2007

Equal Access to Post-Secondary Education: The Sisyphean Impact of Flagging Test Scores of Persons with Disabilities

Helia Garrido Hull
Barry University

ARTICLES

EQUAL ACCESS TO POST-SECONDARY EDUCATION: THE SISYPHEAN IMPACT OF FLAGGING TEST SCORES OF PERSONS WITH DISABILITIES

HELIA GARRIDO HULL*

I. INTRODUCTION 16
II. DISABILITY-BASED DISCRIMINATION THROUGHOUT HISTORY 18
A. Background 18
B. Early Disability Rights Movement and its Impact on Education 22
III. LEGALITY OF FLAGGING TEST SCORES 26
A. Federal Law Applicable To Flagging 26
B. Professional Rules Applicable to Flagging 30
IV. EVIDENCE FROM STANDARD AND NONSTANDARD TEST ADMINISTRATIONS 33
A. Standardization and Accommodation 33
B. Empirical Evidence of Score Comparability 35
C. Legal Challenge and the Partial Demise of Flagging 40
V. FLAGGING THE LSAT AND MCAT 43
A. LSAT and Score Comparability 43
B. MCAT and Score Comparability 47
C. Raising the Flag: Judicial Response to Flagging Professional Exams 49
VI. RECOMMENDATIONS 51
A. Amend Titles II and III of the Americans with Disabilities Act 52
B. Require Testing Entities to Adhere to Their Own Agreements 52
C. Reevaluate OCR's Interim Policy 54
D. Eliminate the Element of Speed from Standardized Tests 54
E. Amend the Standards to Comport with Existing Disability Laws 55
VII. CONCLUSION 57

> Then I witnessed the torture of Sisyphus, as he wrestled with a huge rock with both hands. Bracing himself and thrusting with hands and feet he pushed the boulder uphill to the top. But every time, as he was about to send it toppling over the crest, its sheer weight turned it back, and once again towards the plain the pitiless rock rolled down. So once more he had to wrestle with the thing and push it up, while the sweat poured from his limbs and the dust rose high above his head.[1]

I. INTRODUCTION

Like Sisyphus, condemned for eternity to roll a boulder to the top of a hill only to have it roll back down, disability rights advocates labor under a perpetual undulation of advancement and decay in the rights afforded to disabled individuals. Aided by an emerging social policy of inclusion in the early 1970s, advocates rolled a proverbial rock of equality up from the cavernous depths created by past prejudice in an effort to place disabled individuals on level ground with others in society. A groundswell of conflicting ideologies regarding the impact new civil rights legislation had on the rights of non-disabled individuals, however, quickly caused the rock to start rolling back down the hill. More than three decades later, individuals with disabilities continue to experience educational, political, economical, social, and cultural discrimination.[2] Perhaps nowhere is this discrimination more evident than in the practice of flagging standardized tests.

Standardized college entrance exams are designed to provide a level playing field for all examinees.[3] Ideally, the exam content, administration and scoring are applied uniformly to all examinees so that differences in scores received reflect true individual differences in aptitude among students.[4] However, standardized testing is problematic for many students whose disability prevents them from taking the test as typically administered. Disabled individuals often require some form of

---

*Assistant Professor of Law, Barry University Dwayne O. Andreas School of Law; B.A. Providence College, J.D. Stetson University College of Law. First, the author would like to thank Julie Anthousis for her editorial contribution to this paper and for her dedication, which truly rose above that of any research assistant. Second, the author would like to thank Dean J. Richard Hurt and the Barry Law School Faculty for their support and assistance. Last, but certainly not least, the author would like to thank her husband Eric V. Hull, her children, Kayleigh and Tyler, and her parents, Al and Elia Garrido, for their unwavering love and support.

[1]HOMER, THE ODYSSEY 176-77 (E.V. Rieu & D.C.H. Rieu trans., Penguin Books 1991) (c. 700 B.C.) (footnote omitted). Sisyphus, the mythical king of Corinth, was condemned in Hades and sentenced by Zeus to roll a heavy boulder up a steep hill for all eternity. Each time Sisyphus reached the top, the weight of the rock caused it to roll back down to the bottom of the hill.

[2]*See generally* RUTH COLKER & ADAM A. MILANI, THE LAW OF DISABILITY DISCRIMINATION 1-36 (LexisNexis 5th ed. 2005).

[3]CollegeBoard.com, Standardized Testing: The Big Picture, http://www.collegeboard.com/parents/tests/testing-overview/21292.html (last visited Mar. 1, 2007).

[4]April Zenisky et al., *A Basic Primer for Understanding Standardized Tests & Using Test Scores,* ADVENTURES IN ASSESSMENT, Spring 2004, at 29, *available at* http://www.sabes.org/resources/publications/adventures/vol16/vol16.pdf.

accommodation to complete the examination. In an effort to eliminate testing barriers that might otherwise prevent disabled examinees from demonstrating their actual knowledge and skill on standardized tests, testing services utilize a wide range of testing accommodations for people with disabilities.[5]

To receive an accommodation, disabled individuals are required to disclose information regarding their disability.[6] If a modification is granted, the testing service then decides if the accommodation has the effect of rendering the test results less reliable as predictors of a student's future performance than non-flagged scores.[7] If so, the test scores received are annotated or "flagged" to indicate that the test was taken under nonstandard conditions.[8] Educational institutions requesting the score report are sent the score along with information warning the recipient that the test score should be interpreted with caution.[9] Ostensibly, the purpose of flagging is to maintain psychometric integrity of the test.[10] In reality, the practice discriminates by segregating students with disabilities from the rest of the applicant pool and by informing college admissions personnel that the individual who took the examination is disabled.[11] In view of the social stigma associated with disabilities, and the inherent costs of providing accommodations to disabled students, the opportunity for bias within the admissions selection process is clear. As a result, the practice of flagging standardized tests has come under increasing scrutiny. The practice of distinguishing test takers having a disability from those who do not runs counter to the social policy of inclusion, and prevents disabled individuals from enjoying the benefits of equal citizenship.

---

[5]PANEL ON TESTING OF HANDICAPPED PEOPLE, NATIONAL RESEARCH COUNCIL, ABILITY TESTING OF HANDICAPPED PEOPLE: DILEMMA FOR GOVERNMENT, SCIENCE, AND THE PUBLIC 96-104 (Susan W. Sherman & Nancy M. Robinson eds., 1982).

[6]*See, e.g.*, LAW SCH. ADMISSION COUNCIL, LSAT ACCOMMODATIONS FORM (2006), *available at* http://www.lsac.org/pdfs/2006-2007/AccommodationsForm-2006.pdf; CollegeBoard.com, Accommodations, http://www.collegeboard.com/ssd/student/accom.html (last visited Mar. 1, 2007).

[7]*See* Jordan J. Cohen, A Word From the President, *Meeting a Dual Obligation*, ASS'N AM. MED. C. REP., October 2003, *available at* http://www.aamc.org/newsroom/reporter/oct03/word.htm.

[8]*Id.*

[9]*See* Law Sch. Admission Council, Accommodated Testing, http://www.lsac.org/LSAC.asp?url=/lsac/accommodated-testing.asp (last visited Mar. 1, 2007) ("If you receive additional test time as an accommodation for your disability, LSAC will send a statement with your LSDAS Law School Reports advising that your score(s) should be interpreted with great sensitivity and flexibility.").

[10]Psychometrics deals with "the design and analysis of research and the measurement of human characteristics," including testing of aptitude and intelligence. Robert Williams, Psychometrics and Allied Matters, http://www.geocities.com/bororissa/psycho.html (last visited Mar. 1, 2007); Cohen, *supra* note 7.

[11]Because only students with documented disabilities are eligible to receive an accommodation, flagged test scores necessarily inform the recipients of test scores that the examinee has some form of disability. Cohen, *supra* note 7.

Part II of this paper provides a brief overview of the prejudice disabled individuals have endured throughout history, and discusses some early movements toward change. Part III discusses the legality of flagging test scores and provides an overview of federal laws and professional standards applicable to the practice. Part IV discusses the practice of flagging and the use of accommodations in standardized testing, and evaluates the empirical evidence obtained from standard and nonstandard test administrations in the context of flagging. The section concludes with a brief discussion of why some testing entities stopped flagging test scores. Part V discusses the continued practice of flagging test scores received on the Law School Admission Test (LSAT) and the Medical College Admission Test (MCAT) and examines the empirical evidence used to justify the practice. The section concludes with an analysis of the leading case addressing flagging scores received on professional exams. Part VI provides commentary on the propriety of flagging tests and provides recommendations for change to eliminate the stigmatizing effects of segregating students with disabilities in the admissions process.

II. DISABILITY-BASED DISCRIMINATION THROUGHOUT HISTORY

A. Background

The history of society reflects a history of discrimination against, and misunderstanding of, individuals with disabilities. In ancient Greece, all newborn children determined by state officials to be sickly or deformed were abandoned to die.[12] The Law of the Twelve Tables, legislation that governed ancient Rome for nearly 1000 years, mandated "A father shall immediately put to death a son recently born, who is a monster, or has a form different from that of members of the human race."[13] Despite enjoying an elevated status in society, a priest was expressly prohibited by scripture from bringing sacrificial offerings to his congregation during service if he was afflicted with some form of disability.[14] Some religious scholars have suggested that the prohibition against a disabled priest offering the body and

---

[12]Robert A. Guisepi, The Glory That Was Greece (2001), http://history-world.org/ancient_greece.htm.

[13]The Laws of the Twelve Tables table IV, law 3 (c. 450 B.C.), *reprinted in* 1 THE CIVIL LAW 57, 65 (S.P. Scott ed., Central Trust Co. 1932).

[14]*See Leviticus* 21:16-23 (King James):

> And the LORD spoke unto Moses, saying, "Speak unto Aaron, saying, 'Whosoever he be of thy seed in their generations who hath any blemish, let him not approach to offer the bread of his God. For whatsoever man he be that hath a blemish, he shall not approach: a blind man, or a lame, or he that hath a flat nose, or any thing superfluous, or a man who is broken-footed, or broken-handed, or crookbackt, or a dwarf, or who hath a blemish in his eye, or hath scurvy, or scabbed, or hath his stones broken—no man that hath a blemish of the seed of Aaron the priest shall come nigh to offer the offerings of the LORD made by fire. He hath a blemish: he shall not come nigh to offer the bread of his God. He shall eat the bread of his God, both of the most holy and of the holy. Only he shall not go in unto the vail, nor come nigh unto the altar, because he hath a blemish, that he profane not My sanctuaries; for I the LORD do sanctify them.'"

