TITLE VII: DISCRIMINATORY USE OF TEST SCORES WATCHDOG

Evolution of the Regulation over Employment Test Score Usage from 703(H) to 703(L)

Matthew E. Blakely**

I. INTRODUCTION

More than 200 years ago, Declaration of Independence drafter Thomas Jeffe positioning and classification in the fledgling nation should be based on [1] Proponents of general aptitude examinations, such as -solving abilities under intense time constraints.[2] The College Board attempted to measure through such intelligence tests.[3]

A testing experience in which many college goers can most certainly empathize is the infamous exam conducted nationally by the Educational Testing Service (ETS), the SAT the most pervasively used college entrance exam.[4]

multiple-choice exams.[5] Firm in its belief t
society should be determined by their scores on a series of multiple-choice

[6] the ETS presumably believed that its new college entrance exams provided a remedy for the growing national concern of minorities being afforded an equal opportunity for access to higher education. After the passage of the Civil Rights Act of 1964, the ETS again marketed the SAT as a test that provided equality according to merit, with those marketing efforts resulting in the SAT successfully cornering the college exam market.[7]

Prior to the widespread use of the SAT in 1948, during the World War

discussion about Motorola, as Congress drafted 703(h) in light of its perceived fallacies.[17] The next sub-section introduces the Equal

general powers and purposes. Also, there, a delineation of the grievance procedures for Title VII is explained, which details the manner in which a

Thereafter, a sub-section is then apportioned to the most influential 703(h) case, Griggs, outlining its important doctrines and standards which create stringent hiring guidelines that companies using tests must follow. Next, a sub-section is attributed to analyzing the different types of EEOC employment testing regulations. There, the regulatory agency for Title VII expands on the Griggs doctrines by imposing disparate impact standards that monitor the minority hiring rate of companies.[18] Further, psychological testing studies are examined, as the EEOC and American Psychological Association

And, the final sub-

employers hired based on test score results with percentile adjustments tied to race.[19] This became the easiest employment method for some companies seeking to avoid Title VII disparate impact violations.

Part III of this paper details the modern day Title VII testing provision, as amended in the Civil Rights Act of 1991. Its first sub-section discusses the process of getting the CRA 1991 ratified through the executive and legislative branches. It also includes some congressional discussion about concerns on the state of testing that prompted its amending to the current version, 703(l), and exhibits the new language. And finally, the latter sub-section of Part III discusses the most recent litigation on the issue, observing that many current 92C70003-5sub

II. HISTORICAL DEVELOPMENTS UP TO 1990

A. The Civil Rights Act of 1964

Originally, the Civil Rights Act of 1964 was derived from Fair Employment Practice legislation, a post-World War II proposal created to address the area of equal employment opportunity.[22] President Kennedy sought civil rights legislation by sending a draft proposal to the 88th Congress in 1963.[23] Soon after, H.R. 7152 was introduced by Congressman [24]

Following the assassination of President Kennedy, civil rights legislation received high priority under the Lyndon Johnson administration.[25] The House Rules Committee cleared the bill for House action at the beginning of 1964.[26] The House adopted H.R. 7152 as amended by the Senate, with President Johnson ratifying on July 2, 1964.[27]

H.R. 7152 was a broad civil-rights measure with 10 titles.[28] The equal-employment opportunity provisions were contained in Title VII.[29] Other

designed, intended or used to discriminate because of race, color, religion, sex or national origin.[33]

B. Motorola

Actually, litigation for *Motorola, Inc. v. Illinois Fair Employment*

the hearing officer issued a decision ordering Motorola to hire Myart as an analyzer, and to cease using test No. 10 in its employment screening.[45]

passing result, Motorola nevertheless rendered his score as not passing, solely for discriminatory purposes.[46]

[47]

was

C. 703(h): Congressional Response to Motorola

Considerable congressional discussion was prompted by the initial tests on which whites performed better than blacks were in fact prohibited.[56]

intelligence exams would improve the overall quality of the work force, such tests did not bear any sort of demonstrable relationship to successful performance of the particular types of work in question here.[89]

