discussing with classmates years ago, when I was an undergraduate student. A number of my peers were lamenting how much they hated being graded on a normal curve. If 25, as would be assigned based on the normal curve, my friends were always ranked 26th or 27th and, therefore, earned a B. I had a somewhat different perspective which made no sense to any of my peers. I told them I hated the normal curve method too, but for a different reason. I described a course in which I earned grades of 75, 80, and 79 on exams. I earned an A in that class because my grades were the highest of all the students in the class. Although I was happy to have earned an A, I felt "cheated" because my grade did not reflect the reality that I had mastered only

about three-fourths of the objectives in the class. What did I miss in the 25% of the information I had not mastered? Would I be able to function effectively without that knowledge? Wiesenfeld addressed a similar issue when he talked about the importance of getting the answer right and not just getting partial credit. When scientists and engineers get the answer partially right, catastrophe could result. Wiesenfeld used several illustrative examples, including the collapse of a light tower in Olympic Stadium because an engineer miscalculated how much weight it could hold.

Few students associate a grade with their mastery of curricular objectives. Instead, grades are regarded as a means of getting scholarships, gaining admission to specific graduate programs, or competing for certain jobs. It is my contention that this disconnection between grades and mastery of objectives is related to the normative approach used in assigning grades. Elementary-age students learn early that their learning is contingent, not only on their own mastery of curricular objectives, but on their classmates' as well. I propose the following: use a criterion-referenced approach in assessing student achievement. This will teach students that grades are related to their mastery of objectives. Also, students will understand that instruction will continue until all of the objectives in the curriculum are mastered—not until the grades are assigned. Such an approach will foster an environment of collaboration because students would not be concerned about doing well at the expense of classmates. In such an environment, students would work together to experiment and understand concepts in math and science. They would make the United States proud when they demonstrate their understanding and knowledge on future International Mathematics and Science studies.

References

Glick. (1997, June 23). An encouraging report card. *Newsweek*, 67.
Wiesenfeld, K. (1996, June 17). Making the grade. *Newsweek*, 16.
Wingert, P. (1996, December 2). The sum of mediocrity, *Newsweek*, 96.

## HISTORICAL PERSPECTIVES

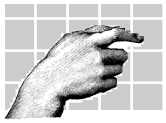
A recent article of interest to instructors to supplement their lectures on the history of intelligence testing, particularly related to Alfred Binet and the Binet-Simon scales, would be "Alfred Binet and the quest for testing higher mental functioning" appearing in the summer, 1996 issue of *Pictorial History of Psychology.* The article includes some very interesting pictures of Alfred Binet and of students who participated in Binet's investigations on eyewitness testimony. Reprints are available from: Dr. Jacqueline L. Cunningham, Department of Behavioral Sciences, Children's National Medical Center, 111 Michigan Avenue, N.W. Washington, DC 20010-2910. Dr. Cunningham has also published articles dealing with the history of Binet's contributions including "Contributions to the history of Psychology: XLVL. The Pioneering Work of Alfred Binet on Children as Eyewitness," *Psychological Reports,* 1988, 62, 271–277 and Binet's "Contextual Study of Memory," *Psychological Reports,* 1995, 77, 955-961.

## CALL FOR ART

Do you have slides that you routinely use in teaching measurement? Do you have other artwork used in the classroom that can be converted to slides? Do you have an idea for a slide that would be particularly useful in teaching measurement?

Inspired by the letter to the *Forum* from Professor Jack Flynn (Connecticut State University), Mark Swerdlik and Ronald Jay Cohen will be editing a set of slides/overhead transparencies for use in the teaching of measurement. Slides, transparencies, or other art should be submitted along with a brief description of how the art is used in your course. The contribution of all contributors will be credited in the published guide to this instructional resource. Additionally, all contributors will receive a complimentary copy of the complete set of materials. We hope to make this instructional resource available to other instructors at cost.

Please send your artwork and accompanying text to: Mark E. Swerdlik, Ph.D., Department of Psychology, Campus Box 4620, Illinois State University, Normal, IL 61790-4620. Questions? Please call Mark Swerdlik at (309) 438-5720 or e-mail him at meswerd@r56000.cmp.ilstuicdu.



## CLASSROOM DEMONSTRATIONS

### TEACHING WITH THE NATIONAL LONGITUDINAL SURVEY OF YOUTH DATABASE

Suzanne M. Phillips
Gordon College

*Integration with Text:*
*Construct Validity Ch. 6 (pp. 193–208)*
*Test Bias Ch. 6 (pp. 200–201, 208–212)*
*Theories of Intelligence Ch. 8 (pp. 263–266)*
*Group Intelligence Tests Ch. 9 (pp. 329–332)*

The National Longitudinal Survey of Youth (NLSY) is a continuing longitudinal study of 12,686 individuals interviewed for the first time in 1979, when they were 14 to 22 years of age. The resulting data have been organized for researchers by the Center for Human Resource Research, The Ohio State University, which distributes a CD-ROM of the data set and clear, detailed documentation, essentially at a materials cost (approximately $20).