Id.

blood of Christ was designed to prevent followers from questioning God's perfection.[15]

Martin Luther's belief that the devil played a role in disability and disease may have exacerbated the prejudice against children who were different.[16] In reference to a learning disabled boy whom he felt was possessed by the devil, Luther declared "If I were the Prince, I should take this child to the Moldau River . . . and drown him."[17] The prejudice against people that were different became lethal during the great witch hunts of the Middle Ages, a period that witnessed the state-sanctioned murder of millions of individuals identified as witches.[18] The *Malleus Maleficaru,* a manual used to identify, prosecute, and dispatch witches, provided a basis for gruesome tortures of individuals whose disabled offspring provided evidence of their association with the devil.[19] Although impossible to quantify, there can be no doubt that many "witches" killed during the hunts were actually individuals with disabilities who exhibited misunderstood behaviors considered by the masses to be socially deviant.

The unwillingness of society to accept flaws in the human form is evident in the near flawless portraits of world leaders through the ages. Perhaps nowhere is this more evident than in the portraits of King Henry VIII's wife, Anne of Cleves, and his daughter, Elizabeth. Both women survived small pox and suffered scarring, yet each is portrayed in period artwork with a perfect complexion.[20]

During the industrial revolution of the eighteenth and nineteenth centuries, fast moving machinery, assembly lines and the need for uniformity created problems for people with disabilities. Those individuals unable to complete tasks in accordance with factory-based standards were considered deviant and excluded from the labor force.[21] Many disabled individuals unable to work were placed into state-built institutions, asylums, hospitals, workhouses and prisons under the guise of providing

---

[15]Michael Gold, *Parashat Emor*, TORAH SPARKS, May 13, 2006, http://www.uscj.org/EMOR_57666958.html.

[16]M. Miles, *Martin Luther and Childhood Disability in 16th Century Germany: What Did he Write? What Did he Say?*, J. RELIGION DISABILITY & HEALTH, No. 4 2001, at 5, *available at* http:www.independentliving.org/docs7/miles2005b.html.

[17]MARTIN LUTHER, COLLOQUIA MENSALIA 387 (Henry Bell trans., London, William Du-Gard 1652) (1566), *quoted in* LEO KANNER A HISTORY OF THE CARE AND STUDY OF THE MENTALLY RETARDED 7 (1964).

[18]Wicasta Lovelace, *Introduction to the Online Edition*, *in* HEINRICH INSTITORIS, MALLEUS MALEFICARUM (Montague Summers trans., online republication of the 1928 ed. n.d.) (1486), http://www.malleusmaleficarum.org/ (last visited Mar. 1, 2007).

[19]*Id.*

[20]*See e.g.*, ELAINE HATFIELD & SUSAN SPRECHER, MIRROR, MIRROR . . . THE IMPORTANCE OF LOOKS IN EVERYDAY LIFE 141-42 (1986), *available at* http://www2.hu-berlin.de/sexology/BIB/HATF2.htm (follow "Chapter 5 MORE INTIMATE AFFAIRS" hyperlink) (suggesting that when artist Hans Holbein was commissioned by Henry VIII to paint a "perfect likeness" of Anne of Cleves, Holbein omitted evidence of Anne's smallpox scars to make the painting more flattering).

[21]Ravi Malhotra, *The Politics of the Disability Rights Movements*, NEW POLITICS, Summer 2001, *available at* http://www.wpunj.edu/newpol/issue31/malhot31.htm.

rehabilitation and protection.[22] Often, however, they endured intense abuse while living under horrible conditions.[23]

At the turn of the twentieth century, Sir Francis Galton's "Eugenics" movement gained popularity as a means to improve the health of society through natural selection.[24] Eugenics encouraged procreation between individuals with desirable characteristics, and discouraged procreation by individuals having inferior or undesirable characteristics through forced sexual sterilization, marriage prohibition, segregation and institutionalization.[25] Disabled individuals soon became viewed as a danger to society, prompting their widespread segregation and placement into asylums, often under dangerous and harsh conditions.[26] England's Mental Deficiency Act of 1913, for example, certified individuals admitted to institutions and created isolated "colonies" of "mental defectives" to ensure that those individuals would never rejoin society.[27] At the time the Act was passed, Winston Churchill, a proponent of the eugenics movement, announced:

> The unnatural and increasingly rapid growth of the feeble-minded . . . classes, coupled . . . with steady restriction among all the thrifty, energetic and superior stocks constitutes a . . . race danger . . . . I feel that the source from which the stream of madness is fed should be cut off and sealed off before another year has passed.[28]

Ironically, Churchill suffered from a learning disability.[29]

The American Eugenics Society was founded in 1926.[30] The movement gained considerable support from the United States Supreme Court's infamous decision in *Buck v. Bell*,[31] which held that a Virginia statute authorizing the forced sterilization of the inmate child of a mother diagnosed with a mental disorder was constitutional.

---

[22]*Id.*

[23]*Id.*

[24]John Holland, Eugenics: America's Darkest Days: Sir Francis Galton, http://iml.jou.ufl.edu/projects/Spring02/Holland/Galton.htm (last visited Mar. 1, 2007).

[25]Ted L. DeCorte, Jr., Menace of Undesirables: The Eugenics Movement During the Progressive Era, http://www.geocities.com/MadisonAvenue/Boardroom/4278/eugenics.htm (last visited Mar. 1, 2007) (Eclectic Buzz / Eclectic Mouse Experience).

[26]John Barrett, History of Discrimination Against Disabled Persons - Part Four, http://www.jackiebarrett.ca/DisabledDiscrimination4.htm (last visited Mar. 1, 2007).

[27]MENCAP, Changing Attitudes to People With a Learning Disability, http://www.mencap.org.uk/html/about_mencap/changing_attitudes.asp (last visited Mar. 1, 2007).

[28]STEVE JONES, THE LANGUAGE OF GENES 19 (1st Anchor Books trade paperback ed. 1995) (1993) (internal quotation marks omitted) (quoting Winston Churchill on compulsory sterilization of the feeble-minded and insane).

[29]*Famous Historical Figures with Disabilities*, SNAP REP., Autumn 1998, *available at* http://www.snapinfo.org/News/Docs/fall98_famous.html.

[30]Am. Philosophical Soc'y, Promoting Eugenics in America, http://www.amphilsoc.org/library/exhibits/treasures/aes.htm (last visited Feb. 5, 2007).

[31]Buck v. Bell, 274 U.S. 200 (1927).

Writing for a near unanimous majority, Justice Holmes opined: "It is better for all the world, if instead of waiting to execute degenerate offspring for crime, or to let them starve for their imbecility, society can prevent those who are manifestly unfit from continuing their kind. . . . Three generations of imbeciles are enough."[32]

By the early 1930s, thirty states had adopted laws permitting involuntary sterilization of the "socially inadequate."[33] That classification included many disabled individuals, including epileptics, the blind and deaf, and the "feebleminded" individuals whose learning disability caused them to perform poorly on IQ tests.[34] By the time the practice stopped some five decades later, approximately 65,000 Americans had been sterilized against their will.[35] In 1939, Adolf Hitler ordered the widespread euthanasia of newborns and children under three years of age who showed symptoms of mental retardation, physical deformity, or disability.[36] The program accounted for nearly a hundred thousand deaths by the time it was stopped.[37] While Hitler's atrocities typically garner more attention, state-sponsored sterilization in the U.S. in many ways paralleled the policies of Nazi Germany.[38]

The widespread social ostracism and abuse of individuals with disabilities began to change as injured soldiers returned home from the major wars of the twentieth century.[39] Starting in the early 1900s, Congress responded to an emerging social consciousness on disability by passing rehabilitation legislation intended to provide

---

[32]*Id.* at 207.

[33]Dolan DNA Learning Center, Sterilization Laws, http://www.eugenicsarchive.org/html/eugenics/static/themes/3.html (last visited Mar. 1, 2007).

[34]Ctr. for Individual Freedom, The Sterilization of America: A Cautionary History, http://www.cfif.org/htdocs/freedomline/current/in_our_opinion/un_sterile_past.html (last visited Mar. 1, 2007). The term feebleminded refers to mental retardation, regardless of functioning level, but is most often associated with mildly retarded, high functioning individuals. *See* Murray Simpson, *Developmental Concept of Idiocy*, 45 INTELL. & DEVELOPMENTAL DISABILITIES 23, 28 (2007), *available at* http://aaidd.allenpress.com/pdfserv/10.1352%2F1934-9556(2007)45%5B23:DCOI%5D2.0.CO%3B2. Based on the current understanding of disabilities, it seems clear that many of the "feebleminded" individuals ostracized from society could have contributed to the work force if given the proper accommodation.

[35]Ctr. for Individual Freedom, *supra* note 34.

[36]The History Place, Nazi Euthanasia, http://www.historyplace.com/worldwar2/holocaust/h-euthanasia.htm (last visited Mar. 1, 2007).

[37]*Id.*

[38]Jacqueline Weaver, *Study Finds Similarities in U.S. and Nazi Eugenics Efforts*, YALE BULL. AND CALENDAR, Feb. 18, 2000, *available at* http://www.yale.edu/opa/v28.n21/story10.html.

