

Expert Review: Validity Generalization & Title VII

The goal of this article written by Dan Biddle, Ph.D. and Patrick Nooren, Ph.D. is to review Title VII requirements for establishing validity evidence, overview federal and professional requirements for validation strategies (specifically VG), outline how some courts have responded to VG strategies, and conclude by providing recommendations for validating tests that come under Title VII scrutiny. To see the Labor Law Journal article in its entirety, visit the Biddle Consulting Group, Inc. website at www.biddle.com/vg.stm.

The 1991 Civil Rights Act requires employers to justify tests with disparate impact by demonstrating they are sufficiently “job related for the position in question and consistent with business necessity.” This requirement is most often addressed by conducting validation studies to establish a clear connection between the abilities measured by the test and the requirements of the job in question.

Building a validation defense strategy in such situations requires employers to address the federal Uniform Guidelines on Employee Selection Procedures (1978), professional standards, and relevant court precedents. In recent years, some employers have attempted to “borrow” validation evidence obtained by other employers for similar positions rather than conduct their own local validation study. This strategy relies on a methodology known as “validity generalization” (VG). Despite the increase in popularity among test

publishers and HR/hiring staff at corporations, relying entirely on VG to defend against Title VII disparate impact suits will likely lead to disappointing outcomes because the courts have generally required employers demonstrate local and specific validation evidence where there is local and specific evidence of disparate impact.

OVERVIEW OF TITLE VII DISPARATE IMPACT DISCRIMINATION

The 1991 Civil Rights Act states disparate impact discrimination occurs when “. . . a complaining party demonstrates that a respondent uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin, and the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity.”¹ Disparate impact occurs when two groups have substantially different passing rates on a test, and is normally evaluated using tests for both statistical (i.e., whether the differences in passing rates are beyond what would be expected by chance) and practical significance (the practical impact or stability of the findings). When tests have such disparate impact, a finding of unlawful discrimination will likely be the judgment, absent an acceptable demonstration of the “job relatedness”

of the test.

The basic necessity of providing “job relatedness” evidence for the test causing disparate impact has been set in stone since the famous U.S. Supreme Court Griggs v. Duke Power² case. However, during a two year period between 1989 and 1991, under the then-reigning U.S. Supreme Court Wards Cove v. Atonio³ case, this standard was lowered. Under the Wards Cove standard, employers only needed to “produce a business justification.” “Producing a justification” is a much less stringent requirement than “demonstrating job relatedness.” Congress overturned this standard in 1991 with the passage of the 1991 Civil Rights Act, which reinstated the original Griggs standard (where it stands today).

Fundamental elements from the Griggs case were encapsulated into the federal treatise to enforce Title VII—the Uniform Guidelines on Employee Selection Procedures, a document jointly developed in 1978 by the U.S. Equal Employment Opportunity Commission, Department of Justice, Department of Labor, and the Civil Service Board, now the Office of Personnel Management (discussed in more detail below).

While the Uniform Guidelines have remained unchanged since 1978, the courts have continued to support one very important component: when an employer uses a specific test for a

...see VG page 2

Inside...

Validity Generalization	1
Recordkeeping	2
New Release: AutoAAP v11	5
Event Calendar	6
Training Calendar	7

The Importance of Proper Recordkeeping in the New OFCCP Era

For several years now, the OFCCP has been touting change in the various ILG and national meetings around the country. For years they were saying that they wanted to focus on systemic disparities within compensation and selection processes such as hiring, promoting and terminating staff. More recently they followed through on their promise to place statisticians in the various OFCCP regions throughout the country with Dr. Michael Sinclair and Dr. Richard Fisher being the lead team out of Washington, D.C.

With the addition of the statisticians, the OFCCP has implemented an extremely aggressive audit campaign over the last two years. In the 2000-2003 timeframe OFCCP was claiming to have collected in the range of \$20-\$25 million per year from Federal contractors. In 2005 they collected more than \$45 million and in 2006 they exceeded \$50 million. What does all this mean to the contractor community? It means that the days of generic audits focused on Placement Goals and Narratives are over and that poor recordkeeping has become an expensive proposition.