Data relevant to a particular topic or exercise can be easily extracted from the CD-ROM and distributed electronically to students. For an exercise in construct validity and factor analysis, I use scores of the 11,914 NLSY youths who took the Armed Services Vocational Aptitude Battery (ASVAB). I extracted scores on the ten ASVAB subtests and showed my students how to factor analyze them using a mainframe SPSS statistical package. I also showed students how to generate a correlation matrix of the ten subtests.

This exercise was helpful in several ways. It allowed students to gain hands-on experience with factor analysis; they were able to see the eigenvalues used by SPSS to determine the number of factors, and they could examine rotated factor loadings to speculate as to the identity of the factors. The analysis indicated the presence of only two factors, the first of which accounted for fully 67% of the variance in subtest scores. Given the purpose of the ASVAB, which is to help sort people into military career paths, the presence of only two factors and the large size of the first factor raise questions about the construct validity of the test. Students who had not been able to understand how factor analysis is used to assess construct validity found this aspect of the exercise helpful. Finally, the demonstration offers a starting place for a discussion of g. The presence of a very large factor, coupled with high correlations between subtests that may seem unrelated (for example, the Science and Vocabulary subtests correlated .83), challenges those who believe that only specific intelligences exist.

Many other teaching applications of this database exist. One would be an examination of issues of race and cognitive ability. The NLSY was the database used so extensively in Herrnstein and Murray's *The Bell Curve* (1994, Free Press). Charles Murray has graciously provided me the supplementary information needed to repeat (or modify) analyses reported in *The Bell Curve*. I am happy to share this supplementary information with interested people; I may be contacted at sphillips@gordonc.edu.

The NLSY database is a rich resource for teaching students about psychological testing. The database can be obtained through:

Center for Human Resource Research
The Ohio State University
921 Chatham Lane, Suite 200
Columbus, OH 43221-2418
web: http://stats.bis.gov/nlshome.htm
e-mail: usersvc@pewter.chrr.ohio-state.edu
phone: (614) 442-7300
fax: (614) 442-7329

### CONSTRUCTING TESTS IN CLASS: MAKING THE THEORETICAL PRACTICAL

Patricia L. Bromley
University of Wisconsin—Platteville

*Integration with Text:*
*Chapters 5, 6, 7*

Constructing and piloting a test in small groups can greatly enhance students' understanding of basic measurement concepts, and give students an appreciation of the practical difficulties in test construction and validation. For several years, I have been having students write and pilot a test as part of the course curriculum in Psychological Measurements. I usually provide the class with a list of five or six possible topic areas, and students are assigned to groups based on their rankings of preference. Usually, students are able to participate in their first or second choice topic area. Class time is used for many of the assignments.

The first step is conceptualization of the test. Students are given the first assignment, which is to obtain at least three articles or books on the topic—anything that might help them understand the construct. The books or articles might also provide suggestions regarding who might use the test and why, and what format(s) might be used. Students are asked to divide the topic area up among the group, and to make sure that at least one group member ascertains whether there already are similar tests available (and the quality and features of such tests). The groups meet during class time and discuss what they have found, and decide what further information would be desirable. They continue developing their test construct as the class covers the introductory chapters of the text.

The next step, item-writing, occurs as the class covers the chapter on test development. The various item formats are discussed in class. For practical reasons, paper and pencil formats must be used for the tests, and the length of the final version of the test is limited to two pages. Groups decide upon an item format, write directions for the examinee,

develop a title for the tests, and begin writing items. They may choose to work collaboratively on item-writing, or they may choose to write items individually and bring them to class for the group to critique. When the group has finished editing, the items are submitted to the instructor for editing.

The groups are then given the assignment of describing, in writing, how they would demonstrate the reliability and validity of their tests. They must state which type(s) of reliability would be most important to establish, and they must describe, in detail, how this would be accomplished. They must do the same for validity. Students are to assume that they should attempt to establish content, construct, and criterion-related validity of their test, and they must describe how this would be done, who would be their panel of "experts," and what other measures would be used in establishing validity. They are also to describe the important face validity issues related to their test. Thus, students must know the literature in their topic area and related topic areas, and must know the quality and type of tests that are currently available in the area.

At this point, the tests are usually ready for preliminary work in class. Depending on the topic, there may be in-class experts (e.g., for a test on roommate satisfaction), or groups may need to go outside of class to establish content validity. When the items and instructions have gone through final revisions, the groups apply for permission from the human subjects committee to pilot the tests outside class. The students ask for permission to do two administrations of their instrument (assuming they wish to establish test-retest reliability), and to administer at least one other instrument along with their own, to partially establish validity. Meanwhile, the tests are piloted in class, and class members are encouraged to give feedback about each test and each item. This part of the project usually amazes group members. Items that students thought were crystal clear are found to be misleading. Sometimes class members surprise colleagues by strongly disliking their response format (true-false tests are particularly difficult to "sell" in class). At this point, using the data from the in-class administration, students perform item-total correlations and other item analyses to determine which items should be retained. If students are granted permission to administer the tests outside class, this is done, as is a second administration. The students then compute reliability and validity coefficients. The last step is an oral presentation to the class, in which each group describes and explains the development of its instrument.