So, the *Griggs* court outlined three factors that are particularly critical in consideration of a potential Title VII testing violation: (1) whether the

established prima facie case by proving that the facially neutral practice is must be tied directly with [97] If the defendant can then the burden once again shifts back to the plaintiff to show that the defendant could have instituted other practices to achieve this necessity which have less of an impact on the minority group.[98]

alternative practices.[99] lat[e] and identify[] the specific employment practices that are allegedly responsible for any [100] After the EEOC establishes this, then the employer carries the burden to demonstrate a valid, nondiscriminatory motive.[101] Liability attaches when the EEOC successfully proves that the

simply compares its hiring rates for the different racial groups, insuring that the protected group is being hired at
group, whites.[118]

2. Validation

The EEOC typically seeks to implement testing evaluative measures in harmony with those implemented by such influential organizations as the [119] In 1978, the APA produced testing validation studies examining the nexus between certain testing instruments with the test- [120]
During the APA studies, employers were advised to consider alternatives that would achieve their business purposes with lesser adverse impact on certain groups.[121] This resulted in employers delicately balancing a

ensure that any disparate impact on such a group would be traceable solely to the skill-level required for adequate job performance, and not racial bias.[122]

APA guidelines provide three circumstances in which employers should be permitted to utilized testing scores: (1) To eliminate grossly under-qualified candidates; (2) For the categorization of applicants based on perceived skill level; (3) For the ranking and filing of the most promising candidates to the least promising.[123]

importance of hiring the absolute high scorer to justify such narrow test score usage.[127]

Validation has become highly technical and complex, and is a constantly changing concept in industrial psychology. The APA declares that there are three concepts that can be used to validate a selection procedure.[128] These concepts reflect different approaches to investigating the job relatedness of selection procedures and may sometimes interrelate.[129] They are (1) criterion-related validity, (2) content validity, and (3) construct validity.[130] In criterion-related validity, a selection procedure is justified by a statistical relationship between the scores on a test and the measure of job performance.[131] In content validity, a selection procedure is justified by showing that it representatively samples significant parts of a job such as a foreign language test for an interpreter.[132] Construct validity involves identifying the psychological trait (the construct) that underlies successful performance on the job, and then devising a selection procedure to measure the presence and degree of the construct.[133] An example of

The APA testing guidelines contain technical standards and documentation requirements for the application of each of the three approaches.[134] One of the problems the guidelines try to address is the interrelated [135] The extreme cases are easy to grasp. A secretary, for example, may have to type. Many jobs require the separation of important matters which must be handled immediately from those which can be handled routinely. For the typing function, a typing test is appropriate. It is justifiable on the basis of content validity because it is a sample of an important or critical part of the job. The second function can be viewed as involving a capability to exercise selective judgment in light of the surrounding circumstances a mental process which is difficult to sample.

In addressing such situations, the guidelines attempt to make it practical to validate the typing test by a content strategy, but do not allow

127. *See id.* at 38291.

128. *See* SOC Y FOR INDUS. AND ORGANIZATIONAL PSYCHOLOGY

clarity on the matter.[161] So, by 1990, Congress sought to make amendments to the Civil Rights Act of 1964.[162]

III. THE CIVIL RIGHTS ACT OF 1991

A. 703(l)

ction with the interpretations of some provisions in the 1964 Civil Rights Act, including the 703(h) test score provision, House Democrats of the George H. W. Bush Administration filed a bill which would endeavor to amend some of the disputed provisions.[163] Among the new sections proposed as an amendment was section 703(l), a new employment test score provision that would supplant 703(h).[164]

Before the bill that would eventually become the 1991 Civil Rights Act passed, it endured a rather protracted legislative process.[165] It took several attempts, over a one-year time span.[166]