Now that OFCCP audits are being driven by statistical analyses, it is necessary to learn the pros and cons of the new regulations and guidelines, including - ***Obligation To Solicit Race and Gender Data for Agency Enforcement Purposes - Office of***

Federal Contract Compliance Programs (Definition of an Applicant); and ***Voluntary Guidelines for Self-Evaluation of Compensation Practices for Compliance With Non-Discrimination Requirements of Executive Order*** (Compensation Analysis Guidelines). In many cases new regulations are viewed as a disadvantage to employer's efforts. However, both of the new initiatives should be viewed as an opportunity for contractors to manage their data more effectively and create more accurate analyses to provide a more accurate picture of a contractor's hiring and pay practices. Contractors now have significantly more influence over what data is used in their required analyses than ever before.

With the new regulations in effect, contractors need to take a close look at who is considered an applicant. In the past, the very broad definition of "anyone who expresses interest" in a job would be used in an Adverse Impact Analysis against hires. The new definition provides contractors with four key filters that define who is an applicant in today's market. Below are the four items that meet the Definition of an Applicant.

1. The individual submits expression of interest through Internet or related electronic data technologies
2. The contractor considers the individual for employment in a particular

position

3. The individual's expression of interest indicates basic qualifications for the position

4. The individual does not remove him/herself from consideration prior to job offer

In the new world of recordkeeping more and more contractors are becoming familiar with the use of Disposition or "Status" codes to help track the status of an applicant during the hiring process. Disposition codes allow the user to analyze for possible Adverse Impact in the various stages of the hiring process. Below is an example of status codes that a hiring manager could use in their system to track the status of an applicant as they progress through the hiring steps to show where they are in the process and ultimately if they were selected or rejected. In this example we list the group (Clerical) and the Code that states the current status for four different applicants. By keeping this information, a contractor would be able to conduct an analysis of who was selected to continue or who was rejected at any given stage in the process.

A Disposition Code or Status Code would be unique to each contractor and each job that they hire into. There should be a code identifying the job level and

...see Recordkeeping page 4

Staff Contributions

Contributors Dan Biddle, Ph.D.
 Patrick Nooren, Ph.D..
 Chris Lindholm
 Christine Anthony
 John Piatt
 Production C. Lori Lee
 Editing Fred Huppert
 Website www.biddle.com

VG

continued from page 1

particular job, and such test has disparate impact, the employer must justify the use of the test by demonstrating that the test is job related. This is because Title VII requires a specific justification for both the test itself as well as how it is being used (e.g., ranked, banded, used with a minimum cutoff, or weighted with other selection procedures) in specific situations where disparate impact exists.

OVERVIEW OF VALIDITY GENERALIZATION

Meta-analysis is a statistical technique used to combine the results of several related re-search studies to form general theories about relationships between variables (e.g., tests, job

...see VG page 3

performance) across different situations. When meta-analysis is applied to tests and job performance in the personnel testing field, it is referred to as VG. While the specific procedures involved in conducting a VG study may vary, the primary reason for conducting VG studies in an employment setting is to evaluate the effectiveness (i.e., validity) of a specific personnel test or type of test (e.g., cognitive ability, personality) and to describe what the findings mean in a broader sense. To accomplish this, a series of validation studies are combined and then various corrections are made to determine the overall operational validity of the test or type of test, with the intent to ascribe universal effectiveness of the test in different situations and/or locations.

To understand VG, some basic statistical concepts need to be introduced. The most integral element to a VG study is a validity coefficient, which is a statistical measure that indicates the strength of a correlation between a certain test and a given job performance criteria (e.g., supervisory ratings). Statistical correlations occur between two variables when high values on one variable are associated with high values on the other variable (and low with low, etc), and range in value between 0 (no correlation) to 1.0 (perfect correlation). In the personnel testing field, correlations that are .35 and higher can be labeled “very beneficial,” correlations ranging from .21 to .35 are “likely to be useful,” those ranging from .11 - .20 are labeled as “depends on circumstances,” and those less than .11 are branded “unlikely to be useful.”