Course evaluations frequently list this series of assignments as the most meaningful part of the course. Students agree that the group projects keep the class from becoming dry. However, difficulties sometime arise with the projects. Over the past few semesters, I have tried various permutations of the approach and can offer additional hints. First, self-selection to groups is to be avoided, because it tends to result in formation of cliques. If students rank-order their topic preferences in class, without opportunity to discuss their preferences with friends, groups seem to function better. I have tried allowing students to generate their own topics, or, to form groups and select a topic, but have found these alternatives both time-consuming and unnecessary. Students who have chosen a preferred topic off a short list seem to be reasonably content in their groups. Second, the length of the semester imposes constraints, in that some groups my not reach the point where they are ready to administer their instruments outside of class. For this reason, I recommend that the tests be designed for use with college students. This way, it is possible to do some data analysis from the in-class administration of the tests, even if students are not able to administer the tests outside of class. A third issue is that the groups may work at different paces. For instance, defining one construct may be more difficult than defining another construct. This makes use of class time for specific assignments tricky. I have dealt with this problem by providing the entire list of assignments at the beginning of the semester, so groups can work ahead. Finally, there will be uneven levels of investment within groups. I have found that tying 100 points to class participation helps ensure that everyone in every group participates actively. The last day of class, each student rates the participation of each other member of the group. These ratings enter into my estimation of the student's level of class participation.

## PSYCHOMETRIC SCALING: LEARNING BY DOING

George Fago
Ursinus College

*Integration with Text: Scaling, Chapter 7*

Psychometric scales or 'psychological tests' are typically of intense interest to beginning psychology majors who tend to see them as mysterious, threatening, or even ridiculous. The frequent lack of face validity of scales themselves coupled with the abstract nature of the theoretical rationale for scaling makes the topic a difficult one for most beginning students to grasp. However, when students move through each step of the process of scaling a particular concept, the entire task becomes cognitively accessible precisely because it is concrete. What follows is an approach for teaching scaling to novice majors in a manner that demands active, participatory learning. In this method, students work collectively through each step in the process of creating a Thurstone Scale with equally appearing intervals, receiving feedback in the process as they generate possible scale items, serve as raters in the normative sample, and select the final scale items.

By convention, scale development is time-consuming and difficult. A minimum of 100 potential items must be generated and each of the items must then be rated by each

## CLASSROOM DEMONSTRATIONS *(Continued from page 5)*

of a minimum of twenty raters. Means, medians, and standard deviations are obtained for each item for use in item selection. This is arduous for a single investigator; to coordinate the activities of twenty student 'investigators' adds one more level of demand to an already arduous task. What makes this at all feasible as part of the course work is that the students themselves generate the items and serve as raters. Scale development takes place as an integral part of classroom and laboratory instruction over the course of several meetings. It requires sequential steps as outlined below, which conserve the instructor's time by using readily available computer technology.

**1.** Selection of the concept to be scaled. Strictly speaking this first step is not necessary, but in practice I find student interest and motivation greater when they have played an active part in choosing the concept. Since the scale will ultimately be used with a campus sample, we select concepts that are relevant to campus issues and concerns. We have typically scaled attitudes towards issues such as 'satisfaction with campus social life' and 'sexual fidelity within personal relationships'.

**2.** Item pool generation. Students are told to individually prepare five potential scale items. A class presentation regarding item preparation and selection follows, and then students in small groups critique each others' items checking for clarity, face validity, spelling, etc. Following the critique, students do any necessary revision of their individual items. When satisfied with their items, they use the campus computer network to e-mail their individual items to me. A lab assistant downloads the individual student files, strips the headings, and combines them into a single text file, a procedure which takes only about a half-hour.

**3.** Item ratings. The item file is circulated by e-mail so that the students can each rate the items according to instructions. These instructions must be very detailed as my experience indicates that this is the step most liable to misinterpretation. The point is made in the instructions as well as in class discussion, that when doing the ratings the rater is to impersonally rate each item in terms of how positive or negative the item appears to be, *not* the extent to which he or she agrees or disagrees personally with the item. The item pool also contains specific instructions as to how the resulting ratings are to be returned. Specifically, each student enters his or her ratings in a text file in a uniform format and e-mails this to me. The student assistant again downloads the files and combines them into a single master raw data file.

**4.** Statistical description of the items. Currently I create the system file and carry out the data analysis, because the available software for students permits the use of only twenty variables. In any event, this procedure is very fast and simple. The resulting means, medians, and standard deviations for each item are saved to a text file.

