Adverse Impact: Why Is It So Difficult to Understand?

Art Gutman
Florida Institute of Technology

The Civil Rights Act of 1991 (or CRA-91) deems adverse impact unlawful if the defendant “fails to demonstrate that the challenged practice is job related and consistent with business necessity.” In *Lanning v. SEPTA (1999)* (or *Lanning I*), the chosen cutoff on an aerobic capacity test yielded passing rates of 55.6% and 6.7%, respectively, for men and women. The district court upheld the test, but the 3rd Circuit remanded with instructions to determine if the test measures the “minimal qualifications necessary for successful job performance.” On remand, the district court ruled that the test passed this freshly minted 3rd Circuit standard, and on October 15, 2002, the 3rd Circuit upheld the district court’s ruling in *Lanning II*.

*Lanning I* is well documented by Sharf in the October 1999 issue of *TIP*. Sharf also provides an overview of *Lanning II* in the current issue of *TIP* (including comments by five labor lawyers). Therefore, for the most part, I will leave the facts in this case undisturbed, other than to express agreement with three general sentiments expressed by Sharf and others. First, regardless of the quality (or lack thereof) of SEPTA’s validity study (which is open to debate), *Lanning I* did not require new adverse impact rules. Second, the DOJ similarly overinterpreted by implying different burdens for “job relatedness” and “business necessity” (that job-relatedness be supplemented with evidence that a cutoff score distinguishes successful vs. unsuccessful performers). Third, I join the chorus who find it baffling that the 3rd Circuit would declare our SIOP Principles as being irrelevant; there is simply too much case law to the contrary.


---

1 Although Griggs and Albermarle also featured high school diplomas, the Supreme Court’s discourse in both cases focused primarily on the testing issues.
defense for adverse impact defense was light in Beazer, heavy in Dothard, and somewhere in between (i.e., moderate) for standardized tests. I further believe that the Wards Cove ruling established (temporarily) a standard that was too light to apply to all adverse impact cases and that the Lanning standard is too heavy to apply to all adverse impact cases.

**Adverse Impact Light (Beazer)**

In Beazer, the Supreme Court upheld exclusion of methadone users because it was deemed “obvious” that drug addiction threatens the “legitimate employment goals of safety and efficiency.” Similar rulings were rendered in several lower court cases. For example, in Hyland v. Fukada (1978), there were obvious reasons for excluding from security work a felon previously convicted of armed robbery and in Davis v. Dallas (1985), there were obvious reasons for excluding recent drug users from police work (i.e., it shows disregard for the law). In these cases, the defense to adverse impact was nothing more than a simple articulation (or explanation) as, for example, in disparate treatment cases such as McDonnell Douglas v. Green (1973) and Texas v. Burdine (1981).

The above cases featured biographical (or historical) variables. Sometimes, the standard of proof for such variables was beyond a simple articulation. However, it rarely (if ever) rose to the level of a full-blown validity study. For example, in Spurlock v. United Airlines (1972), the airline successfully defended a 4-year degree requirement based on expert opinion that a college education is needed to “cope” with classroom training requirements. Here, the 10th Circuit also established a caveat for public safety routinely cited by other courts over the years. Accordingly:

When a job requires a small amount of skill and training and the consequences of hiring an unqualified applicant are insignificant...the employer should have a heavy burden to demonstrate...that his employment criteria are job-related. On the other hand, when the job clearly requires a high degree of skill and the economic and human risks are great, the employer bears a correspondingly lighter burden to show that his employment criteria are job related.

