

Is Something Happening to Measurement Scholarship?

*Lawrence Rudner, Javonà Burke and Leland Rudner
ERIC Clearinghouse on Assessment*

One of Sir Isaac Newton’s most famous quotes is “If I have seen further, it is by standing on the shoulders of giants.” Do we, as measurement specialists, conform to this epitome of scholarship, scientific progress, and humility? Are we advancing knowledge by building on each other’s work? The authors of this piece uncovered some interesting trends with regard to how often measurement research is being referenced.

Using the Institute for Scientific Information’s Social Sciences Citation Index (SSCI) database, we examined the number of times articles in leading measurement journals were referenced by subsequent articles. The SSCI database provides access to current and retrospective bibliographic information and cited references found in 1,700 of the world’s leading scholarly social sciences journals. Using a three-year sliding window, we computed the average number of times articles in *Journal of Educational Measurement (JEM)*, *Educational and Psychological Measurement (EPM)*, *Psychometrika (Psych)*, *Applied Psychological Measurement (APM)*, and *Educational Measurement: Issues and Practice (EM:IP)* were cited in the SSCI database literature. For example, the 29 JEM articles published in 1980 were cited 270 times in 1980, 1981, and 1982 for an average of 9.3 citations per article.

Table 1: Average number of citations within 3 years per article in lead measurement journals

Year published	Year cited	<i>JEM</i>	<i>EPM</i>	<i>Psych</i>	<i>APM</i>	<i>EM:IP</i>
1980	1980–1982	9.3	3.3	29.2	–	–
1985	1985–1987	6.3	5.6	24.1	16.8	2.1
1990	1990–1992	8.3	3.6	14.3	9.2	1.8
1995	1995–1997	2.2	2.6	3.1	3.9	2.1
1996	1996–1998	3.1	4.1	3.7	6.6	5.1
1997	1997–1999	1.5	1.8	2.4	4.7	2.4
1998	1998–2000	1.2	1.8	2.0	2.5	1.3

The trends are very clear. According to the SSCI database, articles published in 1980, 1985, and 1990 were more frequently referenced within three years than articles published during the 1995–1998 interval. The exception is *EM:IP*.

The meaning of these trends is less clear. This may be a weakness in the SSCI database, a trend in education, or a trend in all fields. Perhaps, we are in a lull with regard to interesting questions and therefore there is less need to cascade knowledge. Maybe it is the opposite—there may be so many pressing questions that our research efforts have become more fragmented.

A cursory examination of recent articles suggests a change in the nature of the referenced articles, not in the number of cited articles. The referenced sources appear to be more varied and older. This suggest to us that (1) individual journals, especially *JEM*, are becoming less important as an outlet for scholarly work and (2) the articles themselves are not addressing contemporary assessment issues. In an unscholarly fashion, however, we leave the interpretation and conclusions to the readers. What do you think is happening to scholarship in our field?

MEET YOUR PRESIDENT – H.D. HOOVER

H.D. Hoover is a professor of statistics and educational measurement at The University of Iowa and the director of the Iowa Basic Skills Testing Program, a voluntary statewide testing program in grades K–8. He is the senior author of the *Iowa Tests of Basic Skills* and the *Canadian Tests of Basic Skills*. His research interests include test scaling, test equating, group differences in item and test performance, and the measurement of mathematics achievement. Some of the journals in which his research has appeared include the *Journal of Educational Measurement*, *Applied Psychological Measurement*, *Educational and Psychological Measurement*, *Measurement and Evaluation in Counseling and Development*, *Journal of Early Adolescence*, *Psychology of Women Quarterly*, *Journal for Research in Mathematics Education*, and *The Arithmetic Teacher*. He is a coauthor of the chapter on "Scaling, Norming, and Equating" in the Third Edition of *Educational Measurement* edited by Robert Linn. In addition to currently serving as president of NCME, H.D. serves on the advisory board of the Center for Research on Evaluation, Standards and Student Testing (CRESST), is a past chair of the Directors of State Testing Programs, and was once a junior high teacher.



NCME Paper Repository

The NCME Papers Repository is a place to present papers and presentations on the NCME website. Submitted papers and presentations must be approved by NCME before they are displayed on the publicly searchable repository at <http://www.ncme.org/repository/search.ace>. Please allow some time for NCME approval of a presentation for the file to be shown on the site.