**5.** Item selection. The data analysis from the previous step is circulated to students. The lab assistant prepares a text file containing the mean and standard deviation for each item which is then uploaded and e-mailed to students prior to the next lab. During the lab students are given a short lecture on the criteria for item selection. Then each student individually goes through the item pool selecting the items they deem most appropriate for scale inclusion. Students had previously been told that prizes are available for items that are perfect, i.e., an item which receives identical ratings from all raters. These are extremely rare in practice, but the 'contest' serves not only to spark interest but also to provide a basis for discussion of difficulties in constructing a scale with good validity and reliability. The results of this exercise are shared, prizes if any are announced, and the results form a basis for a concluding discussion of the subjectivity inherent in the methodology.

Student response to this approach has been quite positive. Subsequent testing as well as evaluation of lab reports indicates to me that students have gained a relatively sophisticated understanding and appreciation of the strengths and weaknesses of psychometric scaling.

## "PERSONALITY ASSESSMENT ON TRIAL"

Bruce Bracken
The University of Memphis

*Integration with Text: Chapter 12*

I end my projective assessment class with a mock trial. Half the class defends personality assessment (including projectives) and the other half presents the case that personality assessment should be outlawed. Each "team" appoints an attorney, and each class member becomes an "expert witness" who specializes in one or more salient issues. I serve as the judge (over the years both sides have won, depending on how convincing or original the arguments were). The attorneys can call witnesses, "cross examine", etc. The exercise is much more interesting than having me lecture on the pros and cons of personality assessment. The students have presented the trial at state and regional conferences as well as the National Association of School Psychologists (NASP) and the sessions have been well attended. What I like most about the trial is that the students become better informed, learn both sides of the issue without my biases, and I believe are more "open minded" and tolerant of diverse ideas as a result.

## STORY STARTERS FOR DISCUSSION ON RACIAL DIFFERENCES IN IQ

*Integration with Text: Chapter 8*

Greg McMann of the University of Iowa found the following (offensive) quote from Terman which he uses to stimulate discussion related to issues presented in his discussion of racial differences in IQ. "Their dullness seems to be racial, or at least inherent in the family stocks from which they come. The fact that one meets this type with such extraordinary frequency among Indians, Mexicans, and Negroes suggests quite forcibly that the whole question of racial differences in mental traits will have to be taken up anew . . . which cannot be wiped out by any schemes of mental culture. Children of this group should be segregated in special classes . . . They cannot master abstractions, but they can often be made efficient workers" (Terman, 1916; cited in Bowles & Gintis, 1976, p. 123). An additional line to this quote that appeared in the original Stanford-Binet manual, no less, "There is no possibility at present of convincing society that they should not be allowed to reproduce, although from a eugenic point of view they constitute a grave problem because of their unusually prolific breeding" (Terman, 1916; cited in Kamin, 1974, p. 6).

Another quote summarizing the results of Ellis Island testing—circa 1919/20—focuses on the genetic, intellectual inferiority of Italians, Jew, Poles, "Mediterranean peoples."

Kamin's (1974) book describes a 1913 paper ("The Binet test in relation to immigration", *Journal of Psycho-Asthenics*, 18, 105–107) in which Goddard reported that "83% of the Jews, 80% of the Hungarians, 79% of Italians, and 87% of the Russians were feeble-minded" (Kamin, po. 16). Kamin cites many other sources of social and legal consequences of IQ tests in a 1975 article in the *Journal of School Psychology*, 13 (4), 317–323.

Some possible discussion questions, developed by Dr. McMann, to use with the above quotations include:

1) Given that Terman's views periodically have been echoed by others during the past eighty years (most recently, by Herrnsteim and Murray, authors of *The Bell Curve)* would you describe Terman's remarks in the 1916 Stanford-Binet manual as "visionary" or "racist"?

2) If the original intent of Terman was to devise a scale that would help segregate minorities into special education classes (in the interest of developing their potential as "efficient workers"), is disproportionate representation of minorities in special education classes merely an "unintended outcome" of the use of IQ tests for disability diagnosis?

Another exercise, suggested by Dr. Jeff Braden of the University of Wisconsin, is to divide the class up into pro and con, have them take sides, and then following the debate take the other side in their closing arguments. That might help students see the complexity of the issues, and appreciate that simply dismissing the work at "racist dogma" does a disservice to everybody.

*The following is reprinted from the 1997 issue of the Division Two Newsletter*

## MENTORING SERVICE ON TEACHING NOW AVAILABLE

The Society's Mentoring Service on Teaching is now available to provide support for individuals beyond that available in their own departments. The Mentoring Service is designed especially for beginning faculty members and graduate teaching assistants, but may be utilized by more experienced faculty as well.

A number of experienced individuals, all winners of the Society's Teaching Award, have been solicited to serve as teaching mentors. These individuals bring experience from different types of institutions and have different areas of expertise. Mentors can provide support and advice on the following issues: lecture style, syllabus writing, grade distributions, alternate methods of assessment, cheating, selecting textbooks, lab equipment, computer technology, professional development, time management and balancing responsibilities, interprofessional relationships, publishing materials on pedagogy, writing textbooks, and curriculum-related issues.