[39]*See* Polly Welch & Chris Palames, *A Brief History of Disability Rights Legislation in the United States, in* STRATEGIES FOR TEACHING UNIVERSAL DESIGN 5 (Polly Welch ed., 1995), *available at* http://www.udeducation.org/resources/readings/welch.asp; Nancy Murray, President, The Arc of Greater Pittsburgh, Address at the University of Pittsburgh Institute of Politics: Historical Overview of Disability Policy (May 2, 2003), *available at* http://www.wheelchairnet.org/WCN_Living/Docs/Historicaloverview.html.

opportunities for physically disabled individuals.[40] Over time, new laws were adopted to create opportunities for individuals with learning disabilities.[41]

Despite these ostensible advances, studies suggest that most people continue to harbor negative attitudes toward individuals with disabilities.[42] In some areas of the world, the barbaric practices of infanticide of the disabled and social purification may still continue.[43] Whether these prejudices are attributed to societal factors, and therefore subject to change, or represent an indelible condition of the human psyche is a matter of much debate.[44] Regardless, despite enduring centuries of societal and state-sanctioned ridicule, stigmatization, and physical abuse, disabled individuals remain at risk of discrimination in society.[45]

B. Early Disability Rights Movement and its Impact on Education

The disability rights movement in America has its roots in the establishment of the American School for the Deaf in 1817.[46] Another century passed, however, before Congress passed the first federal legislation impacting individuals with disabilities. The Smith-Sear Veterans Vocational Rehabilitation Act of 1918 established the first federal vocational rehabilitation program for soldiers with disabilities.[47] In the following decades, the legislation expanded and evolved from a

---

[40]San Francisco State Univ. Disability Programs and Res. Ctr., A Chronology of the Disability Rights Movements, http://www.sfsu.edu/%7Ehrdpu/chron.htm (last visited Mar. 1, 2007).

[41]*Id.*

[42]Kristin M. Lucas, Non-Handicapped Students' Attitudes Toward Physically Handicapped Individuals (May 4, 1999) (on file with the Missouri Western State University National Undergraduate Research Clearinghouse), *available at* http://clearinghouse.missouriwestern.edu/manuscripts/110.asp

[43]Michael Sheridan, *Deformed Babies Killed in Front of Mothers in Bid for Super Race*, AUSTRALIAN, Oct. 16, 2006, at 11 (World), *available at* http://www.theaustralian.news.com.au/story/0,20867,20587474-2703,00.html.

[44]*Compare* G. H. Neumann, *Prejudices and Negative Attitudes Towards the Disabled—Their Origin and Methods of Elimination*, 16 REHABILITATION (STUTTG) 101 (1977) (arguing that inborn human inclinations for a specific reaction towards marginal groups and fear-reaction towards strangers contribute to the formation of prejudices), *with* R. Zimmermann & H.J. Kagelmann, *Reactions Vis-a-Vis the Disabled: Critical Comments On G. H. Neumann's Article:* Prejudices and Negative Attitudes Towards the Disabled—Their Origin and Methods of Elimination *From the Viewpoint of Behavioral Science and Biology*, 17 REHABILITATION (STUTTG) 77 (1978) (arguing that prejudice against the disabled is a product of society and is subject to change).

[45]Interestingly, disability rights groups opposed Judge Samuel L. Alito's nomination to the United States Supreme Court out of fear that his narrow interpretation of the powers that authorize Congress to pass disability rights laws would remove protections afforded to the disabled. *See, e.g.*, Judge David L Bazelon Ctr. for Mental Health Law, Samuel Alito's Record on Disability Issues, http://www.bazelon.org/takeaction/alerts/alitosrecord-details.htm (last visited Mar. 1, 2007).

[46]American School for the Deaf, Museum/History, http://www.asd-1817.org/history/index.html (last visited Mar. 1, 2007).

[47]San Francisco State Univ. Disability Programs and Res. Ctr., *supra* note 40.

narrowly focused job movement for the physically disabled to comprehensive programs serving all people with disabilities. Amendments to early disability legislation culminated in passage of the Rehabilitation Act of 1973, (Act), the first federal civil rights legislation promulgated to specifically prohibit discrimination against individuals on the basis of physical, mental or emotional disabilities.[48] The Act prohibits federal entities from discriminating in the services that they provide on the basis of disability.[49] Because most educational institutions receive some form of federal funding, the Act fundamentally altered the landscape of education in America.[50] Two years later, Congress passed the Education for All Handicapped Children Act (now known as the Individuals with Disabilities Education Act, (IDEA)), which requires public elementary and secondary school systems to identify children with disabilities and to develop appropriate Individualized Education Plans (IEPs) for each child in exchange for receiving additional federal funds.[51] In 1990, the Americans with Disabilities Act, (ADA) was passed to promote the full participation of disabled individuals in all aspects of society by prohibiting discrimination by private entities, including private educational entities not covered by prior legislation.[52] Most importantly, the ADA required all schools to provide reasonable accommodations to students with disabilities.[53]

The positive impact early disability legislation had on education is evident from the dramatic rise in the number of disabled students attending undergraduate programs. Between 1978 and 1994, the number of first-time, full-time students with disabilities attending colleges and universities tripled from 2.6 percent to 9.2 percent.[54] Today, approximately one third of all high school graduates with disabilities have taken at least some post-secondary classes.[55] These changes,

---

[48]Rehabilitation Act of 1973, Pub. L. No. 93-112, 87 Stat. 355 (codified as amended at 29 U.S.C. §§ 701-97 (2006)).

[49]29 U.S.C. § 794 (2006).

[50]Nat'l Council on Disability, Access to Education by People with Disabilities: Illustrations of Implementation from the United States–Quick Reference Guide (Aug. 2, 2005) (Topical Overview for Delegates to the United Nations 6th Ad Hoc Committee on the Protection and Promotion of the Human Rights of People with Disabilities), *available at* http://www.ncd.gov/newsroom/publications/2005/pdf/access2education.pdf.

[51]Education for All Handicapped Children Act of 1975, Pub. L. No. 94-142, 89 Stat. 773; Education of the Handicapped Act Amendments of 1990, Pub. L. No. 101-476, 104 Stat. 1103, 1142 (codified as amended at 20 U.S.C. §§1400-06 (2006)); *see also* National Council on Disability, *supra* note 50.

[52]Americans with Disabilities Act of 1990, Pub. L. No. 101-336, 104 Stat. 327 (codified as amended at 42 U.S.C. §§ 12101-12213 (2006)); *see also* Nat'l Council on Disability, *supra* note 50.

[53]Nat'l Council on Disability, *supra* note 50.

[54]NAT'L COUNCIL ON DISABILITY, SOC. SEC. ADMIN., TRANSITION AND POST-SCHOOL OUTCOMES FOR YOUTH WITH DISABILITIES: CLOSING THE GAPS TO POST-SECONDARY EDUCATION AND EMPLOYMENT 56 (2000), *available at* http://www.ncd.gov/newsroom/publications/2000/pdf/transition_11-1-00.pdf.

[55]Doug Lederman, *College and the Disabled Student*, INSIDE HIGHER ED., July 29, 2005, http://www.insidehighered.com/news/2005/07/29/disabled.

however, have created new challenges for school administrators. Because schools are not allowed to pass off the cost of providing reasonable accommodations to their students, administrators must consider the potential financial impact of providing accommodations.[56] Although many accommodations cost little or nothing and require only simple modifications to the course structure, others are very expensive.[57] This is problematic because providing a costly accommodation to meet the needs of one disabled student may indirectly harm other non-disabled students.

For example, when a school spends thousands of dollars from its finite budget to provide a signer to a hearing impaired student, or to modify the structure of a building to make it more accessible, those funds are left unavailable to pay for other resources or instructional personnel that may improve the educational experience of other students.[58] Thus, the cost of providing accommodations to disabled students represents a legitimate concern for schools, particularly post-secondary schools that are not compensated for such expenditures. For some schools, the response is to simply ignore the issue.[59] The extent to which the concern impacts the admission of disabled students is impossible to quantify because admissions decisions are typically cloaked in secrecy.

Candidates for admission to undergraduate, graduate or professional degree programs often take some form of standardized test that purports to evaluate their potential for academic success. Because there is no way to accurately weigh the difficulty in course work or grade inflation across educational institutions, scores from these tests are considered by admissions committees in the selection process. Theoretically, this is because standardization places all test takers on an even playing field. However, studies show that scores are weighed differently at different schools. For example, a study conducted by the National Association for College Admission Counseling found that large universities, along with highly selective colleges, were more likely to place greater emphasis on scores received on standardized tests in admissions decisions than other institutions.[60]

---

[56]Paul D. Grossman, *Making Accommodations: The Legal World of Students with Disabilities*, ACADEME, Nov.-Dec. 2001, *available at* http://www.aaup.org/publications/Academe/2001/01nd/01ndgro.htm.

[57]Kevin H. Smith, *Disabilities, Law Schools, and Law Students: A Proactive and Holistic Approach*, 32 AKRON L. REV. 1, 75-76 (1999) (noting that the cost of providing certain accommodations to disabled students is prohibitive, and may cause schools to take money allocated for other educational programs to pay for the accommodation).

[58]*Id.* at 76.

[59]Scott Jaschik, *Enforcing the Disabilities Law*, INSIDE HIGHER ED., July 19, 2006, http://www.insidehighered.com/news/2006/07/19/ada (discussing the U.S. Justice Department's ongoing investigation of post-secondary institutions that have failed to comply with the ADA standards for accessibility); Scott Jaschik, *Who Decides?*, INSIDE HIGHER ED., July 31, 2006, http://insidehighered.com/news/2006/07/31/houston (discussing college policies that allow professors complete discretion on whether to comply with requests from students with disabilities and noting that some professors refuse to provide accommodations to students with documented disabilities).

[60]DAVID A. HAWKINS & MELISSA CLINEDINST, NAT'L ASS'N FOR COLL. ADMISSION COUNSELING, STATE OF COLLEGE ADMISSIONS 2006, at 31-32 (2006), *available at* http://www.nacacnet.org/MemberPortal/ProfessionalResources/Research/SOCA.htm (follow "Annual State of College Admission Report" hyperlink); s*ee also*, The Princeton Review,

The use of standardized tests is problematic for many disabled individuals because the form of the test or the manner in which it is administered may create barriers that prevent disabled students from demonstrating their true skills and abilities. To overcome these obstacles, reasonable testing accommodations are often made available for students with documented disabilities.[61] Such modifications may alter the exam presentation format, the manner in which an examinee may respond to a test question, the time period for taking the test, the location of the test administration, or other methods of properly compensating for an individual's disability.[62] The type of accommodation afforded depends on the type and severity of the individual's disability, and is typically evaluated on a case-by-case basis.[63] For example, a student with a learning disability may receive extra time to complete an exam, while another person with a more severe form of the same learning disability may receive extra time and other accommodations to take the same test.[64]

When an accommodation is provided on a standardized test, in some cases the test results forwarded to academic institutions are annotated, or "flagged," to indicate that the test was taken under nonstandard conditions, along with a warning that the scores should be interpreted cautiously. The practice of flagging unquestionably stigmatizes disabled individuals by informing admissions representatives that the applicant is disabled, by separating their tests scores from the pool of applicants, and by raising questions about the validity of their test scores.[65] In the highly competitive admissions process the opportunity for bias, conscious or unconscious, presented by the practice of flagging is clear.[66] In recognition of this undesired

---

Dispelling the Myths about the LSAT and Law School Admissions, http://www.princetonreview.com/home.asp (follow "law school" hyperlink; then follow "LSAT" hyperlink; then follow "learn more about the LSAT" hyperlink; then follow "Dispelling the Myths About the LSAT" hyperlink) (last visited Mar. 1, 2007) ("[T]he LSAT is the most important element of your law school application—regardless of whether or not you have a great academic background and GPA.").