Regardless of the size of the validity coefficient (e.g., .15 or .35), it needs to be “statistically significant” beyond a 5% level of chance to be “valid” in a Title VII situation (a requirement also

adopted by federal and professional standards), and this determination depends on the sample size involved in the study (with higher validity coefficients required for smaller studies). For example, a coefficient of .20 with a sample of 69 has a corresponding statistical significance probability value (referred to a “p-value”) of .0496 (using a one-tail test for significance), which could be argued as defensible under Title VII. However, the same coefficient of .20 with a sample of only 68 has a resulting probability value of .051, which is not statistically significant (because it exceeds the .05 threshold needed for labeling the finding as a “beyond chance occurrence”).

Another statistical concept that is important for understanding VG is statistical power. In a practical sense, statistical power refers to the ability of the study to find a statistically significant finding if it exists to be found. Validity studies that have large sample sizes (e.g., 500 subjects) have high statistical power, and those with small samples have low

statistical power. For example, assume that a personnel researcher wanted to find out if a certain test had a validity coefficient of .25 or higher, and there were only 80 incumbents in the target position for whom test and job performance data was available, they could be about 73% confident (i.e., have 73% power) of finding such a coefficient (if it was there to be found). With odds of about 3 to 4, the researcher has a “decent shot” at finding validity. With twice the sample size (160 subjects), power would increase to about 94%, which would provide the researcher a near certain ability to find out whether the test was valid at that particular location. And, if the researcher conducts such a study and finds no validity (by obtaining a coefficient that was not statistically significant), they would be comfort-able in concluding that validity did not exist at that location, or was sufficiently suppressed by statistical artifacts.

The issue of statistical power frames a problem with personnel researchers that VG attempts to address. By rolling up and combining several independent studies, VG attempts to cast a vision of the “big picture” of what validity for that test might look like over various situations (with some including small samples). Consider the sample VG data in Table 1.

In these sample data, the average sample size was about 134 subjects, yielding about 90% statistical power (on average) for each study to detect a validity coefficient of about .25 in each respective local situation. Notice that 12 of the 22 studies (over half) showed no validity (i.e., had corresponding probability of less than .05 in local settings). Eight (8) studies had correlations that

**TABLE 1
SAMPLE VALIDITY GENERALIZATION RESULTS**

Study #	Validity Coefficient	Sample Size	Power (I-tail)	p-value	Valid?
1	0.030	120	87%	0.37	No
2	0.135	130	89%	0.06	No
3	0.180	140	91%	0.02	Yes
4	0.290	150	93%	0.00	Yes
5	0.340	120	87%	0.00	Yes
6	0.180	130	89%	0.02	Yes
7	0.150	140	91%	0.04	Yes
8	0.110	150	93%	0.09	No
9	0.090	120	87%	0.16	No
10	0.126	130	89%	0.08	No
11	0.210	140	91%	0.01	Yes
12	0.390	150	93%	0.00	Yes
13	0.198	120	87%	0.02	Yes
14	0.164	130	89%	0.03	Yes
15	0.109	140	91%	0.10	No
16	0.094	150	93%	0.13	No
17	0.020	120	87%	0.41	No
18	0.114	130	89%	0.10	No
19	0.164	140	91%	0.03	Yes
20	0.070	150	93%	0.20	No
21	0.010	120	87%	0.46	No
22	0.010	130	89%	0.46	No

Recordkeeping

continued from page 2

each step in the hiring process allowing for consistent use of the identifier for every staff member that would enter or select that code in the tool where data is tracked. As the applicant progresses through each stage, the code would change to reflect their status. If an individual applicant is selected out of the process, the code would identify that they failed that particular step allowing for a hiring manager to know who passed and failed each step in the process.

Examples of Disposition or Status Codes:

1. **CL - MQ** (Clerical - Did not meet Minimum Qualifications)
2. **CL - PT** (Clerical - Passed Test)
3. **CL - PI** (Clerical - Passed Interview)
4. **CL - R** (Clerical Failed 2nd Interview/Rejected)

In the Table A you will see that the selection rate of females compared to

males is 66% which is less than the expected 80% threshold and is therefore considered a red-flag by the OFCCP.