The Mentoring Service is administered by the Office of Teaching Resources in Psychology (OTRP) at no charge to Society members. If you would like to contact a teaching mentor, you will first need to request an information form from the OTRP. When the OTRP Executive Director receives your completed form, she will match you with a mentor based on your expressed needs. Once you receive the name of your mentor, you can contact the person by phone, e-mail, or snail mail at your convenience. Discussions between you and your mentor can continue as long as both of you find them productive!!

To request an information form, contact: Margaret A. Lloyd, Executive Director, Office of Teaching Resources in Psychology, Department of Psychology, Georgia Southern University, P.O. Box 8041, Statesboro, GA 30460; telephone: (912) 681-5423; e-mail: MLLOYD@GASOU.EDU.

Development of this service is due to the work of Barbara Nodine, Patricia Keith-Spiegel, Margaret Lloyd, Virginia Andreoli-Mathie, Dana Dunn, Tresmaine Grimes, Steve Hobbs, Steve Schiavo, and Stacey Zeremba.

## THE UNIVERSAL NONVERBAL INTELLIGENCE TEST: AN ASSESSMENT SOLUTION FOR AN EVOLVING POPULATION

Bruce A. Bracken
University of Memphis

As the 21st Century approaches and the U.S. population continues to evolve, psychologists need more than ever to cross language barriers if they are going to effectively serve students. In anticipation of and response to this pressing need, the Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, in press) was developed. The UNIT is intended for children between the ages of 5 through 17 years who speak English as a second language, or who are hearing impaired, have language-related learning disabilities, language-limiting psychiatric disorders (e.g., autism, selective mutism), or who would otherwise be disadvantaged if administered a traditional, language-loaded intelligence test.

Unlike many "nonverbal" tests that employ language-reduced instructions and performance tasks, the UNIT is totally and truly a nonverbal measure of intelligence. The UNIT is administered through the use of standardized gestures, task demonstrations, and unscored sample/teaching items. The UNIT was developed, first and foremost, as a comprehensive measure of global intelligence. Additionally, the UNIT was designed to assess four theoretically important facets of intelligence: reasoning, memory, symbolic processing, and nonsymbolic processing. A description of the six UNIT subtests and their scale assignment follows.

**Symbolic Memory.** The Symbolic Memory subtest contributes to the Memory and Symbolic Processing scales. The examiner displays a stimulus plate with an array of green and black universal figures depicting men, women, boys, girls, and babies. After a five-second exposure the examinee arrange response cards to match the stimulus sequence. Because of the multiple salient stimulus characteristics to be coded, recalled, and replicated, this task is a more complex measure of memory and intelligence than traditional short-term memory tasks (e.g., digit recall tasks).

**Cube Design.** This subtest is a three-dimensional adaptation of familiar block design tests found on the Kohs Block Designs or the Wechsler and Binet scales. On this subtest, the examinee attempts to construct a three-dimensional design to match a pictured stimulus on the top, front, and right visible sides. Block design subtests have long been held in esteem as sound measures of general intelligence; the UNIT Cub Design subtest advances that tradition in its three-dimensional format.

**Spatial Memory.** This complex memory task presents arrays of green and black "polka dots" positioned in various locations on a three-by-three or a three-by-four grid. The examiner presents a stimulus plate for a five-second exposure, after which the stimulus is removed from view. The examinee attempts to replicate the polka-dot placement, number, and color on a table-top response grid. The gestalt-like presentation of bicolored dots presents a task that is nonsymbolic in nature; that is, one that is not easily coded or mediated verbally. Spatial Memory contributes to the Memory and Nonsymbolic Scales.

**Analogic Reasoning.** Progressive matrices have had a psychometric tradition as robust measures of global intelligence. The UNIT Analogic Reasoning subtest provides a complex measure of reasoning by assessing conceptual and symbolically mediated pictorial analogies (e.g., Hat is to head as foot is to . . . ). This subtest contributes to the Reasoning and Symbolic Scales.

**Object Memory.** As another measure of complex short-term memory, this subtest presents a five-second pictorial display of easily recognized objects. The examinee is then shown a response plate that depicts all of the original objects, though in different locations, along with several distractors. The examinee is directed to place response chips on all of the objects that were presented on the stimulus plate. Performance on Object Memory is facilitated by symbolic processing (e.g., labeling, categorizing) of the stimuli, and thus the subtest contributes to both the Memory and Symbolic Scales.

**Mazes.** The Mazes subtest is an adaptation of historical maze tests (e.g., Porteus Mazes), and provides a nonverbal measure of reasoning and planning abilities. This UNIT subtest is uniquely scored. The examinee is credited with each consecutive correct decision made while traversing through the maze; once an incorrect decision has occurred, the examinee does not accrue any additional credit. This scoring system emphasizes examinee planning and careful decision making. The Mazes subtest contributes to the Reasoning and Nonsymbolic Scales.