[61]*See*, *e.g.*, ETS.org, Testing Resources for Test Takers with Disabilities: Documentation Criteria, http://www.ets.org (follow "Test Takers with Disabilities" hyperlink; then follow "Documentation Criteria" hyperlink) (last visited Mar. 1, 2007) (discussing how a disability is documented).

[62]CARA CAHALAN ET AL., PREDICTIVE VALIDITY OF SAT® I: REASONING TEST FOR TEST-TAKERS WITH LEARNING DISABILITIES AND EXTENDED TIME ACCOMMODATIONS 1-3 (Coll. Bd. Research Report No. 2002-5, ETS RR-02-11, 2002), *available at* http://www.collegeboard.com/research/pdf/rr20025 _11437.pdf.

[63]*Id.*

[64]*Id.*

[65]*See* ELLEN B. MANDINACH ET AL., THE IMPACT OF FLAGGING ON THE ADMISSIONS PROCESS: POLICIES, PRACTICES, AND IMPLICATIONS 32 (Coll. Bd. Research Report No. 2002-2, ETS RR-02-03, 2002), *available at* http://ftp.ets.org/pub/res/researcher/RR-02-03-Mandinach.pdf (noting that 72.9% of admissions officers surveyed admitted that when they saw a flagged test score, they assumed that the flag indicated that the test taker was disabled).

[66]*See id.* (noting that 4.0 percent of respondents admitted that they viewed the flag as an indication that the test score received is a less reliable or less accurate predictor of a student's potential for success and that 2.3 percent of respondents admitted that a flagged score may decrease the applicant's opportunity for admission in the program). Due to the covert nature

result, several testing entities ended the practice of flagging.[67] However, both the Law School Admissions Council (LSAC), which administers the Law School Admission Test (LSAT), and the Association of American Medical Colleges (AAMC), which administers the Medical College Admissions Test (MCAT), continue to use flags to denote that a test was taken under nonstandard conditions.[68] This practice is inconsistent with the spirit and intent of the ADA, and should be prohibited.

III. LEGALITY OF FLAGGING TEST SCORES

Proponents of flagging assert that the practice is necessary because available empirical data demonstrates that scores obtained from tests taken under nonstandard conditions may not be comparable to scores obtained from tests taken under standard conditions and, therefore, may not accurately predict future success in school.[69] Some accommodations, it is argued, fundamentally alter the nature of the construct measured and necessitate the use of flags to protect the integrity of scores received.[70] Opponents of flagging assert that the accommodations are necessary to level the playing field, but argue that flagging effectively restores the imbalance by allowing those who make crucial admissions decisions to know that an individual is disabled.[71] Given the prejudice against disabled individuals that has been exhibited across human history, it is argued, flagging puts some disabled students at a competitive disadvantage in the admissions process and indirectly exposes them to discrimination. Although a healthy debate has surfaced regarding the need for the flag, there has been little discussion on the legality of the practice.

A. Federal Law Applicable To Flagging

Section 504 of the Rehabilitation Act of 1973 prohibits post-secondary educational institutions receiving federal funds from discriminating against students

---

of the admissions process, obtaining empirical evidence of bias in the selection of candidates is almost impossible.

[67]*See* Press Release, Coll. Bd., The College Board and Disabilities Rights Advocates Announce Agreement to Drop Flagging from Standardized Tests (July 17, 2002), *available at* http://www.collegeboard.com/press/releases/11360.html [hereinafter Coll. Bd. Press Release]; Press Release, ACT, Inc., ACT Will End Practice of Flagging Test Scores Under Extended Time (July 26, 2002), *available at* http://act.org/news/releases/2002/07-26-02.html [hereinafter ACT Press Release].

[68]*See* Suria Santana, *Executive Council Decides to Continue MCAT Flagging*, AAMC REP., Nov. 2003, *available at* http://www.aamc.org/newsroom/reporter/nov03/mcat.htm; LAW SCH. ADMISSIONS COUNCIL, LSAT & LADAS INFORMATION BOOK 8 (2006), *available at* http://www.lsac.org/pdfs/2006-2007/informationbk2006.pdf (noting that students who receive extra time to take the test will have their scores annotated to reflect the accommodation).

[69]*See* Cohen, *supra* note 7.

[70]*See generally* Stephen G. Sireci, *Unlabeling the Disabled: A Perspective on Flagging Scores From Accommodated Test Administrations*, 34 EDUC. RESEARCHER 3 (2005), *available at* http://www.aera.net/uploadedFiles/Publications/Journals/Educational_Researcher/3401/2672-02_Sireci.pdf; AM. EDUC. RESEARCH ASS'N ET. AL., STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING 105 (1999).

[71]*See generally*, Sireci, *supra* note 70; AM. EDUC. RESEARCH ASS'N ET. AL., *supra* note 70.

on the basis of disability.[72] However, despite an emerging national social policy of inclusion for disabled individuals, agencies charged with implementing section 504 refused to promulgate any regulations until they were sued by a disabled research patient in 1976.[73] Current United States Department of Education, (DOE), section 504 regulations prohibit post-secondary institutions from denying admission or otherwise excluding qualified disabled individuals from educational programs based on the student's disability.[74] DOE's section 504 regulations also require recipients of federal funds to make modifications to their academic requirements that are "necessary to ensure that such requirements do not discriminate or have the effect of discriminating, on the basis of [disability], against a qualified . . . applicant" with a disability.[75] With respect to post-secondary education, a qualified person with a disability is one "who meets the academic and technical standards requisite to admission or participation in the recipient's education program or activity."[76] Post-secondary institutions receiving federal funding are required to utilize tests whose results "accurately reflect the applicant's aptitude or achievement level or whatever other factor the test purports to measure, rather than reflecting the applicant's impaired sensory, manual, or speaking skills (except where those skills are the factors that the test purports to measure)."[77] Further, institutions

> [m]ay not make use of any test or criterion for admission that has a disproportionate, adverse effect on [disabled] persons . . . unless (i) the test or criterion, as used by the recipient, has been validated as a predictor of success . . . and (ii) alternate tests or criteria that have a less disproportionate, adverse effect are [un]available.[78]

With few exceptions, educational institutions subject to section 504 are prohibited from making any "preadmission inquiry [into] whether an applicant for admission is a handicapped person."[79]

In 1997, DOE's Office for Civil Rights, (OCR), which overseas the fair and equitable provisions of accommodations, met with education experts and representatives from testing entities to discuss the legality of flagging test scores.[80]

---

[72]29 U.S.C. § 794 (2006) provides in pertinent part: "No otherwise qualified individual with a disability . . . shall, solely by reason of her or his disability, be excluded from the participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance . . . ." *Id.*

[73]In *Cherry v. Mathews*, 419 F. Supp. 922 (D.D.C.1976), the court held that Congress had intended regulations to be issued and ordered the United States Department of Health, Education and Welfare to immediately issue regulations.

[74]34 C.F.R. § 104.42(a), .43(a) (2006).

[75]34 C.F.R. § 104.44(a) (2006).

[76]34 C.F.R. § 104.3(l)(3) (2006).

[77]45 C.F.R. § 84.42(b)(3) (2006).

[78]§ 84.42(b)(2).

[79]§ 84.42(b)(4).

[80]*See* MANDINACH ET AL., *supra* note 65, at 6.

OCR's primary concern was that identifying disabled students through use of a flag could result in discrimination if admissions decisions were biased in favor of non-flagged scores.[81] Test administrators feared that removal of the flag would be professionally irresponsible in the absence of proof that the test scores obtained on standard and nonstandard tests were comparable.[82] Despite a lack of evidence regarding comparability between scores received with and without an accommodation, OCR issued an interim policy that allowed flagging to continue only "[u]ntil such time as a more viable policy can be worked out."[83] OCR later announced that a post-secondary education institution does not violate this regulation by "using test scores indicating that the test was taken under nonstandard conditions, so long as the test score is not the only criterion used for admission, and a person with a disability is not denied admission because the person with a disability took the test under nonstandard testing conditions."[84] OCR never repealed its policy despite finding that admissions personnel had violated the law by treating flagged scores differently.[85]

The Individuals with Disabilities Education Act (IDEA) (previously known as the Education of the Handicapped Act)[86] requires public elementary and secondary school systems to identify children with disabilities and to develop an appropriate Individualized Education Plan (IEP) for each child.[87] Importantly, the IDEA does not require colleges and universities to be proactive in identifying students with disabilities.[88] As a result, post-secondary institutions are required to provide

---

[81]*See id.*

[82]*See id.*

[83]MARTHA ROSS REDDEN ET AL., AM. ASS'N OF COLLEGIATE REGISTRARS & ADMISSIONS OFFICERS, AM. COUNCIL ON EDUC., RECRUITMENT, ADMISSIONS AND HANDICAPPED STUDENTS: A GUIDE FOR COMPLIANCE WITH SECTION 504 OF THE REHABILITATION ACT OF 1973, at 22 (1978), *quoted in* Diana C. Pullin & Kevin J. Heaney, *The Use of "Flagged" Test Scores in College and University Admissions: Issues and Implications Under Section 504 of the Rehabilitation Act and the Americans with Disabilities Act*, 23 J.C. & U.L. 797, 811 (1997), *quoted in* MANDINACH ET AL., *supra* note 65, at 6.