Uploading Papers & Presentations: To upload papers to the repository for review by NCME, go to <http://www.ncme.org>, and click on the Annual Meeting section. Then from the menu at the left, select "Submit Papers & Presentations". To upload a paper, you will need to choose a username and a password, and enter them along with your email address, where indicated. Please make a note of your chosen username and password, as you will need this information if you wish to log in again and upload a new copy of your paper or presentation.

Once you have entered the papers repository, you may:

- submit a new paper or presentation, or
- update the file for a previously uploaded presentation.

To submit a new paper or presentation, click on "Submit a Paper or Presentation". Enter a title, the names of authors with first author at the top, and click "Browse" to select a file from

your computer to upload to the repository. Click "Open" or "Select", and the name of the file should show up in the window next to the "Browse" button. Lastly, select up to 3 "keywords" among the menus of topics at the bottom of the page. These "keywords" will be used for searches of the repository, so classify your presentation or paper accordingly. Click "Upload for Review" **once** to submit your paper or presentation. Some large files may take a few minutes to upload—this is normal.

Updating Previously Uploaded Papers or Presentations: To update a previously uploaded file, log in to the repository. You should see a list of "Your Recent Submissions". Click on "Upload New Copy or Delete" next to the file you wish to update. Change any information about the file on the next page. If you wish to upload a new file to replace this presentation, check the box labeled "Check to Re-upload document". If you wish to keep this file and only change author, title, or other information, do not check this box. If you are uploading a new file, select it with the "Browse" button. When you have finished entering the new information, click "Update Submission" **once**. You will be returned to the repository menu when the update is complete.

Forget your password and/or username? If you forget your username or password, you may go to this address: <http://www.ncme.org/repository/password.ace> and have the information emailed to you.

NEWSLETTER ADVISORY BOARD

BETTY BERGSTROM, Computer Adaptive Technologies, Inc.
GREGORY CIZEK, University of North Carolina, Chapel Hill
JOAN HERMANN, CRESST/UCLA
SHARON LEWIS, Council of the Great City Schools
DUNCAN MACQUARRIE, Tacoma Public Schools
KAREN MITCHELL, National Academy of Sciences
LORA MONFILS, Graduate Student – Rutgers University
S.E. PHILLIPS, Consultant
NAMBURY RAJU, Illinois Institute of Technology
LAWRENCE RUDNER, ERIC/University of Maryland
STEVE SIRECI, University of Massachusetts, Amherst
LISA F. SMITH, Kean University
JON TWING, NCS Pearson
DENNY WAY, Educational Testing Service

DOUGLAS F. BECKER, EDITOR, Riverside Publishing

Send articles or information for this newsletter to:

Douglas F. Becker Phone: (800) 767-8420, ext. 7006
Riverside Publishing Fax: (630) 467-7126
425 Spring Lake Drive e-mail: douglas_becker@hmco.com
Itasca, IL 60143

The *NCME Newsletter* is published quarterly. The *Newsletter* is not copyrighted; readers are invited to copy any articles that have not been previously copyrighted. Credit should be given in accordance with accepted publishing standards.

LEGAL CORNER: California ELL Case Settlement

S.E. Phillips, Consultant

The San Francisco, Oakland, Hayward, and Berkeley Unified School Districts joined together to oppose a state legislative requirement that an achievement test designated by the State Board of Education be administered annually to all students in grades 2–11. In particular, the districts objected to the requirement that English Language Learners (ELLs) with less than 30 months of public school instruction (ELLs<30) be required to take the state-designated test. Test scores were used for school accountability with scores for students enrolled for less than 12 months excluded from the computation of a school's accountability index. The designated achievement test was the Stanford Achievement Test Ninth Edition (SAT9) plus an additional set of items selected to measure state standards not measured by the SAT9.

The case began when San Francisco refused to administer the SAT9 to ELLs<30 unless specially designated by their teachers. The state sought a court order to enforce the testing legislation. At the time of the scheduled trial in fall 2000, ELLs in all districts except San Francisco (and Oakland in 1999) had three years of testing data for their ELL students.