Three forms of the UNIT were developed to accommodate psychologists' schedules and varied needs. The Abbreviated Form includes just two subtests (i.e., Symbolic Memory and Cube Design) and requires approximately 15 minutes to administer. The Abbreviated Form yields a Full Scale IQ, with an average internal consistency of .91 across the age span. A four-subtest Standard Battery (Symbolic and Spatial Memory and Cube Design and Analogic Reasoning) can be administered in approximately 30 minutes. The average internal consistencies of the Reasoning, Memory, and Full Scale scores for the Standard Battery are .90, .88, and .93, respectively. The Extended Battery, which requires approximately 45 minutes to administer, includes all six subtests. The average reliability of the Reasoning, Memory, and Full Scale scores on the extended battery are .86, .90, and .93, respectively.

In an era when an estimated 200 languages are spoken by students in the Chicago Public Schools alone, nonverbal tests and assessment techniques offer a promising method for assessing students' intellectual abilities in a fair and equitable manner. Similarly, for students who have language-related learning disabilities, psychiatric conditions, or hearing impairments, nonverbal instruments and assessment techniques may prove helpful. The UNIT offers psychologists a brief, yet comprehensive, reliable, and valid nonverbal measure of intelligence for clients who cannot be assessed, for whatever reason, through the use of traditional language-loaded measures of intelligence. As such, the UNIT is designed to meet the needs of an evolving U.S. population.

## RECENT ADVANCE IN INTELLIGENCE TESTING

Jack A. Naglieri
Ohio State University

*Integration with Text: Chapter 9*

"Understanding of Intelligence Has Changed Over Time; Old Standards Haven's" reads a recent issue of the *Wall Street Journal* (June 5, 1997). This headline brings out the important fact that the tests used to measure intelligence have remained locked in time since 1916 when the Stanford-Binet and 1939 when the Wechsler-Bellevue were published. These "standards" have now been revised three and four times, respectively, but still retain their essential elements as measures of general ability subdivided mainly by test content. In recent years, however, it has become apparent that a shift toward a test build on contemporary knowledge and emphasis on specific abilities should be considered.

The need to modernize intelligence tests was noted by Brody in his book *Intelligence* (1992). He stated that "our intellectual progress has [not] had a major impact on the development of tests of intelligence" (p. 355)—but this is only partially true. The progress made in psychology in the last 50 years has influenced development of new tests as suggested by Naglieri (1997). Important contributions have been made by (a) the Kaufman Assessment Battery for children and Kaufman Adolescent and Adult Intelligence Test (Kaufman & Kaufman, 1983; 1993); (b) The Woodcock-Johnson Test of Cognitive Abilities (Woodcock & Johnson, 1989); (c) The Differential Ability Scales (Elliott, 1990); and most recently, (d) the Cognitive Assessment System (CAS) (Naglieri & Das, 1997). In the remainder of this short paper I will focus on the CAS.

The CAS is the most recent tool designed to go beyond general ability and traditional IQ and move toward a modern view of intelligence. While recognizing the contribution and utility of tests like the Wechsler and Binet the CAS is based on the assumption that these methods can be significantly improved upon. To do so, Naglieri and Das (1997a) suggest that one should begin with a modern multidimensional theory of human ability, focus on basic cognitive processes shared by all persons regardless of their cultural or racial background, eliminate achievement (arithmetic, vocabulary, for example) from the measurement of ability, and show relevance to instruction/intervention. Our theory called PASS (Planning, Attention, Simultaneous, Successive) and the test we have built according to the theory (CAS) (Naglieri & Das, 1997b) is designed to achieve these goals.

The PASS theory is the result of the merging of both theoretical and applied psychology with roots in the work of A. R. Luria (e.g., 1966; 1973; 1980; 1982), the "most frequently cited Soviet scholar in American, British, and Canadian psychology periodicals" (Solso & Hoffman, 1991, p. 251). Using Luria's view of the functional organization of the brain, we suggest that intelligence can be reconceptualized as cognitive processes and defined as follows:

Human intelligence is based upon the four essential activities of Planning, Attention, Simultaneous and Successive (PASS) cognitive processes that employ and alter an individual's base of knowledge. Planning is a mental process by which the individual determines, selects, and uses efficient solutions to problems. Planning is responsible for the generation and control of activity, the utilization of attentional, simultaneous, and successive processes, and the retrieval and use of knowledge. This includes, intentionality, the development of plans of action, the evaluation of the effectiveness of these plans, regulation of the plans, verification, self-correction, and impulse control. Attention is a mental process by which the individual selectively attends to a particular stimulus and inhibits attending to competing stimuli. Attention is involved in all activities requiring the focus of cognition and becomes increasingly difficult when nontarget stimuli are more salient than the target stimuli. Simultaneous processing is a mental process by which the individual integrates stimuli into groups. The essence of simultaneous processing is the elements of the stimuli must be interrelated for successful completion regardless of the content. Successive is a mental process by which a person integrates stimuli in their specific serial order that forms a chainlike progression. The distinguishing quality of successive processing is that each element is only related to those that precede it and these stimuli are not interrelated.