[84]*See* Duke University (N.C.), Complaint No. 04-91-2124, 4 NAT'L DISABILITY L. REP. 87 (Office for Civil Rights, Region VII April 2, 1993).

[85]*See* SUNY Health Science Center at Brooklyn—College of Medicine (N.Y.), Complaint No. 02-92-2004, 5 NAT'L DISABILITY L. REP. 77 (Office for Civil Rights, Region II Aug. 18, 1993). In this matter, the admissions committee admitted that they either devalued students' asterisked MCAT scores or weighted them in a different and lesser manner than non-flagged scores. OCR announced that a post-secondary educational institution violates 34 C.F.R. §§ 104.4(a), 104.4(b)(1)(ii), 104.4(b)(1)(iv), and 104.42(b)(1)(v) by "adopt[ing] a practice of devaluing the MCAT scores of individuals with disabilities who have taken the MCAT's under nonstandard conditions, thereby subjecting these individuals to differential treatment on the basis of disability." 5 NAT'L DISABILITY L. REP. 77.

[86]Education of the Handicapped Act Amendments of 1990, Pub L. No. 101-476, 104 Stat. 1103, 1142 (codified as amended at 20 U.S.C. §§1400-1406 (2006)).

[87]*See* 20 U.S.C. § 1401(14) (2000).

[88]OFFICE FOR CIVIL RIGHTS, U.S. DEP'T OF EDUC., STUDENTS WITH DISABILITIES PREPARING FOR POSTSECONDARY EDUCATION: KNOW YOUR RIGHTS AND RESPONSIBILITIES (rev. ed. 2007) (2002), *available at* http://www.ed.gov/about/offices /list/ocr/transition.html.

assistance only if the student voluntarily discloses his or her disability.[89] For various reasons, including a desire for privacy, fear of discrimination or reprisal, or simply embarrassment, some students may elect not to disclose their disability and struggle through the post-secondary curriculum without accommodations. Thus, for students who need accommodations but waive their right thereto, the grades received may not accurately reflect the student's true abilities.

The ADA expands on the essential concepts of the Rehabilitation Act by prohibiting discrimination on the basis of disability in virtually all aspects of society.[90] Title II prohibits discrimination on the basis of disability by public entities.[91] Department of Justice, (DOJ), regulations implementing Title II, mandate that public post-secondary educational institutions may not impose or apply eligibility criteria for admission that screen out or tend to screen out disabled applicants who are otherwise qualified.[92] Furthermore, such institutions must make reasonable modifications in policies, practices, or procedures to avoid discrimination on the basis of disability unless the institution can demonstrate that making the modifications would fundamentally alter the nature of the service, program, or activity provided.[93]

Title III of the ADA closely mirrors the provision in Title II, but applies to private entities.[94] DOJ regulations implementing Title III require private entities that administer examinations relating to applications for post-secondary education to offer the examinations "in a place and manner accessible to persons with disabilities or [to provide] alternative accessible arrangements for such individuals." [95] Each entity is required to provide reasonable modifications to the examination and appropriate auxiliary aids and services unless the entity can demonstrate that doing so would fundamentally alter the construct measured or would result in an undue burden.[96] Such modifications may include the provision of additional time to complete the examination.[97] Most importantly, each entity must assure that "the examination results accurately reflect the individual's aptitude or achievement level or whatever other factor the examination purports to measure, rather than reflecting

---

[89]*Id.*

[90]42 U.S.C. §12182(a) (2006) provides: "No individual shall be discriminated against on the basis of disability in the full and equal enjoyment of the goods, services, facilities, privileges, advantages, or accommodations of any place of public accommodation by any person who owns, leases (or leases to), or operates a place of public accommodation." *Id.*

[91]42 U.S.C. § 12132 (2006).

[92]28 C.F.R. § 35.130(b)(8) (2006). The application of this statute to state universities has been held an invalid exercise of congressional power and, thus, unconstitutional. *See* Press v. State Univ. of N.Y., 388 F. Supp. 2d 127 (E.D.N.Y. 2005).

[93]§ 35.130(b)(7).

[94]§ 12182(a).

[95]28 C.F.R. §36.309(a) (2006).

[96]§36.309(b)(3).

[97]§36.309(b)(2).

the individual's impaired sensory, manual, or speaking skills (except where those skills are the factors that the examination purports to measure)."[98]

Existing disability laws demonstrate a Congressional intent to place all individuals on an equal level in regard to education. However, neither the Rehabilitation Act nor, the ADA, require testing entities to give examinations that provide equal results for disabled and non-disabled test takers. Rather, entities must ensure that the test administered measures the skills of disabled and non-disabled test takers equally. To that end, entities are required to take the steps necessary to eliminate artificial barriers for disabled individuals that necessarily result from using tests that have been standardized based on the average non-disabled individual.[99] Unlike the provision of auxiliary aids, the requirement to select and administer tests that equally measure the actual abilities of disabled and non-disabled test takers is imposed regardless of the burden placed on the testing entity and regardless of whether the accommodation fundamentally alters the construct measured.[100] If testing entities met this requirement, flagging would have no purpose.

Although passed well after the practice of flagging had commenced, the ADA does not expressly address the legality of flagging scores received on tests taken by disabled individuals who were provided an accommodation. Because no other federal or state law expressly prohibits the practice of flagging standardized test scores, testing entities have looked within the educational profession for guidance in developing and administering tests.

B. Professional Rules Applicable to Flagging

The Standards for Educational and Psychological Testing (Standards) were developed jointly by the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME) to provide professional and technical guidance to testing entities to promote the sound and ethical use of tests and to provide criteria for the evaluation of testing practices. [101] The Standards recognize that tests designed for use with the general population may be inappropriate for use with individuals with disabilities if the person's disability impacts the results but is

---

[98]§36.309(b)(1)(i).

[99]*See* NAT'L CTR. FOR LEARNING DISABILITIES, HIGH STAKES ASSESSMENTS AND STUDENTS WITH LEARNING DISABILITIES: ASSESSMENT SYSTEMS MUST ENSURE NONDISCRIMINATION AND HIGH PARTICIPATION FOR ALL STUDENTS (n.d.), *available at* http://www.ncld.org/index.php?option=content&task=view&id=271 (last visited Mar. 1, 2007) ("Students with disabilities are usually not included in the sample population used in test development nor are students with disabilities, when included, given appropriate accommodations. This results in a lack of test validity . . . .") Given that the LSAT was formulated in 1974, well before the ADA, it is highly unlikely that the developers considered how long it would take students with particular disabilities to complete a set number of questions. *See id.*

[100]Unlike 28 C.F.R. § 36.309(b)(3), which allows a testing entity to refuse to provide an auxiliary aid as an accommodation if doing so fundamentally alters the construct measured or causes an undue burden, no such defenses are provided under 28 C.F.R. § 36.309(b)(1)(i) regarding the type of test administered.

[101]*See* AM. EDUC. RESEARCH ASS'N ET. AL., *supra* note 70, at v, 1.

otherwise irrelevant to what the test purports to measure.[102] The Standards also recognize that disabilities differ in degree and severity, requiring testing entities to tailor accommodations to the unique needs of each student.[103] For example, where a student requests additional time to complete an exam to compensate for his or her disability, Standard 10.6 urges testing entities to use available empirical evidence and professional judgment to determine the specific amount of additional time to allow.[104] Testing entities are discouraged from simply providing test takers with disabilities some multiple of the standard time allowed.[105] The goal is to ensure that an accommodation adopted is "appropriate for the individual test taker, while maintaining all feasible standardized features."[106] Standards 10.4 and 10.11 specifically address flagging test scores, and provide:

> 10.4[:] If modifications are made or recommended by test developers for test takers with specific disabilities, the modifications as well as the rationale for the modifications should be described in detail in the test manual and evidence of validity should be provided whenever available. Unless evidence of validity for a given inference has been established for individuals with the specific disabilities, test developers should issue cautionary statements in manuals or supplementary materials regarding confidence in interpretations based on such test scores. [107]
>
> . . . .
>
> 10.11[:] When there is credible evidence of score comparability across regular and modified administrations, no flag should be attached to a score. When such evidence is lacking, specific information about the nature of the modification should be provided, if permitted by law, to assist test users properly to interpret and act on test scores.[108]

Interestingly, the comment to Standard 10.4 notes that where a testing entity intends that a modified version of a test should be interpreted as comparable to an unmodified one, the testing entity should provide evidence of score comparability.[109] However, Standard 10.11 urges the use of flags where there is no evidence of score comparability between scores received on tests administered under standard and nonstandard conditions.[110] These Standards provide the means for testing entities to

---

[102] *See id.* at 100-01.

[103] *See id.* at 102.

[104] *See id.* at 107.

[105] *Id.*

[106] *Id.*

[107] *Id.* at 106.

[108] *Id.* at 108.

[109] *See id.* at 106.

[110] *Id.* at 108.

avoid performing the detailed, and likely costly, studies that are required to accurately demonstrate differences in test results.

Testing entities use a single test format because doing so, in theory, provides a means of objectively comparing each student's ability.[111] Providing an appropriate accommodation does not alter what is measured, because the accommodation only eliminates disability-related barriers that are irrelevant to what is measured by the test.[112] Therefore, test results obtained with or without the aid of an appropriate accommodation should be comparable. Proving this, however, requires extensive research. If there is difference between scores achieved on standard and nonstandard test administrations, the difference results from the failure to provide an accommodation that is appropriate.[113] Unfortunately, testing entities may avoid Standard 10.1 by relying on language in Standard 10.11 that urges the use of flags when there is no evidence of score comparability.[114] In essence, even if testing entities cannot demonstrate an actual difference in test results received by disabled and non-disabled test takers, they are urged to single out disabled individuals based on a perceived difference. Moreover, by providing an accommodation and then flagging the test score received, testing entities in effect reject a fundamental tenet of the ADA, i.e., that the provision of reasonable accommodations places the disabled students on a level playing field with other non-disabled students.[115]

The use of flags on tests taken under nonstandard time conditions allows testing entities to disregard Standards 10.6 and 10.10, which collectively urge entities to ensure that each accommodation granted is individually tailored to the student's unique needs as demonstrated by empirical evidence.[116] Rather than conduct the research necessary to determine the exact amount of time needed to appropriately compensate a test taker for his or her disability, entities approve a requested time accommodation and then flag the test results because they have no evidence to show whether the amount of time provided was or was not appropriate.[117] Moreover, as discussed below, studies on score comparability demonstrate that testing entities typically provide time accommodations based on some multiple of the standard time period, e.g., time and one-half or double time, rather than providing an amount of time appropriate for the individual as urged by the Standards.[118]

Because no explicit prohibition against flagging exists, testing entities that believe flagging is the only way to maintain the integrity of their test results continue the practice.[119] Whether the practice violates federal law depends, in part, on the degree to which scores from accommodated test administrations are comparable to

---

[111] *See, e.g.*, Sireci, *supra* note 70, at 4.