Civil Rights Act of 1990), did not pass in the House of Representatives

numbers [167] Then, at the first House meeting for the 102nd Congress session in 1991, the House Democrats filed H.R. 1 (The Civil Rights Act of 1991).[168] The Bush Administration felt that H.R. 1 was identical to that of the previously vetoed bill, so they introduced their own version of the Civil Rights Act of 1991 (S. 611) in March 1991.[169]

proposed bill with yet another bill, showing a lack of confidence in S. 611.[170] Senator D

Rights Act of 1991 (S. 1745), the third consecutive different bill proposed on the matter.[171]

Bush informed Danforth that S. 1745 would be acceptable, so long as
r to the previous bill.[172] Thus,

in S. 1745, and the bill became Public Law 102-166 in late 1991.[173]

In addition to the actual passage of the bill, the rationale of 703(l) may be delineated by noting some of the pertinent discussions in the House Report regarding P.L. 102-166 in 1991. The dialogue occurring during the committee hearings addressed several concerns, such as the general validity of employment tests, as well as the need for an omission of the very
[174] In the House Report, it was emphasized that the usage of tests by employers was permissible so long as they are valid, objective, and justified for business necessity.[175] An example provided was an employer using the test as a measure of aptitude:

description for the position.[176]

Public law 102-166, eventually signed and approved by the President, came

GATB test results stated that without such score adjusting, there would be *significant* discrimination against certain minority groups in score reporting to the participating employers.[185] Hence, employers realized that if evaluative tests were to be continually used during the hiring and promotional process and of course they were it was prudent for the implementation of an alternative method that could circumvent the wording of the newly implemented 703(l) to achieve the requisite diversity.

B. Contemporary Litigation

In the present day, a new popular method of classifying test scores has [186]

banding (i.e., grouping) of a batch of scores that fall within a certain range.[187] Employers who use it believe that any difference in scoring within the band is statistically insignificant, as some factors such as margin of error and scoring variance on multiple tries are inherent in any 2(m)17(et051004A1160B

scoring.[192] On the contrary, the plaintiffs in private sector claims frequently encounter courts applying a strict scrutiny analysis, since the test score is typically just a component of the conflict there.[193]

In addition, it appears that contemporary courts are *very* reluctant to rule against the employer.[194] This is presumably a combination of various clear-cut Title VII facial violations. Also, present employers as a whole exhibit a more sympathetic appearance towards minority inclusion.[195] As a result, unlike prior decades, there has been relatively minimal success for plaintiffs alleging Title VII 703(l) violations since 1991.[196] All of this, perhaps, could signal that the modern judiciary is fairly content with the manner in which employers are currently utilizing testing in its evaluative practices.

1. Public Sector

icer testing results on a

201

The city further substantiated this band by saying that there existed a margin of error inherent in scoring any exam, and that minor differences in test scores do not reliably predict differences in job performance.[202] The 9th Circuit Court stated that despite this procedure being contrary to a more strict ranking approach, the uniform guidelines do not forbid such an alternative selection procedure unless it is proven invalid.[203]

Also, the 9th Circuit Court used a strict scrutiny analysis to uphold the band scoring as well.[204] It was held that public sector employers can use race as a factor in selecting between qualified applicants pursuant to a

remedial action is necessary.[205] as found in the fact that there historically lacked adequate representation of minorities in the police force tantamount to the minority population in the surrounding community.[206] So, the banding was and is considered to be a way to circumvent the strict language of the 703(l).[207]

Also, similar to *Officers For Justice*, the 7th Circuit upheld another *Chicago Firefighters Local 2*.[208] There, several white police officers claimed to have been passed over for a promotion illegally after being surpassed by minority candidates who had scored lower on the promotional exam.[209] The city replied that despite the

located on the same band, thus any variance was statistically insignificant.[210] The court ruled in favor of the city, determining that this

number grade into a letter grade.[211]

201. *Id.* at 722-23.
202. *See id.* at 724.
203. *Id.* at 728.
204. *Id.* at 726.
205. *Id.*
206. *Id.*
207. *Id.* at 726-27.
208. *See* Chicago Firefighters Local 2 v. City of Chicago, 249 F.3d 649, 649-