If the only data maintained was the overall Hires or Applicants data, without identifying where applicants are in the process, then according to the Table A there would be Adverse Impact and possible discrimination in the eyes of the Department of Labor and possibly in any civil suit brought forth by potential applicants.

Collecting data on employment procedures is required under the *Uniform Guidelines on Selection Procedures* and it just makes sense to always track detailed information on company selections in order to minimize scrutiny. If Disposition Codes are tracked then it would be possible to isolate where Females were dropping out. Table B indicates where females may have passed the basic qualifications and maybe even an interview but failed

the test at a higher rate than males. If this information isn't tracked, how would one know that the test stage is where the failing rates of females needs to be analyzed?

If Females are dropping out in the Interview or Test process then an analysis can be run of who was selected vs. rejected at that individual step and try to explain why (see Table C). If Females are dropping out in the test, the test needs to be validated. If the test is validated then can be proven that there is no discrimination even if the summary results say there might be. If a test cannot be validated, then the Department of Labor will likely come looking for a settlement that could include back pay with interest for Females. If the Disposition/Status Codes are never collected, how could a defense be built or a proactive analyses conducted to ensure that there isn't any discrimination? ❌

TABLE A

Overall Selections	Pass (Hired)	Fail (Not Hired)	Applied	Pass Rate	80% Test
Male	60	20	80	75%	
Female	10	10	20	50%	50/75 = 66%
	70	30	100		

TABLE B

Applicant Gender	Applicant Race	Disposition Code	Final Status
Male	White	CL - PI	Rejected
Female	Black	CL - PI	Rejected
Male	Hispanic	CL - R	Considered
Female	Asian	CL - R	Hired

TABLE C

Interviewed Passed/Selected	Passed Test	Failed Test	Took Test (Total)	Pass Rate	80% Test
Male	60	20	80	75%	
Female	10	10	20	50%	50/75 = 66%
	70	30	100		

AutoAAP® 11 Released

Biddle Consulting Group, Inc. (BCG) is proud to announce the release of AutoAAP® Version 11. Over two years in development, AutoAAP 11 brings BCG and our clients into the second generation of our Microsoft .Net product - featuring an updated process for developing AAPs in addition to dozens of new features and reports.

The AutoAAP platform has undergone many changes over the years, beginning with its roots as a DOS application over 25 years ago, all the way up to the latest Microsoft standards. Over 300 clients will benefit from the new desktop or client server installation.

In version 11, clients can utilize new tools for correcting data flaws, importing multiple plan structures in a single import, view detailed census data, generate summary reports for the Workforce, Job Group and Placement Goals and even compile the entire Narrative and numerical reports into a single Adobe .PDF file.

BCG will be conducting a series of training sessions for existing and prospective software users. In 2007, a series of free, brief online training sessions (webinars) are being offered, in addition to our in-depth 2-day training classes held periodically at the BCG headquarters in Folsom, CA. Check our Training Calendar for sign-up opportunities for the planned sessions to be held in 2007.

Affirmative Action Plan Development Software



Key Features


- **Expanded Database** – The new version offers more data fields including five organization levels where there were three previously
- **Job Codes and Job Titles** – Import Job Codes associated with Job Titles for better data reconciliation
- **Organization-Level Narrative** – Now you can access the master narrative and make changes right in the software
- **Drag and Drop Import** – You can now drag and drop your import fields
- **New Data Checks** – Now you can check and fix on-screen data right after you import
- **Import Data Once** – You can create establishment, functional and roll-up plans from one import
- **Create All Plans in One Click** – In the Create Plans screen you can link all Selection Site values into AAPs with one click
- **New Job Screen** – Manage all information associated with a job title on one screen
- **Import a Job File** – You can now import a list of Job Titles and all the corresponding information to match employee and transaction files
- **Enhanced Zip Code**
- **Analysis** – Develop overall Zip Code Analyses as well as Analyses by individual Job Groups
- **Feeders using History Data** – Use the new automated function to generate Feeders from imported Promotions data
- **New Tests** – Combine the Whole Person and 80% Rule into one utilization test or create an Exact standard deviation and a Chi-Square standard Deviation
- **New Reports** – We now offer a Utilization Summary and Employee Data report
- **Adverse Impact Analysis** – There is a new Applicants vs. Hires query as well as more test options
- **Compensation Analysis** – The new report lets you select a percentage difference to test for significant differences in pay
- **New Plan Checks** - At the end of the plan process users can run checks to ensure that all plan steps are complete
- **Time Stamps** – All reports now have the option to show a current date and time stamp
- **New Compile Plan Function** – Compile narrative and reports into a single Adobe .PDF for a complete electronic AAP