The plaintiffs argued that administration of the SAT9 to ELLs<30 was unfair because the test measured English language skill in addition to content knowledge. They argued that students with limited English proficiency should either be tested in their native language or be exempt from testing. State law provided that ELLs<12 be administered a second achievement test in their native language when available. ELLs<12 were also eligible for test modifications when tested in English.

Plaintiffs' experts opined that ELLs<30 would suffer psychological harm from taking the SAT9 in English because their low scores would be stigmatizing, would diminish their self-esteem, and would cause them to be inappropriately placed in special education programs and portrayed as having inferior employment skills. In addition, plaintiffs' experts argued that ELLs<30 would score at the chance level, resulting in unreliable test scores. Plaintiffs also objected to rules prohibiting educators from encouraging ELL parents to invoke statutory provisions allowing them to exclude their children from testing.

In response, the state argued that a reasonable interpretation of state law indicated an intent to measure academic skills in English, that a fair accountability system requires the inclusion of all students, that the districts and children benefit from the receipt of state funds targeted toward the improvement of academic skills for low-scoring students, that the districts failed to show that any ELLs were harmed by the test administration, and that the data demonstrated that most ELLs scored above chance, and their test scores were reliable. In addition, the state argued that 30 months was an arbitrary exclusion criteria and that there was significant overlap in the performance of ELLs<30 and ELLs>30. Further, over the

three-year period in which the SAT9 had been administered statewide, ELLs had made substantial gains in some districts.

The case settled out of court just prior to trial. The districts agreed to administer the state-designated achievement test to all ELL students as provided by state law. The state agreed to clarify the rules regarding educator communications with parents about exemptions; consider, among other factors, English Language Development scores when considering school waiver requests; and make other minor modifications to program procedures.

Call for Manuscripts – JEM

The *Journal of Educational Measurement (JEM)* seeks manuscripts containing generalizable contributions to educational measurement. In keeping with the purposes of NCME, preferred topics are those likely to be of interest to persons concerned with the practice of measurement in field settings as well as measurement theorists. In addition to presenting new contributions to measurement theory, *JEM* is a vehicle for the improvement of measurement application in a variety of educational settings.

Manuscripts should be submitted in triplicate to: Barbara G. Dodd, Editor, *JEM*, University of Texas at Austin, Dept. of Educational Psychology, SZB 504, Austin, TX, 78712-1296. Two of the copies will be used for blind review and, therefore, should not include the author's name or institution. A more detailed description of the manuscript requirements can be found in the spring 2001 issue of *JEM* under Information for Contributors.

NCME Session on Test Accommodations

One highlight of the 2001 NCME annual meeting was an innovative session on test accommodations, subtitled "Waving a White Flag or Raising a Checkered One?" The session was organized by Gregory Cizek of the University of North Carolina–Chapel Hill and Douglas Becker of Riverside Publishing.

To begin the session, those in attendance were provided with a brief scenario that described a hypothetical examinee, named Chris, whose score report had been flagged to indicate that the test was taken under nonstandard conditions. The examinee suspected that the reason potential internship opportunities had failed to materialize was due to the flagging.

Four presenters at the session had been provided with the scenario in advance and asked to comment on the situation from various perspectives. Kurt Geisinger of LeMoyne College, who has provided expert testimony on behalf of examinees in similar situations, provided insights into logic and validity arguments relevant to flagging cases. Diana Pullin of Boston College summarized key legal principles that those who develop, administer, and use tests must consider as suggested by current statute and case law, and she commented

on some of the legal issues that remain unresolved. Rina Sjolund of ACT presented some of the operational and validity concerns that come into play from the perspective of a testing company. Finally, Gerald Tindal of the University of Oregon presented a three-prong framework for viewing accommodations that incorporated: understanding of the specific disability, the requirement of the Standards for Educational and Psychological Tests, and the relationship between specific accommodations and valid inferences from resulting test scores.

At the end of the session, those in attendance were asked to complete a questionnaire to gather their ideas about the flagging issue. The following presents a summary of the respondents (n=40) opinions.

- 1a) Should Chris have been permitted to take the FET under non-standard testing conditions?

Yes [39] No [1]

- 1b) If you answered "yes" to question 1a, which testing accommodations do you believe should be provided for Chris? (check all that apply)

Additional time [32]

Frequent breaks [40]

Individual administration [27]

Other [5]

- 2) Should Chris' test score have been reported with an indication that it was obtained under nonstandard conditions (i.e., "flagged")?