Additionally, in U.S. v. Buffalo (1978), a district court upheld a high school diploma requirement for police officers based on federal commission reports in 1967 and 1968 suggesting that “a high school education is a bare minimum requirement for successful performance of the policeman’s responsibilities.” And in Davis v. Dallas, a “poor driving” policy was upheld based on research indicating that past habits predict future habits and an education requirement (45 hours of college credit with C or better grades) was upheld based on the task force reports cited in U.S. v. Buffalo. In Davis, the 5th Circuit also echoed the Spurlock ruling, stating:  

2 Notice how the term “manifest” relationship is used in the context of a defense that is lighter than Griggs-Albermarle.
The instant context is distinguishable from those presented in the foregoing decisions, where we struck down education requirements for jobs which involved neither professional-type positions nor an especially unusual degree of risk or public responsibility. We regard this distinction as crucial, and affirm the district court’s finding that appellees’ educational requirement bears a manifest relationship to the position of police officer.

In short, especially when public safety has been implicated, courts have used lighter standards for biographical variables than seen in the Uniform Guidelines.³

**Adverse Impact Moderate (Griggs-Albermarle)**

Prior to Title VII, Duke Power hired blacks for low-wage labor jobs but not for higher-wage operations jobs. On the same day Title VII became law (July 2, 1965), applicants for operations jobs had to possess a high school diploma and pass two cognitive tests. Exclusion rates were 94% for blacks and 42% for whites for the tests and 88% for blacks and 66% for whites for the diploma. Although both requirements were struck down, Griggs was primarily a cognitive testing case. Duke Power relied on Title VII language making it legal to use professionally developed ability tests. According to the 1966 EEOC Guidelines, such a test must:

[F]airly measure the knowledge or skills required by the particular job or class of jobs which the applicant seeks, or which fairly affords the employer a chance to measure the applicant’s ability to perform a particular job or class of jobs.

Speaking for a unanimous Supreme Court, Justice Burger ruled that Title VII covers the “consequences of employment practices, not simply the motivation” of employers. Therefore, given adverse impact, defendants must prove a manifest relationship between the challenged practice and the “employment in question.” Other terms used by Burger for this “manifest relationship” included “job relatedness” and “business necessity.”

Facing trial on virtually the same facts as Duke Power, the Albermarle Paper Company hastily conducted a criterion-related validity study 1 month prior to trial. According to the Supreme Court, this study had four major defects, including (a) a lack of quality, or “odd patchwork;” (b) unknown job-performance criteria and “subjective supervisory rankings;” (c) a focus on higher-level jobs rather than the “entering low-level jobs” at issue; and (d) a validation sample that included only “job-experienced white workers.” Relying on the 1974 EEOC Guidelines, the Supreme Court then defined how a manifest relationship should be proven:

The message of these Guidelines is the same as that of the Griggs case—that discriminatory tests are impermissible unless shown, by professionally acceptable methods [italics added] to be “predictive of or significantly correlated [italics added] with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are evaluated.” [29 CFR section 1607.4(c)].

Clearly, terms like “manifest relationship,” “job-relatedness,” and “business necessity” became interchangeable codes for validity studies. Griggs and Albermarle also served as the basis for the 1978 Uniform Guidelines, which established ground rules for validity studies. The Guidelines also left room for “new strategies for showing the validity of selection procedures.... as they become accepted by the psychological profession.” And in subsequent case law, courts have paid deference to newer strategies as, for example, in the 1987 SIOP Principles.

Davis and Teal addressed other related issues such as use of training data for criterion-related validity studies of police selection (Davis) and validation of individual steps of multiple hurdles even when the selection process as a whole does not produce adverse impact (Teal). Taken as a whole, it was clear from these cases that the Griggs-Albermarle tradition required proof of validity in accordance with the “professionally acceptable methods” adopted by our profession.

**Adverse Impact Heavy (Dothard)**

A dictionary definition of the term “business necessity” implies selection criteria necessary for business survival. Obviously, this exceeds the implications of a validity study. That is, even if content and criterion-related validity are established, there is no automatic implication that failure to use the targeted selection criteria will destroy the business. Dothard illustrates that such a heightened standard does exist for physical characteristics such as height and weight and that this standard matches the bona fide occupational qualification (BFOQ) defense for facial exclusion based on gender or age.4