VG

continued from page 3

would be considered too low ($< .11$) to be acceptable in litigation settings. The average validity coefficient across the 22 studies is about .15, which is just barely above the level needed to be statistically significant at the .05 level. However, when these studies are combined into a VG analysis and various corrections are applied, this average validity coefficient increases to between .24 and .48 (based on the type of corrections applied assuming typical reliability estimates and range restriction values). Due to these upward corrections, VG analyses estimate the level of validity that might be found absent the suppressive factors that negatively impact validity studies.

Unfortunately, while these “corrected” VG studies can often offer researchers useful insights into the strength of the relationship between the test and job performance in the studies included in the VG analysis, there is no guarantee that employers would find the level of validity promised by the result of a VG study if a study was performed in a new local setting. This is primarily because a host of situationally-specific factors exist in each and every new situation that may drastically impact the validity of a test. In addition, there are a number of issues with typical VG studies that may further limit their relevance and reliability when ascribing test validity into new situations. 

To see the Validity Generalization vs. Title VII: Can Employers Successfully Defend Tests Without Conducting Local Validation Studies? article from the Labor Law Journal, Winter 2006 issue, please visit the Biddle Consulting Group, Inc. website at www.biddle.com/vg.stm and click on the link to Dan Biddle and Patrick Nooren’s article at the bottom of the page. The article highlights relevant court cases, tables, citations, references to the Uniform Guidelines, Joint Standards, and SIOP Principle, recommendations, and conclusions.

2007 Events Calendar

<i>Conferences</i>	<i>Presentations</i>
<p style="text-align: center;">Biddle Consulting Group, Inc. will be a corporate sponsor at the following conferences and will have a vendor booth at each show listed below.</p> <p style="text-align: center;">: February 27-28 : <i>Western Region APCO Conference Long Beach, CA</i></p> <p style="text-align: center;">: April 23-25 : <i>Navigator Conference Las Vegas, NV</i></p> <p style="text-align: center;">: April 30-May 2 : <i>Association of Legal Administrators (ALA) Annual Conference & Expo Las Vegas, NV</i></p> <p style="text-align: center;">: June 11-12 : <i>National Emergency Numbers Association (NENA) Annual Conference Charlotte, NC</i></p> <p style="text-align: center;">: August 6-8 : <i>APCO International Conference & Expo Baltimore, MD</i></p> <p style="text-align: center;">: August 21-24 : <i>National Industry Liaison Group (NILG) Conference New York, NY</i></p> <p style="text-align: center;">: September 30-October 2 : <i>IPMA-HR Annual Conference Chicago, IL</i></p>	<p style="text-align: center;">: March 2 : <i>Seattle Industry Liaison Group</i> Topic: Compensation Analyses and the OFCCP <i>Speaker: Patrick Nooren, Ph.D., Executive Vice President Biddle Consulting Group, Inc. Seattle, WA</i></p> <p style="text-align: center;">: March 9 : <i>Idaho Industry Liaison Group</i> Topic: Establishing Basic Qualifications <i>Speaker: Dan Biddle, Ph.D., President Biddle Consulting Group, Inc. Boise, ID</i></p> <p style="text-align: center;">: April 26 : <i>American Association for Affirmative Action (AAAA) Annual Conference</i> Topic: Adverse Impact Analyses Across Individual Selection Decisions: The Mantel-Haenszel & Breslow-Day Procedures <i>Speaker: Patrick Nooren, Ph.D., Executive Vice President Biddle Consulting Group, Inc. Austin, TX</i></p> <p style="text-align: center;">: May 16 : <i>South Carolina ILG</i> <i>Speaker: Dan Biddle, Ph.D., President Biddle Consulting Group, Inc. Columbia, SC</i></p> <p style="text-align: center;">: August 21-24: <i>NILG Conference</i> <i>Speaker: Dan Biddle, Ph.D., President</i> Topics: TBD <i>Biddle Consulting Group, Inc. New York, NY</i></p>