Yes [21] No [19]

- 3) What, if any, additional information about Chris' testing conditions should be provided to those who receive the score? (check only one)

No additional information [17]

A description of the testing conditions [7]

An evaluation of the validity of the score interpretation [4]

A comparison of examinees' performances under standard and nonstandard conditions [1]

Other [3]

In addition to the above, 3 respondents indicated **both** "A description of the testing conditions" and "An evaluation of the validity of the score interpretation" while 2 respondents indicated **both** "An evaluation of the validity of the score interpretation" and "A comparison of the examinees' performances under standard and nonstandard conditions."

Additionally, respondents were asked the following question: "In a sentence or two, what do you recommend as the best

strategy for assuring accurate interpretations of test scores obtained under non-standard conditions?" Only 11 responses were provided to this question. The responses (with n-counts in parentheses) could be categorized as revealing that respondents believed:

- Flagging is acceptable in its current state [2];
- Flagging should never be necessary on a power test [3];
- Differing kinds of accommodations exist, and only those that require different score interpretations should be flagged [3];
- More research is required to understand the relationship between test performance and real-life or on-the-job performance before flagging should be considered [3].

Clearly, the diversity of responses indicates the unresolved state of affairs with respect to accommodations. It seems equally clear that additional research and dissemination of best practices in the area of test accommodations represent opportunities for promoting validity and fairness in educational measurement.

New NCME Website Editor Sought

NCME seeks nominations for the position of Website editor. The Website is intended to disseminate information about NCME and its activities, and to serve as a resource and a means of communication for NCME members. The editor's primary role is to continue development of the site and to determine what materials to post, within the framework of NCME policy. This includes screening submissions from NCME members and posting information at the request of the Board. In addition, the Website editor implements special projects, which may be initiated by the Board or by the editor. For example, a depository for NCME conference presentations and a software "bulletin board" have recently been initiated on the site. The editor is assisted by an editorial board that he or she appoints, and receives guidance from the NCME Publications Committee and the NCME Board. The Website editor position requires editorial skills, Web skills, initiative, and creativity. The term for the editor is three years, beginning no later than January 2002. NCME provides funds to support the operation and maintenance of the Website.

By July 15, 2001, please relay nominations, including information about the nominee's qualifications, to Rebecca Zwick, Chair, NCME Publications Committee, Department of Education, University of California at Santa Barbara, Santa Barbara, CA, 93106-9490, or send email to: rzwick@education.ucsb.edu.

Call for Manuscripts – EM:IP

Educational Measurement: Issues and Practice seeks manuscripts and special issue proposals that illuminate issues in educational measurement and inform readers about the practice of measurement. Examples of manuscripts

appropriate for *EM:IP* include those dealing with specific measurement techniques for various educational objectives or controversial measurement issues. Articles on examples of good practices in educational measurement and articles presenting contrasting views concerning the educational role of testing are welcome. If you are unsure of the relevance of your paper to *EM:IP*'s mission, please send an email to the editor, Jeri Benson, at jbenson@coe.uga.edu, and she would be happy to discuss your paper.

Council Releases Report

Striving for Excellence: A Report on Stanford Achievement Test Results in the Great City Schools is a comprehensive examination of student achievement in urban school districts using the SAT-9. The Council of the Great City Schools (CGCS) conducted this study in collaboration with Harcourt Educational Measurement.

This study was conducted to provide (1) a snapshot of student achievement in the nation's urban schools, (2) an account of how urban students perform on a traditional standardized test, (3) a way for urban districts to compare themselves, and (4) the first step of a longitudinal analysis of urban student progress.

Striving for Excellence examines the achievement of over 760,000 urban test-takers in grades 2 through 11 on eight subtests – Total Reading, Reading Vocabulary, Reading Comprehension, Language, Total Math, Problem Solving, Math Procedures, and Science. SAT-9 data in this report were disaggregated by race/ethnicity, gender, and poverty.

It is important to note that the demographic composition of test-takers in the Great City Schools (GCS) is very different from test-takers nationally. Approximately six out of ten test-takers in the national sample were White, while six out of ten test-takers in the GCS were African American or Hispanic. In addition, about two-thirds of students in the GCS were eligible for a free lunch, compared with about one-third nationally.