More specifically, the BFOQ defense requires proof that it is reasonably necessary to exclude all or most members of a class. For example, borrowing from the public safety defense used in Spurlock, and extended in Hodgson v. Greyhound (1974), Usery v. Tamiami Trail Tours (1976), and in Dothard, the EEOC, in a 1981 modification of its age discrimination Guidelines, stated the following:

An employer asserting a BFOQ defense has the burden of proving that (1) the age limit is reasonably necessary to the essence of the business, [italics added] and either (2) that all or substantially all individuals excluded

---

4 Although BFOQ applies to religion and national origin in theory, all BFOQ cases to this point have featured gender and age. By definition, there is no defense, BFOQ or otherwise, for facial discrimination based on race or color.
from the job involved are in fact disqualified, or (3) that some of the individual so excluded possess a disqualifying trait that cannot be ascertained except by relevance to age. If the employer’s objective is the goal of public safety, the employer must prove that the challenged practice does indeed effectuate that goal and that there is no acceptable alternative which would better advance it or equally advance it with less discriminatory impact.

With minor exceptions, these Guidelines mimic those promulgated by the EEOC for BFOQ defenses for facial discrimination based on gender.

Dothard was actually two cases in one. First, minimum height and weight standards excluded significantly more women then men as prison guards. The State of Alabama argued that height and weight are indicators of strength, prompting the following ruling from Justice Rehnquist: “If the job-related quality that the appellants identify is bona fide, their purpose could be achieved by adopting and validating a test for applicants that measures strength directly.”

Second, after abandoning the height/weight criteria, the state facially excluded all women, the logic being that they posed a threat to prison safety in all-male maximum security prison where 20% of the inmates are sex offenders. The state won on this (BFOQ) defense. Thus, it was easier to justify facial exclusion of women than exclusion based on physical characteristics.

More generally, defendants have rarely succeeded in defending purely physical criteria. For example, in Boyd v. Ozark Airlines (1977), it was proven that shorter pilots could not safely operate all cockpit instruments and in Fitzpatrick v. Atlanta (1993), firefighters needed to be beardless for proper functioning of facial safety equipment. In contrast, in Horace v. Pontiac (1980), the defendant argued that being tall is necessary for police officers to fend off criminals and to gain respect, to which the 6th Circuit, as in Dothard, countered with more direct methods of assessing these capabilities. In short, in both Boyd and Fitzpatrick, height was a target attribute, whereas in Dothard and Pontiac, it was a surrogate (or proxy) for other attributes. Critically, in all of these cases, the standard was business necessity in the dictionary meaning of the term, not in the way it was used in the Griggs-Albermarle tradition.

Wards Cove Featured Adverse Impact “Light”

In Watson v. Fort Worth Bank (1988), a black woman was passed over for promotion for a fourth time, each time in favor of a white applicant and each time because of subjective ratings by white supervisors. Only eight justices heard this case and each agreed that subjective selection decisions are subject to adverse impact analysis. However, a plurality of four justices (O’Connor, Rehnquist, Scalia, & White) feared that employers would resort to quota
selection and endorsed fundamental changes in the *Griggs-Albermarle* tradition. Or as stated by Justice O’Connor:

> [W]hen a plaintiff has made out a prima facie case of disparate impact, and when the defendant has met its burden of *producing* [italics added] evidence that its employment practices are based on legitimate business reasons, the plaintiff must “show that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer’s legitimate interest in efficient and trustworthy workmanship.”

To “produce” means to articulate or explain, as in *McDonnell-Burdine* disparate treatment scenarios. As noted earlier, it is also the standard used in several of the adverse impact “light” scenarios, most notably in *Beazer*. Recognizing that adverse impact is a heterogeneous phenomenon, Justice Blackmun countered that:

> [W]ith the type and size of the business in question, as well as the particular jobs for which the selection process is employed. Courts have recognized....nationwide studies and reports...expert testimony...and psychologist’s testimony explaining job-relatedness....[etc.]