For more information on any of the conferences or other events listed, or to have BCG speak at your event, please call Biddle Consulting Group, Inc. toll-free at (800) 999-0438 or e-mail us at staff@biddle.com

2007 Training Calendar

<i>Webinars</i>		<i>Seminars</i>
<p><i>AutoAAP 11.0 Training</i></p> <p style="text-align: center;"><i>AAP Development Using AutoAAP 11.0</i> Cost: Free</p> <p style="text-align: center;">: March 2 : : March 9 : : March 23 : : April 6 : : April 20 :</p> <hr/> <p><i>Compensation Guidelines: Update</i></p> <p style="text-align: center;">: June 12 : <i>Update on OFCCP Compensation Guidelines</i> Cost: Free</p> <hr/> <p><i>Definition of an Applicant</i></p> <p style="text-align: center;">: May 22 : <i>Definition of an Applicant: The Changing Landscape</i> Cost: Free</p> <hr/> <p><i>EEO-1 Changes</i></p> <p style="text-align: center;">: February 15 : <i>The New EEO-1 Changes and How It Affects Your Company</i> Cost: Free</p>	<p style="text-align: center;"><i>Compensation Analysis: How To</i></p> <hr/> <p style="text-align: center;">: March 14 : <i>Completing OFCCP-Compliant Compensation Analyses Using COMPARE</i> Cost: Free</p> <p style="text-align: center;">: March 28 : <i>Conducting Internal Compensation Analyses</i> Cost: \$99</p> <p style="text-align: center;">: June 13 : <i>Completing OFCCP-Compliant Compensation Analyses Using COMPARE</i> Cost: Free</p> <p style="text-align: center;">: June 27 : <i>Conducting Internal Compensation Analyses</i> Cost: \$99</p> <p style="text-align: center;">: September 12 : <i>Completing OFCCP-Compliant Compensation Analyses Using COMPARE</i> Cost: Free</p> <p style="text-align: center;">: September 26 : <i>Conducting Internal Compensation Analyses</i> Cost: \$99</p> <p style="text-align: center;">: December 3 : <i>Completing OFCCP-Compliant Compensation Analyses Using COMPARE</i> Cost: Free</p> <p style="text-align: center;">: December 10 : <i>Conducting Internal Compensation Analyses</i> Cost: \$99</p>	<p style="text-align: center;"><i>AAP Training Workshops</i></p> <p style="text-align: center;"><i>Two (2) Day Seminars AAP Methodology & Software Training Folsom, CA</i> Cost: \$990</p> <p style="text-align: center;">: April 25-26 : : June 19-20 : : September 25-26 : : December 4-5 :</p> <hr/> <p style="text-align: center;"><i>Regional Seminars</i></p> <p style="text-align: center;">One (1) Day Seminars Systemic Discrimination and the Transformed OFCCP: Navigating the Changing Landscape of Affirmative Action Compliance</p> <p style="text-align: center;"><i>presented by:</i> Biddle Consulting Group, Inc. Jackson Lewis LLP</p> <p style="text-align: center;">Cost: \$595 (Early Bird \$495)</p> <p style="text-align: center;">: April 17 - New York, NY : : April 18 - Arlington, VA : : April 20 - Atlanta, GA :</p> <p style="text-align: center;">: May 1 - Los Angeles, CA : : May 2 - San Francisco, CA : : May 3 - Seattle, WA :</p> <p style="text-align: center;">: October 10 - Denver, CO : : October 11 - Dallas, TX : : October 12 - Chicago, IL :</p> <p style="text-align: center;">Register Today: www.biddle.com/registration</p>

For more information on any of the seminars or webinars listed, please call Biddle Consulting Group, Inc. toll-free at (800) 999-0438 or e-mail us at staff@biddle.com