[112] *See, e.g.*, AM. EDUC. RESEARCH ASS'N ET AL., *supra* note 70, at 101.

[113] *See, e.g.*, *id.* at 106.

[114] *See id.* at 108.

[115] *See e.g.*, 28 C.F.R. § 36.309(b)(1)(i) (2006).

[116] *See* AM. EDUC. RESEARCH ASS'N ET AL., *supra* note 70, at 107-08.

[117] *Cf. id.* at 108 (allowing use of flag where evidence of comparability is lacking).

[118] *See id.* at 107; *see also* discussion *infra* Part IV.

[119] *See, e.g.*, Sireci, *supra* note 70, at 3.

scores from standard administrations. That is, if the scores are comparable, then flagging discriminates by unnecessarily segregating individuals with disabilities from the remainder of the applicant pool. However, if the scores are not comparable because the accommodation fundamentally alters the construct measured, then the use of flags to denote the difference likely does not violate the law. Thus, to assess the propriety of flagging one must first assess the empirical data available regarding scores obtained from standard and nonstandard test administrations.

IV. EVIDENCE FROM STANDARD AND NONSTANDARD TEST ADMINISTRATIONS

Proponents of flagging rely on a small number of studies on standardized tests, which conclude that scores achieved under standard and nonstandard testing conditions may not be comparable.[120] Standardized tests are widely used because they purport to provide objective measurements of an individual's aptitude for success in the field for which the test applies.[121] In the context of post-secondary education, standardized tests are designed to provide a statistically accurate prediction of a student's expected first year grade-point average.[122] Thus, the primary concern for educational institutions utilizing such test results is that the scores provide accurate information. To understand the argument that accommodations may invalidate the test scores received, one must consider the psychometric principles underlying standardized testing.

A. Standardization and Accommodation

Tests are required to be both reliable and valid.[123] "Reliability refers to consistency of results [whereas v]alidity refers to what a test measures and for whom it is appropriate." [124] A given test may provide accurate information for one purpose but not for another.[125]

In the context of flagging, test validity is of primary concern. The Standards define validity as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests," i.e., a test is considered valid if it measures what it claims to measure.[126] Validity may be assessed by correlating criterion with other criteria known to be valid.[127] When the

---

[120]*See, e.g.*, *id.* at 7.

[121]*See generally* Zenisky et al., *supra* note 4.

[122]*See, e.g.*, Anthony G. Picciano, EDTATS Primer: Review of Statistics – Reliability and Validity, http://www.hunter.cuny.edu/edu/apiccian/edstat25.html (last visited Mar. 1, 2007).

[123]*See, e.g.*, *id.*

[124]*Id.*

[125]*See* Vi-Nhuan Le & Stephen P. Klein, *Technical Criteria for Evaluating Tests*, *in* MAKING SENSE OF TEST-BASED ACCOUNTABILITY IN EDUCATION 51, 52 (Laura S. Hamilton et al. eds., 2002), *available at* http://www.rand.org/pubs/monograph_reports/MR1554/MR1554.ch3.pdf.

[126]*See* AM. EDUC. RESEARCH ASS'N ET. AL., *supra* note 70, at 9.

[127]*See* CollegeBoard.com, ACES: Validity Handbook: Types of Validity Evidence http://www.collegeboard.com/highered/apr/aces/post2006/vhandbook/evidence.html#criterion validity (last visited Mar. 1, 2007).

criterion measured is collected after the measure being validated, the goal is to establish predictive validity.[128] In the context of standardized testing, the test is used to predict how well a student will do at a later date.[129] For purposes of post-secondary admissions evaluation, standardized tests scores are typically correlated with high school or undergraduate grades to predict success at the next academic level.[130]

In some cases, tests may not accurately measure a desired construct, i.e., the concept or characteristic that a test is designed to measure, if the test omits something that should be included or adds something that is unnecessary, or both.[131] The validity of standardized test scores may be influenced by many factors, but in the context of flagging, construct-irrelevance variance poses the most significant problem.[132] Construct-irrelevance variance refers to the situation where scores are influenced by factors irrelevant to the construct being measured, e.g., when a test, designed to measure intelligence, is influenced by reading comprehension.[133] Because the standardized test format or method of administration may prove problematic to a disabled individual, accommodations are often provided in an attempt to eliminate construct-irrelevance variance.[134] For example, a vision impaired student may be provided with a test in Braille to ensure that the score results obtained are the result of his or her actual knowledge rather than a result of the student's inability to view the test questions.

Regardless of form, testing accommodations are designed to remove disability-related barriers to performance and allow a disabled person to demonstrate his or her "true" abilities.[135] In theory, the accommodation levels the playing field without altering the measurement goals of the assessment so that scores from the accommodated tests accurately measure the same construct as scores from the un-accommodated test.[136] When used effectively, accommodations increase the validity of the inference made from a test score.[137] However, where the modification alters what a test purports to measure, inferences made from the test result may be

---

[128]*See id.*

[129]*See* Dennis Gilbride, Reliability and Validity, http://suedweb.syr.edu/faculty/ddgilbri/assess/ReValid.html (last visited Mar. 1, 2007).

[130]*See* Robert L. Linn, *Admissions Testing: Recommended Uses, Validity, Differential Prediction, and Coaching*, 3 APPLIED MEASUREMENT IN EDUC. 297, 302-03 (1990).

[131]Samuel Messick, *Validity, in* EDUCATIONAL MEASUREMENT 13, 34 (Robert L. Linn ed., 3d ed. 1989).

[132]*See id.* at 34-35.

[133]Sireci, *supra* note 70, at 4.

[134]*Id.*

[135]SUSAN HATCHER & ANGELA WAGUESPACK, ACADEMIC ACCOMMODATIONS FOR STUDENTS WITH DISABILITIES, (Nat'l Ass'n of Sch. Psychologists, Helping Children at Home and School II: Handouts for Families and Educators S8, 2004), *available at* http://www.naspcenter.org/principals/nasp_accom.pdf.

[136]Sireci, *supra* note 70, at 4.

[137]*See generally id.*

invalid.[138] Thus, even though accommodations are intended to level the playing field, in some cases the accommodation may actually alter the construct measured and place individuals at a competitive advantage by allowing them to obtain a score higher than that which reflects their actual ability. Studies conducted to evaluate the impact of providing accommodations have focused primarily on the provision of additional time to disabled test takers.[139] Proponents of flagging base their argument mainly on predictive validity evidence drawn from a handful of studies that have compared test scores obtained under extended time conditions with those obtained under standard times.[140] As explained further below, the results of those studies are far from conclusive and do not support the argument for flagging.

B. Empirical Evidence of Score Comparability

The most common form of accommodation utilized in post-secondary test administration is the provision of extra time for learning disabled students.[141] Most of the data relevant to the issue of flagging comes from studies conducted on the comparability of scores obtained under standard and nonstandard time conditions.[142]

Ziomek and Andrews investigated differences between scores received on the ACT by students who took the test twice.[143] A subset of the study participants having disabilities were placed into three groups: those who took the test twice under extended time conditions both times (group I), those who initially tested under standard time conditions and then extended time conditions (group II), and those who took the test under extended time conditions first and then under standard time conditions (group III).[144] The ACT score scale ranges from 1 to 36.[145] Group I participants exhibited an average scaled score gain of 0.9, as compared to 0.7 for non-disabled students who took the test twice under standard conditions.[146] Group II students obtained a composite scale score gain of 3.2 points.[147] Interestingly, score gains depended in part upon the nature of the disability. Group II students with attention deficit disorder had higher gains (4.7 points) than those students with dyslexia (3.2) or learning disabilities (2.7) when they took the test with additional

---

[138]*Id.* at 4.

[139]*Id.* at 7.

[140]*Id.* at 7-9.

[141]*Id.* at 7.

[142]*See, e.g.*, *id.* at 7-9.

[143]ROBERT L. ZIOMEK & KEVIN M. ANDREWS, ACT ASSESSMENT SCORE GAINS OF SPECIAL-TESTED STUDENTS WHO TESTED AT LEAST TWICE 2 (ACT, Inc., ACT Research Report Series 98-8, 1998), *available at* http://eric.ed.gov/ERICDocs/data/ericdocs2/content_storage_01/ 0000000b/80/11/3c/d7.pdf.

[144]*Id.* at 3.

[145]ACT, Inc., ACT Research: Information Brief 2001-1, http://www.act.org/research/ briefs/2001-1.html (last visited Mar. 1, 2007).

[146]ZIOMEK & ANDREWS, *supra* note 143, at 5.

[147]*Id.*

time.[148] Group III students' scores were actually 0.6 points *lower*, indicating that testing time is a significant factor for disabled students.[149]

The study did not attempt to explain how much of the score gain was attributed to the provision of additional time or the nature of the individual's disability. The authors noted that with the same diagnosis, and across all groups, scores varied depending upon the testing conditions, and suggested that the performance might be associated with the degree of severity of the diagnosis.[150] The authors concluded that flagging should continue due to the difference in scores received from accommodated and non-accommodated tests.[151] Interestingly, rather than provide any analytical reason for their conclusion, the authors quote from two sources to support their position. The first, a paper, presented to the National Academy of Sciences Board of Testing and Assessment, "concluded that, '[a]fter years of research, the profession has insufficient evidence to conclude the scores given [sic] under nonstandard administrations mean the same thing as scores obtained under standard administrative conditions.'"[152] The second quote, taken from a book on testing disabled individuals, "concluded that '[t]he primary source of noncomparability that is directly associated with test scores is the extended time available in the nonstandard test administrations.'"[153] Interestingly, the author of the first paper also noted that flagging "apparently violates regulations written following passage of the [Rehabilitation Act [of 1973] and ADA] Acts."[154] The author of the book clarified that the source of the non-comparability is not the provision of extra time, but the provision of more time than the student needs to be fully compensated for his or her disability.[155] Thus, the conclusions drawn by Ziomek and Andrews are directly contradicted by the sources used in support of the conclusion. At best, the study demonstrates that individuals with disabilities require additional time on tests to demonstrate their actual ability and that the amount of time needed depends in part on the type and severity of the individual's disability.