That court analogized that placing all grades of 80-89 on a band called

yet due to the variance in each test-taker performance, this is the fairest way to judge their general range of abilities.[212] After all, the court opined, it is likely that *not* using such a general model of scoring may in itself be misleading, considering the many different factors which may have played into the scoring result (e.g., slightly varied questioning on different test versions, test-taker suffering from any sort of sickness or distraction, etc.).[213]

And one of the most recent cases deciding an alleged test score misuse by a public sector employer is the 2003 1st Circuit case of *Cotter v. City of Boston*.[214] There, suit was brought by ten white officers claiming to have been injured by the police department, who promoted three black officers scoring on exactly the same band over the plaintiffs.[215] The city defended the promotion of the blacks based on its need to remain in compliance with

at a minimum 80% rate of the most promoted group, the white group.[216] The plaintiffs, perhaps employing a legal stratagem in view of the perceived futility of a 703(l) testing claims (which plaintiffs had yet to succeed on in that jurisdiction), brought a 14th Amendment equal protection contention.[217]

The court reminded that for equal protection claims, it is not required that every citizen is treated identically, rather, there needs to be adequate explanation for treating groups differently.[218] Accordingly, the city felt it could show that its conduct of promoting only blacks from the scoring band was narrowly tailored to further a compelling governmental interest.[219] The

discrimination, avoiding lawsuits, and the operational needs of the department.[220] In regard
the court concurred that such action was a slow and gradual process, and it

could take years to successfully integrate an organization that had been segregated for a century in time span.[221]

Cotter also introduced a test which set out guidelines for what was

> [The extent to which] (i) The beneficiaries of the order are specially advantaged; (ii) the legitimate expectancies of others are frustrated or encumbered; (iii) the order interferes with other valid state or local policies; and (iv) the order contains (or fails to contain) built-in mechanisms which will, if time and events warrant, shrink its scope and limit its duration.[222]

In *Cotter*
because only the required number of black officers bringing the department

the candidates were on the same band, the plaintiffs could not claim that they were in fac
over them (thus no legitimate expectancies should have been frustrated).[223] No valid policies were disturbed, and the city had no quotas or long-term affirmative action guidelines established, hence an inference that such practices were limited in scope and duration.[224]

Thus, *Cotter*
which provides a checklist for employers to conscientiously monitor.[225] Intriguingly, present-day courts appear more rigid in their technique of analyzing cases, providing consistency with multi-prong tests in place for case analysis. As for the private sector lawsuits, though the litigation there is fairly infrequent, it is clear that there also appears to be more structure in recent cases with regards to how judges generally derive their decisions.[226] And, also comparable to public sector litigation, the inherent advantage seems to go to the employers, as plaintiffs tend to find little success against private companies as well.[227]

2. Private Sector

A casual observer to some of the post-1991 private sector discriminatory test claims would probably feel that plaintiffs tend to make much ado about nothing, as they appear to not have a very good

221. *Id.* at 169.
222. *Id.* at 171.
223. *Id.*
224. *Id.* at 172.,Ñ
225. *See id.* at 171.
226.

employer may validate a test through content, construct, or criterion-related studies.[236]

In *Hawkins v. Home Depot U.S.A.*, there was a Title VII claim from a black employee who was fired upon failing a departmental reassignment test after his prior position held was eliminated.[237] The plaintiff asserted that he was discriminated against because he was black, and also questioned whether he had actually failed the test, as he was not able to obtain a copy of it from the employer.[238] The employer moved for summary judgment,[239]

The court held that the plaintiff failed to show by the preponderance of the evidence that he had qualified for the newly created position, which required the passing of a sales-associate exam, due to the new responsibility of customer contact (a duty that the plaintiff previously did not have).[240]

that he in fact passed the test, or that his test results were in some way

instead just looking to fine-tune another variation of a disparate impact analysis tha

mystery, certainly, in an era of microscopic media attention and severe competition amongst companies, the typical employer has produced a more public-friendly countenance, in consideration that any sort of controversy with regards to racial discrimination could easily result in substantial losses