The four PASS processes are operationalized in the twelve-subtest CAS (Naglieri & Das, 1997b). The CAS has four normed scales called Planning, Attention, Simultaneous, and Successive, following directly from the theory, and a Full Scale (all set at a mean of 100 and SD of 15). The scale is standardized on a representative sample of 2,200 children aged 5 through 17 years. Ample evidence of internal reliability and validity is provided in the CAS

*(Continued on page 10)*

Interpretive Handbook (Naglieri & Das, 1997a). For example, a study of the relationship between CAS and achievement was conducted using a representative group of children from the standardization sample. The results provided in the Interpretive Handbook show that the CAS correlated highly (.73) with achievement—considerably higher than traditional IQ tests that correlate approximately .50 with achievement (Brody, 1992). Other validity evidence provided suggests that the PASS scales are sensitive to the cognitive deficiencies of children with Attention Deficit-Hyperactive Disorder and Reading Disabilities and have relevance to intervention.

While the CAS approach based on PASS provides a broad view of ability that extends beyond that offered by general intelligence tests, it has the added benefit of having relevance to instruction and intervention. This was recently demonstrated, for example, in two published papers (Naglieri & Gottling, 1995, Naglieri & Gottling, 1997) involving the performance of children in mathematics computation. Both of these studies showed that children who were poor in planning improved considerably more than children who were good in planning when given an instruction that facilitated a planful and strategic approach to class work in math computation. That is, the characteristics of the children (high or low in planning) were relevant to predicting outcomes.

The results of initial work on the CAS suggest that the PASS theory may be a viable alternative to the standard measures of general ability. The specific abilities (processes) have relevance to differential diagnosis, the PASS processes are excellent predictors of achievement, and the broad scope of abilities measured is sensitive to the deficits of certain exceptional children. All these factors combine to make the transition from assessment to intervention more possible. They also suggest that because our understanding of intelligence has changed over time, it is appropriate to consider modern approaches to intelligence.

Jack A. Naglieri, Ph.D.
Ohio State University

## AN INTERVIEW WITH ALAN KAUFMAN

Michael F. Shaughnessy
Eastern New Mexico University

*Integration with Text: Chapter 10*

Recently, this writer interviewed Dr. Alan S. Kaufman. Dr. Kaufman is probably the most famous individual currently associated with IQ and achievement testing today. In this interview, which appears in the issue of *Educational Psychology Review* (1996, Vol. 8, No. 2), Dr. Kaufman discusses the various tests that he has constructed. These include:

The Kaufman Assessment Battery for Children (K-ABC)
The Kaufman Brief Intelligence Test (K-BIT)
The Kaufman Adolescent and Adult Intelligence Test (KAIT)
The Kaufman Functional Academic Skills Test (K-FAST)
The Kaufman Short Neuropsychological Assessment Procedure (K-SNAP)
The Kaufman Survey of Early Academic and Language Skills (K-SEALS)
The Kaufman Test of Educational Achievement (T-TEA)

In this interview, he discusses the current "state of the art" of testing and discusses portfolio assessment, the trend away from standardized testing and the trend toward inclusion, and the impact it will have on testing, evaluation, and assessment. He also provides an opinion about the Wechsler Intelligence Scale for Children-III and the Fourth Edition of the Stanford Binet. He also discusses specific tests for the mentally retarded and gifted and offers suggestions as to how we should best test and evaluate for these exceptionalities. Lastly, he reviews his work in the field, specifically regarding the K-ABC and culture-fair, culture-free testing.

Those who are interested in a copy of the interview can write to:

Dr. Michael F. Shaughnessy
Eastern New Mexico University
School of Education
Portales, NM 88130
or e-mail at: shaughnm @ e-mail.enmu.edu

The Kaufman Adolescent and Adult Intelligence Test (KAIT) is a relatively new test that provides a very viable alternative to the overly long, burdensome, and tiresome Stanford Binet and an alternative to the WAIS-R (which is in the process of being re-normed).

The KAIT can be given by experienced examiners in less than one hour. It is based on a fluid and crystallized model of intelligence and yields a Composite IQ.

There are six subtests in the core battery—Definitions, Rebus Learning, Logical Steps, Auditory Comprehension, Mystery Codes, and Double Meanings. There is also an expanded battery which is comprised of Famous Faces (a general information subtest), Memory for Block Designs, and a Rebus and Auditory Delayed Recall, based on two of the previously administered core battery subtests.

Drs. Alan and Nadeen Kaufman have attempted to provide a valid, reliable IQ measure that can be objectively scored and administered quickly. In today's stress-filled world where time is money, the KAIT is a viable alternative for schools, vocational centers, and medical facilities (the test also contains a Mental Status Subtest).

American Guidance Service publishes this test which contains both timed and untimed tests. Colleagues of mine have commented on the appropriateness of certain subtests—

*(Continued on page 12)*

## RESEARCH SAMPLER

An article by Steven A. Meyers (Roosevelt University) published in *Teaching of Psychology*, 1997, 24, 105–115 summarizes ways to increase individual student involvement in small-group activities for psychology classes.