Despite Justice Blackmun’s insightful observation, a year later in *Wards Cove v. Atonio (1989)*, the newly seated Justice Kennedy provided the fifth vote to turn the O’Connor plurality opinion into case law. Arguably, *Wards Cove* was not even suitable for adverse impact analysis.5

The *Wards Cove* ruling permitted employers to win cases that were clearly losers under *Griggs-Albermarle* rules. For example, in *Evans v. Evanston (1989)*, the 7th Circuit, in a pre-*Wards Cove* ruling, favored women adversely impacted by arbitrary cut-off scores on an agility test. However, after *Wards Cove*, the court issued the following apologetic ruling:

> The plaintiff had challenged the entire test, and the city put in a great deal of justificatory evidence which succeeded in justifying everything about the test except the scoring method. That was enough to satisfy the burden of production and shift inquiry to whether...the plaintiff proved—since after *Wards Cove* it is the plaintiff that has the burden of persuasion—that the test, because of its method of scoring, did not serve the legitimate ends of the employer but instead unreasonably excluded women.

And in *Allen v. Seidman (1989)*, the 7th Circuit reversed a pre-*Wards Cove* ruling favoring black applicants for promotion, issuing the following apologetic ruling:

> In a test notably devoid of objective standards, where far from using blind grading the testers based an unknown part of the grade on the results of an unstructured personal interview, the danger is acute that racial bias of

---

5 Gutman (2000) argues that *Watson* was a viable adverse impact testing case, but that *Wards Cove* better fit the historical image of a pattern or practice motive case.
which the testers may well be unconscious may influence the grade....It is hard to believe that the FDIC can’t do better than the Program Evaluation test, which its own consultants had criticized repeatedly.

_Wards Cove_ was easily the most important of six 1988 Supreme Court rulings Congress opposed in the aborted Civil Rights Restoration Act of 1990 (or CRRA-90). Indeed, the failure to compromise on this ruling prompted President Bush to veto CRRA-90, which Congress nearly overrode, missing by a single vote in the Senate. Interestingly, the Democrats opened the CRRA-90 debates pleading for the adverse impact defense from _Dothard_ (that challenged practices are essential for job performance), whereas the Republicans pleaded for the defense from _Beazer_.

**Lanning Features Adverse Impact “Heavy”**

Because of the near override of Bush’s veto, the Democrats and Republicans worked harder in 1991 and were successful in compromising on the _Wards Cove_ ruling (and other issues). However, in CRA-91, Congress expressly limited the legislative history for _Wards Cove_ to an “Interpretive Memorandum” from the Congressional Record (volume 127, page 15276). This Memorandum contains two substantive paragraphs. The first substantive paragraph addresses identification of the cause(s) of adverse impact. Accordingly:

When a decision-making process includes particular, functionally integrated practices which are components of the same criterion, standard, method of administration, or test, such as the height and weight requirements designed to measure strength in _Dothard v. Rawlinson_, 433 US 321 (1977), the particular, functionally integrated practices may be analyzed as one practice.

The second substantive paragraph addresses (for present purposes) a more important issue; the defense to adverse impact. Accordingly:

The terms “business necessity” and “job-related” are intended to reflect the concepts enunciated by the Supreme Court in _Griggs v Duke Power Co._, 401 US 424 (1971), and in other Supreme Court decisions prior to _Wards Cove Packing Co. v. Atonio_, 490 US 642 (1989).

Additionally, statutory language from Sec.105(A)(i)) in CRA 91 states that adverse impact is _unlawful_ if the defendant “fails to demonstrate that the challenged practice is job related and consistent with business necessity.”

Taken as a whole, I believe CRA-91 left us with two major sources of ambiguity.

First, it is unclear what “concepts enunciated by the Supreme Court in _Griggs_...and in other...decisions prior to _Wards Cove_” means. To illustrate, shortly before the Senate voted 93 to 5 to pass CRA-91, Senator Dole explained his impending yes vote by suggesting that “the present bill has codified the ‘business necessity’ test in _Beazer_ and reiterated in _Wards Cove_”
[Congressional Record, 1991, S15476]. In other words, his interpretation of “prior to Wards Cove” equated to adverse impact “light.” On the other hand, as he was signing the bill into law, President Bush’s view was that CRA-91 contains “a compromise provision that overturns Wards Cove by shifting to the employer the burden of persuasion on the ‘business necessity’ defense,” a quote that implies a defense more akin to Griggs-Albermarle.