Camara investigated improvement in scores on the Scholastic Aptitude Test I Reasoning Test (SAT I), between students with learning disabilities and students

---

[148]*Id.*

[149]*Id.*

[150]*Id.* at 7.

[151]*Id.* at 7, 9.

[152]*Id.* at 9 (second alteration in original) (quoting William A. Mehrens, Flagging Test Scores: Policy, Practice and Research 36 (1997) (unpublished manuscript, on file with Michigan State University, East Lansing, Mich.).

[153]*Id.* (quoting WARREN W. WILLINGHAM ET AL., TESTING HANDICAPPED PEOPLE 185 (1988)).

[154]Mehrens, *supra* note 152, at 36, *quoted in* Defendant's Exhibit 22, Doe v. Nat'l Bd. of Med. Exam'rs, No. CIV. A. 99-4532, 1999 WL 997141 (E.D.Pa. Nov. 1, 1999), *quoted in Doe*, 1999 WL 997141, at *8 (external alterations in original), *vacated*, 199 F.3d 146 (3d Cir. 1999).

[155]*See* WILLINGHAM ET AL., *supra* note 153.

without disabilities.[156] The study evaluated score changes for students who completed the SAT I during the spring of their junior year and repeated the test in the fall of the senior year.[157] The test subjects were separated into four groups: students without disabilities who took the test under standard conditions in each administration (group I), students with learning disabilities who took the test each time under extended time conditions (group II),[158] students with learning disabilities who first took the test under standard time conditions and then under extended time conditions (group III), and students with learning disabilities who took the test under extended time conditions first and then under standard time conditions, (group IV).[159] The results showed that group I students achieved a mean score gain of 13 and 12 points on the verbal and math scales, as compared to a gain of 15 and 12 points for group II students.[160] Group III students recorded mean score gains of 45 and 38, while group IV students recorded mean score losses of 9 and 6.[161] With the exception of group IV students, all scores improved with re-examination. For group III students, score gain increased in direct relation to the amount of extra time used on the test.[162] The results showed that all students benefit from taking the test a second time, and that disabled students benefit from additional time. The authors noted:

> A major problem with any analysis of the effects of accommodations for disabled examinees, such as the effects of extended time, is the difficulty in disaggregating the extent the modification compensates for the disability from the extent that it may overcompensate and introduce construct irrelevant variance (attributed to extra time to more carefully read, review, and respond to items) into the score.[163]

In a follow up study, Bridgeman investigated scores received on the SAT I when non-disabled students were given additional time. Those scores were compared to SAT I scores obtained under normal time administrations.[164] Allowing non-disabled

---

[156]WAYNE J. CAMARA ET AL., EFFECTS OF EXTENDED TIME ON THE SAT® I: REASONING TEST SCORE GROWTH FOR STUDENTS WITH LEARNING DISABILITIES (Coll. Bd., Research Report No. 1998-7, 1998), *available at* http://www.collegeboard.com/research/abstract/3850.html (follow "PDF" hyperlink).

[157]*Id.* at 4.

[158]*See Id.*

[159]*Id.* at 5.

[160]*Id.*

[161]*Id.*

[162]*Id.* at 11.

[163]*Id.* at 13.

[164]BRENT BRIDGEMAN ET AL., EFFECT OF FEWER QUESTIONS PER SECTION ON SAT® I SCORES (Coll. Bd., Research Report No. 2003-2, ETS RR-03-08, 2003), *available at* http://www.collegeboard.com/research/abstract/23501.html (follow "PDF" hyperlink).

students time and one-half to complete the exam resulted in score gains of 20 points on the math section and 10 points on the verbal section.[165]

Some have interpreted the results from the Camara and Bridgeman studies to suggest that learning disabled students benefit more from the provision of extra time, and may receive an advantage by receiving an accommodation that includes additional time.[166] However, such a conclusion is not supported in view of other facts. Because the Camara study did not attempt to segregate students based on the type or severity of their disability, it is impossible to determine whether the additional time provided adequately compensated an examinee for his or her disability or provided him or her with an advantage.[167] The nature and severity of disability differs markedly between individuals, and the same accommodation may produce different results.[168] Thus, the provision of additional time may place one disabled person on a level playing field with other test takers, while providing the same amount of additional time to a less disabled person may place him or her at a competitive advantage. Further, unlike the Bridgeman study, which provided each student with time and one-half to complete the exam, the Camara study set no time limit. Because the authors noted that scores improved in direct relation to the amount of time utilized, yet the scores were averaged together, it is impossible to know the extent to which those scores obtained using more than time and one-half skewed the average score gain observed. Finally, unlike in the Camara study, the Bridgeman study participants did not have the benefit of taking the test a second time. Camara found that disabled students who took the test two times under extended time conditions achieved score gains of 15 and 12 points on the verbal and math portions of the test, respectively, but achieved score gains of 45 and 38 on the same parts when they took the test first under standard time conditions and then again with extended time. Thus, when score gain associated with repeating the test under similar conditions (15 and 12) is removed from the score gain observed for use of an accommodation (45 and 38), the actual differences observed between disabled and non-disabled students taking the test with additional time is not as significant. That is, when the gain attributed to familiarity with the test itself is removed, students with disabilities exhibited average score gains of 30 and 26 points,[169] while students without disabilities exhibited gains of 20 and 10 points. Moreover, when one considers that some of the Bridgeman study participants achieved higher scores on part of the test than that achieved by disabled students under similar time conditions in the Camara study, the conclusion that disabled students receive a competitive advantage when granted an accommodation weakens considerably.[170]

---

[165]*Id.* at 10.

[166]Sireci, *supra* note 70, at 3, 8.

[167]CAMARA ET AL., *supra* note 156, at 13.

[168]*See* AM. EDUC. RESEARCH ASS'N ET. AL., *supra* note 70, at 101-04.

[169]*See generally id.* at 101-08.

[170]*See* BRIDGEMAN ET AL., *supra* note 164, at 10 (reporting that some students achieved score gains of 26 points on the math section with additional time, which is three points higher than the average score gain achieved by disabled students on the same math section with the same additional time).

Despite the complete absence of studies indicating how much time an individual with a particular type of disability requires to be fully compensated for his or her disability, flagging continues. Clearly, there are differences that render an accommodation appropriate for one individual and not another. As a result, flagged test scores may substantially misrepresent the extent to which a particular accommodation impacts the score received. Worse, a flag may operate to invalidate an otherwise valid test score if the accommodation is appropriate. Recognizing this problem, the authors in both studies recommended further research into the nature and extent of accommodations provided and their effects on test scores and grades.[171]

Cahalan evaluated the differential predictive validity of scores received on the Scholastic Aptitude Test I (SAT I), Reasoning Test between students with learning disabilities and non-disabled students under differing time conditions.[172] The test subjects were separated into three groups: test-takers without a disability who took the test under standard time conditions (group I); test-takers with a disability who took the test with extended time (group II); and test-takers with a disability who took the test without extended time (group III).[173] The study compared each student's SAT I score, with their first year college grade point averages (FGPA) to determine whether the scores received on the SAT I accurately predicted the student's FGPA.[174] The results showed that the correlation between the actual FGPA received and that predicted from the SAT I score received were slightly lower for some members of group II students than group I students. Specifically, scores received on the SAT I by group II students tended to over predict FGPA for males with disabilities, but accurately predicted FGPA for females with disabilities.[175] Interestingly, the study also found that when FGPA was predicted using both the SAT I scores and high school grade point averages, the correlation between the actual FGPA and the FGPA predicted by the SAT I score were the same for group I and group II students.[176] This result may reflect the fact that high schools are required under the IDEA to identify and assist disabled students, which in turn may allow those students to obtain grades that accurately reflect their abilities. The study also found that scores received on the SAT I in group I tended to over predict FGPA for males and under predict FGPA for females, and that the correlation was stronger when high school grade point averages were also considered.[177]

At best, the results from the Cahalan study suggest that scores from disabled male students who take the SAT I under extended time conditions may not be comparable to scores obtained by non-disabled students who take the test under normal time conditions. However, the authors cautioned that such a conclusion might not be reasonable.[178] Importantly, the authors noted that the study was limited

---

[171] CAMARA ET AL., *supra* note 156, at 14.

[172] CAHALAN ET AL., *supra* note 62.

[173] *Id.*

[174] *Id.* at 9.

[175] *Id.* at 1.

[176] *Id.* at 9.

[177] *Id.*

[178] *Id.* at 9-10.

due to the small number of disabled individuals in the test and that the sample size precluded examination of group differences such as undergraduate major, level of SAT score (high, medium or low), and severity of learning disability.[179] The degree to which differences in courses taken or the severity of disability affected each student's FGPA was not evaluated. Also, because the study did not segregate those disabled students who were provided with accommodations in class during their first year of college study from those who were not, it is impossible to determine how the failure to receive accommodations impacted the FGPA. Correlating the SAT I score to the FGPA, without considering first-year accommodations provides little insight into whether an accommodated test score actually over predicts an individual's likelihood of success. Further, validation studies on FGPA are restricted because of the lack of data regarding the number, level and rigor of courses learning disabled students take in their freshman year.[180] For some students, the fact that FGPA was lower than predicted by the SAT I may have had nothing to do with aptitude. Rather the difference may reflect the fact that the students were not provided with accommodations during their first year of studies. Indeed, the authors noted that further study should be performed that examines the accommodations disabled students request and receive at college and the effect such accommodations have on their FGPA.[181] The authors warned that the study results should be interpreted with caution due to the multiple types of accommodations utilized, variety and severity of disability, and the controversy regarding how each accommodation changes the test's constructs.[182]

Collectively, the studies conducted to date fail to demonstrate that there is a meaningful difference between test scores obtained from standard test administrations and those obtained from nonstandard administrations. At best, they provide ambiguous data which provides evidence of comparability as well as evidence of non-comparability. Properly viewed, the studies actually support the removal of flags because they provide no scientifically acceptable evidence to show that providing an appropriate accommodation changes that test construct measured or provides an unfair advantage to individuals with disabilities. Although few cases have been filed to prevent the practice of flagging, one lawsuit convinced some testing entities to stop the practice.