The September 1996 issue of *GRE Data Views* concludes that little relationship exists between the personal statement, typically used by graduate and undergraduate admissions committee as an index of writing ability, and other indicators of writing ability. The correlation between the scores given by trained readers to personal statements and an essay was low (.15) indicating that the personal statement and essay were not measuring the same ability. This research was conducted as part of a study of the relationship between a number of nontest indicators of writing ability and the new GRE writing test scheduled to be introduced in 1999. Copies of this issue are available from Graduate Record Examinations, Princeton, NJ 08541.

*Integration with Text: Chapter 4*

Most faculty members have at one time or another been faced with the dilemma of resolving issues related to dual relationships with students. A recent article, "A Model for the Ethical Management of Faculty-Student Dual Relationships" by Marka Biaggio, Tana Lucic Paget, and M. Sue Chenoweth of Pacific University and published in *Professional Psychology: Research and Practice*, 1997, 28 (2), 184–189 discusses the nature of these dual relationships and provides guidelines for their ethical management.

Mark L. Mitchell of Clarion University of Pennsylvania has written a paper that describes a free Macintosh software program that assists students to distinguish between positive and negative correlations and understand differences between correlation coefficients of different sizes. Copies of the paper are available from Mark L. Mitchell, Department of Psychology, Clarion University, Clarion, PA 16214; e-mail: mitchel@mail.clarion.edu.

*Integration with Text: Chapter 4*

Technology Tools for Today's Campuses is available in the special projects section on the Horizon Home Page (http://sunsite.unc.edu/horizon). The goal of Technology Tools for Today's Campuses is to provide faculty with accounts of colleagues who have agreed to share with others their experiences in using technology in their classes. The seventy two articles have important information to help you decide if you want to use listservs, e-mail, the WWW in your teaching and also have valuable links to example syllabi, student papers written on the Web, and references. The Horizon Home Page will provide you with information on how you can obtain this publication free of charge in CD-ROM format from Microsoft Corporation.

T. B. Rogers of the University of Calgary published a recent article, "Teaching Ethics and Test Standards in a Psychological Testing Course: A Test Taker's Bill of Rights" in *Teaching of Psychology* (Vol. 24, No. 1, 1997, pp. 41–46). In this article, Dr. Rogers shares his Test Takers Bill of Rights which incorporates test standards and ethical principles into a format that is more easily read, understood, and evaluated by students. Copies of reprints are available from: Dr. T. B. Rogers, Department of Psychology, The University of Calgary, Calgary, Alberta, Canada T2N1N4; e-mail: tbrogers@acs.ucalgary.ca.

*Integration with Text: Chapter 2*

### GUIDE TO EVALUATING TESTS

Dr. Larry M. Rudner has written a very useful handout for use by students when evaluating tests entitled, "Questions to Ask When Evaluating Tests". The handout provides a very helpful summary of key measurement concepts (e.g. reliability, validity) and would be useful to students when completing any assignment involving evaluating psychological or educational tests. In fact, the Buros Institute provided this handout to all of their reviewers of tests for the most recent *Mental Measurement Yearbook*. Copies of the handout are available from: Buros Institute of Mental Measurements, 135 Bancroft Hall, University of Nebraska—Lincoln, Lincoln, NE 68588-0308.

*Integration with Text: Chapter 1*

**NEW ASSESMENT INSTRUMENTS** *(Continued from page 10)*

most particularly the Auditory Comprehension. In this subtest, the subject listens to a tape-recorded news story and is then asked specific questions about what has been heard. This is a "real life" type of task that reflects on real life situations, i.e., the ability to listen, concentrate, pay attention, and respond to what has been heard. The cassette tape provides for standardization and the authors strongly recommend that it be used rather than relying on one's own voice.

The test is in easel format and the protocol provides for easy simple scoring. In fact, experienced examiners can score (and in certain instances must score) as they proceed through the test.

For specific groups, for example, the mentally retarded, the test may not provide the wealth of information that is provided by other tests. For example, the WAIS-R may provide more information for vocational rehabilitation purposes. This writer has used the test with gifted, learning disabled, mentally retarded and "at-risk" children and would be happy to "compare notes" with other users and data with other users. This Measurement Forum should perhaps serve as a "clearinghouse" of sorts for this type of endeavor.

In sum, all things considered, the KAIT is a most welcome addition to a school psychologists group of assessment devices and a most viable alternative to the Stanford-Binet, and the Slossen and other less valid, less reliable instruments.

## MEASUREMENT MEETINGS

20th Annual National Institute on the Teaching of Psychology, January 3–6, 1998, St. Petersburg, Florida. Contact: Joanne Fetzner, (217)398-6969

Assessment '98 Assessment for Change—Changes in Assessment, January 16–18, 1998, St. Petersburg, Florida. Contact: ERIC/CASS Assessment '98, 1(800)414-97699 or (910)334-4114

First Annual National Conference on Assessment Psychology, April 1-4, 1998, Orlando, Florida. Contact: Dr. Alan Raphael, (305)372-0010

**MAYFIELD PUBLISHING COMPANY**
1280 Villa Street
Mountain View, CA 94041

Bulk Rate
US Postage
Paid
Permit No. 529
Fremont, CA