The report presents the data in two ways: by NCE and by performance level. Both ways are used to present data by grade (2–11) and by subtest. Only grades 4, 8, and 10 are used to report racial, gender, and poverty data. Selected grades were used in order to simplify and shorten the report.

There was encouraging news in a number of areas. First, mean NCEs in the GCS hovered around 42 and 45, scores well within the average range. Second, White test-takers in the GCS not only outperformed all other urban test-takers but also scored higher than the national NCE mean, which is 50. This finding is consistent with other Council research. The organization's examination of ACT assessment data, for instance, showed that White students in the GCS outscored all other test-takers nationally. The finding indicates that urban schools can and do educate students to the high standards in some circumstances.

Third, there is evidence in the report that the achievement gap by race is no worse in the GCS than it is nationally. Finally,

there was no evidence among the students in this study that academic achievement gets worse the longer they remain in school.

The discouraging news is that achievement gaps between White and non-White urban test-takers are significant. Overall, reading scores were lower than math scores and achievement gaps between White and non-White students were wider for reading than for math. African American students scored lower than White students, and Hispanic students scored lower than African American students. Achievement gaps between White and Hispanic students were wider than achievement gaps between White and African American Students.

The results were also predictable with respect to the effects of poverty on student performance. The greater the concentration of poverty in the school district, the lower the student achievement. Urban students in high poverty districts were more likely to perform below the basic level than students attending schools in districts with moderate concentrations of poverty.

This report also allows one to see similarities and differences between results reported by NCE and by performance level. NCE results cluster more around the mean while performance data show more variation. The percentage of test-takers in the GCS, for instance, scoring below the basic level ranged from 31 to 77 percent. The percentage of test-takers nationally scoring below the basic level ranged from 26 to 75 percent.

There were nearly 800,000 test-takers in the GCS included in these analyses. These students come from a relatively small number of urban school districts. Caution should be exercised accordingly when reading this report and examining its findings and conclusions.

Striving for Excellence does not contain many surprises, but it does provide an objective examination of student achievement in the nation's urban schools using comparable data. While *Striving for Excellence* highlights what we already know about the correlation of race/ethnicity and poverty to student achievement, the data provide a baseline for improvement. This report is a part of a series of studies that will track improvements by urban districts as they work to increase overall student achievement and eliminate achievement gaps.

Editor's note: The preceding was an executive summary of a report recently released by the Council of the Great City Schools. Authors included: Michael Casserly, Executive Director; Sharon Lewis, Research Director; Nicole Baker, Research Specialist; and Jack Jepson, Research Specialist. For additional information on this report and others, contact:

Sharon Lewis
Research Director
Council of the Great City Schools
1301 Pennsylvania Avenue NW
Suite 702
Washington DC 20004
(202) 393-2427 phone
(202) 393-2400 fax
Website: <http://www.cgcs.org/>

CAT, Inc. Announces Research Grant

Computer Adaptive Technologies (CAT) is sponsoring a limited number of research grants. The purpose of the grants is to encourage research related to Internet-based testing.

Each grant recipient will receive a limited license to use the CAT*Software System*[™] for a period of two years. This technology allows Internet-based item banking, test publication, and test administration. The limited license will allow publication of up to 20 tests and 5,000 test administrations via the CAT Browser-Based Test Administrator over the two-year period. At the end of the grant period, the researcher is requested to submit a written report of the research findings to Computer Adaptive Technologies.

Applicants must be faculty or affiliates (e.g., graduate students or clinicians) at a college or university. The grant will be made to the principal investigator through his or her academic institution. To apply for a grant, the

principal investigator must submit a letter of intent, a proposed research plan (no more than five pages), and any supporting documentation.

Submit completed grant applications to:

Betty Bergstrom, Ph.D.
Vice President, Program Management and
Psychometric Services
Computer Adaptive Technologies, Inc.
bbergstrom@catinc.com

Grants are competitive. CAT retains the rights to any suggestions for improvement in our software and must be appropriately credited in any research publications.

CAT Corporate Worldwide Headquarters
1007 Church Street, 7th Floor, Evanston, Illinois 60201
Phone: (847) 866-2001, Fax: (847) 866-2002
Toll-free (in North America): (800) 255-1312
Website: <http://www.catinc.com/>