The second source of ambiguity concerns the meaning of the phrase “job related and consistent with business necessity.” Although both “job-relatedness” and “business necessity” are terms featured in Griggs (along with “manifest relationship”), they are used interchangeably, not only in Griggs, but also, in most other cases (including Beazer). The Lanning rulings (and the DOJ Amicus Brief) implies a higher standard than Griggs-Albermarle. In the Griggs-Albermarle era, proof of validity was tantamount to proof of “job-relatedness.” Thus, requiring more than proof of validity (i.e., minimum qualifications) raises the standard in the direction of adverse impact “heavy,” as used in Dothard.

Conclusions

First, we owe a vote of thanks to Brother Sharf for kindling the interest in Lanning in his 1999 TIP article and for rekindling the interest in the current issue. The reader should pay attention to the comments by the five labor lawyers. Among these, my favorite quote is by Attorney David Copus, who states:

[I]t seems to me that somehow the court has lost sight of the fact that any test, physical or otherwise, is merely a rough predictor designed to increase the likelihood of success on the job. We just want the test to be better than a toss of a coin.

Unless a correlation coefficient is at or near perfection (fat chance), what Attorney Copus says is true by definition. Short of such perfection, a significant correlation implies nothing more than better-than-chance prediction. Yet, the 3rd Circuit’s obsession with a single applicant who failed the test and went on to great heights is, as Attorney Copus states, “goofy.”

Second, it has always been goofy to me that the Supreme Court, Democrats, and Republicans alike lost sight of the fact that the Griggs-Albermarle standard was applicable primarily, if not exclusively, to standardized tests.

Third, the at-issue jobs have strong implications for public safety; these are transit authority cops, not coal producers or mill workers. Courts have always given extra leeway to employers in such cases, not only for adverse impact defenses, but also for BFOQ defense. To alter this precedent is also goofy.

Fourth, public safety aside, what about expense? Before police officers attend training school, they will likely receive drug tests, background checks, and polygraph tests and, after conditional job offers, psychological tests, clin-
ical interviews, and medical exams. That’s a lot of money to spend, particularly if an applicant is less likely to be selected. Suppose, for example, there are 100 “minimally qualified” applicants for 20 training slots, and past data reveal that a pool of 40 is sufficient to fill those slots. It makes sense to focus on the 40 best performers. Ironically, to force any organization to undergo such needless expense is a financial threat to its survival. How goofy is that?

Finally, I find it goofy that a court would take it upon itself to decide the merits or demerits of our SIOP Principles. It might well be that the validity study conducted by SEPTA was not a good one; we can debate that among ourselves. Moreover, had the 3rd Circuit, based on the opinions of the various experts, concluded that the SETPA validity study was no better than in Albermarle and left it at that, the noise level would have been much lower. Our concerns stem from the fact that the 3rd Circuit created an unnecessarily heavy standard that never before existed for standardized tests, and they lectured us on the meaning of our own constructs.

References


Case Law Citations

Boyd v. Ozark Airlines (CA8 1977) 568 F.2d 50.
Davis v. Dallas (CA5 1985) 777 F.2d 205 (No. 97-843).
Evans v. City of Evanston (CA7 1989) 881 F.2d 382.
Fitzpatrick v. City of Atlanta (CA11 1993) 2 F.3d 1112.
Lanning v. SEPTA (CA3 1999) 181 F.3d 478.
Spurlock v. United Airlines, Inc. (CA10 1972) 475 F.2d 216.
Usery v. Tamiami Trail Tours Inc. (CA5 1976) 531 F.2d 224.
Washington (Mayor, DC) v. Davis (1976) 426 US 229.