C. Legal Challenge and the Partial Demise of Flagging

In *Breimhorst v. Educational Testing Service*,[183] a disabled individual challenged Educational Testing Service's (ETS) practice of flagging test score reports taken under nonstandard conditions. Breimhorst, an individual without hands, took the Graduate Management Achievement Test (GMAT), on a computer using a track ball and was provided additional testing time to compensate for his disability.[184]

---

[179] *Id.* at 10.

[180] CAMARA ET AL., *supra* note 156, at 14.

[181] CAHALAN ET AL., *supra* note 62, at 9.

[182] *Id.* at 10.

[183] Breimhorst v. Educ. Testing Serv., No. C-99-CV-3387, 2000 WL 34510621 (N.D. Cal. Mar. 27, 2000).

[184] *Id.* at *2

Recipients of his test scores received a notation that indicated: "Scores obtained under special conditions."[185] ETS denied Breimhorst's requests to remove the information and to stop the practice of flagging, and Breimhorst filed suit.[186] The lawsuit alleged that the practice of flagging violates, *inter alia*, sections 302, 309, and 503(b) of the Americans with Disabilities Act, and section 504 of the Rehabilitation Act, by improperly suggesting that people with disabilities obtain an unfair advantage when they receive accommodations on standardized tests.[187] The lawsuit further alleged that ETS had no evidence to demonstrate that test results obtained from tests taken under nonstandard conditions are incomparable to those obtained from tests taken under standard conditions.[188] Further, the suit "allege[d] that ETS'[s] flagging policy [denied disabled individuals the] right to reasonable . . . accommodations by [indirectly] discouraging [disabled individuals] from requesting . . . accommodations," and punishing them by disclosing their disability to admissions personnel through use of a flag.[189] ETS moved for judgment on the pleadings, asserting that the practice of flagging violated no law.[190] The court granted ETS's motion regarding section 302 after finding that that section was not applicable to flagging, but denied the motion with regard to the other claims.[191]

In reaching its decision, the *Breimhorst* court held that the ADA "require[s] the test provider to take steps to best ensure that the tests equally measure the skills of disabled and non-disabled test takers."[192] The court noted that the "requirement is imposed regardless of the burden it causes the test provider."[193] Interestingly, the court opined,

> If test providers meet this burden, then there would be no reason to flag the test results of disabled test takers who receive accommodations. . . .
>
> . . . .
>
> . . . All test scores would accurately reflect the abilities of the test taker, regardless of whether the test was taken with accommodations for the test taker's disability.[194]

The court cautioned, however, that flagging may be appropriate if ETS takes steps to ensure equality, but the results still demonstrate that there is a significant

---

[185] *Id.*

[186] *Id.*

[187] *Id.* at *1.

[188] *Id.*

[189] *Id.* at *7.

[190] *Id.* at *2.

[191] *Id.* at *1, *7.

[192] *Id.* at *5 (emphasis omitted) (referring to 28 C.F.R. § 36.309(b)(1) (2006)).

[193] *Id.*

[194] *Id.* at *5, *8.

difference in test scores obtained from accommodated and non-accommodated test scores.[195]

Faced with the prospect of proceeding to trial, the parties settled after ETS agreed to stop flagging the GMAT, the Graduate Record Exam (GRE), the Test of English as a Foreign Language (TOEFL), and other exams it owned and administered.[196] The settlement did not include tests given for medical or law school, the SAT or the ACT college entrance exams.[197] However, the parties agreed to select a national panel of experts, jointly selected by Disability Rights Advocates and the College Board, to study the issue of flagging on the SAT I.[198] A majority of the panel, which included a leading psychometrician, recommended that the College Board "discontinue the practice of flagging the SAT I based on scientific, psychometric, and social evidence."[199]

After reviewing the empirical evidence, the panel concluded that there was "no evidence to suggest that the magnitude of the overall difference in predictive validity between standard and extended time administration warrants a cautionary flag to be attached to the scores of students who took the test under the condition of extended time."[200] The panel opined, that the goal of maintaining the integrity of the SAT I "should not result in a bias against applicants taking the SAT I with extended time when scientific, psychometric, and social evidence challenge the continued practice of flagging."[201] The panel noted that because the vast majority of students with learning disabilities are those with reading disabilities or dyslexia who require the accommodation of extra time, flagging tests taken under extended time "discriminates against a specific group of individuals[,] amplifies stereotypes, discourages students from applying for needed accommodations, and represents a profound and artificial barrier preventing students with disabilities, most often those with learning disabilities, from equal access to colleges and future careers."[202] Based on the panel's recommendation, the College Board agreed to stop the practice of flagging on all tests it owned, including the SAT, administered after October 1, 2003.[203] Shortly thereafter, the American College Testing, Inc., agreed to stop

---

[195]*Id.* at *5.

[196]*See* Tamar Lewin, *Disabled Win Halt to Notations of Special Arrangements on Tests*, N.Y. TIMES, Feb. 8, 2001, at A1.

[197]*Id.*

[198]*Id.*

[199]NOEL GREGG, ET AL., THE FLAGGING TEST SCORES OF INDIVIDUALS WITH DISABILITIES WHO ARE GRANTED THE ACCOMMODATION OF EXTENDED TIME: A REPORT OF THE MAJORITY OPINION OF THE BLUE RIBBON PANEL ON FLAGGING 2 (2002), *available at* http://www.dralegal.org/downloads/cases/breimhorst/majority_report.pdf.

[200]*Id.* at 8.

[201]*Id.* at 2.

[202]*Id.* at 4.

[203]Coll. Bd. Press Release, *supra* note 67. The College Board also owns and administers the Pre-SAT, NMSQT and AP exams. *See* COLL. BD., FREQUENTLY ASKED QUESTIONS ABOUT THE COLLEGE BOARD'S DECISION TO DROP FLAGGING FROM STANDARDIZED TESTS (2002),

flagging the test scores obtained on the ACT for students who took the test under extended time conditions.[204] Although some initial concern was voiced that the number of students seeking accommodation would increase in the absence of the flag, preliminary data indicates that no such increase has occurred.[205]

Despite the significant change brought about by *Breimhorst*, both the Law School Admission Council (LSAC), which owns and administers the LSAT, and the Association of American Medical Colleges (AAMC), which owns and administers the MCAT, steadfastly refuse to discontinue the use of flags on scores received under nonstandard test conditions.[206]

V. FLAGGING THE LSAT AND MCAT

Like the SAT, ACT, and other standardized college entrance exams, the LSAT and MCAT are used to predict a student's success during the first year of studies. The test score received on either exam is weighed differently by each institution, but constitutes a major factor in most admissions decisions.[207] However, unlike other tests, both the LSAT and the MCAT are intentionally speeded.[208] Due to a general perception that the provision of additional time changes the construct measured, i.e., speed of analysis or reasoning, both the LSAC and AAMC continue to flag scores obtained through nonstandard administration of the LSAT or MCAT. As explained below, the decision is unwarranted in view of the empirical evidence available.

A. LSAT and Score Comparability

The LSAT is designed such that the score obtained accurately reflects an individual's ability to make logical decisions under pressure.[209] In theory, students

---

available at http://www.collegeboard.com/prod_downloads/about/news_info/SATFlagFAQ72402B.pdf.

[204]ACT Press Release, *supra* note 67.

[205]ACT, Inc., ACT's Decision to Stop Flagging ACT Scores Achieved with Nonstandard Time: Questions and Answers, http://act.org/aap/disab/flag.html#increase (noting that the number of requests for accommodations has increased in proportion to the increased numbers of students taking the ACT).

[206]Santana, *supra* note 68.

[207]*See* Council on Legal Education Opportunity, Applying to Law School: Law School Admissions Test, http://www.cleoscholars.com/applying_to_law_school/lsat.cfm (last visited Mar. 1, 2007) (noting that a student's LSAT score is one of the most important factors an admissions committee will initially assess).

[208]WILLIAM D. HENDERSON, SPEED AS A VARIABLE ON THE LSAT AND LAW SCHOOL EXAMS 2 (Law Sch. Admission Council, Research Report Series No. 03-03, 2004), *available at* http://www.lsacnet.org/Research/Speed-as-a-Variable-on-the-LSAT-and-Law-School-Exams.pdf (noting that speed is a factor tested on the LSAT). A test is generally considered to be speeded if less than 100 percent of the examinees reach 75 percent of the test items and less than 80 percent of the examinees finish the test. *See* NORMAN G. PETERSON, REVIEW OF ISSUES ASSOCIATED WITH SPEEDEDNESS ON THE GATB TESTS 2 (1993), *available at* http://www.onetcenter.org/dl_files/Speed_GATB.pdf.

[209]HENDERSON, *supra* note 208, at 22 (noting that although the test is intended to test reasoning ability it is speeded and suggesting that test-taking speed may help explain the uneven predictive power of the LSAT).

should be able to process a certain amount of information in a certain amount of time to reach a proper result. Speeded standardized tests necessarily discriminate against individuals with disabilities who, as a result of their disability, are unable to work or process information at the speed of the average non-disabled test taker. Because the LSAT is speeded, and because available empirical data has been interpreted to suggest that tests taken under nonstandard test conditions may not be comparable to tests scores obtained under standard conditions, the LSAC continues to flag scores obtained on the LSAT under nonstandard time conditions.

In the only comprehensive study regarding flagging on the LSAT by Thornton researchers evaluated the predictive validity of LSAT scores obtained under extended time conditions.[210] The purpose of the study was to determine whether LSAT scores obtained under nonstandard time conditions are comparable to those obtained under standard time conditions.[211] The study considered each student's LSAT score, Undergraduate Grade Point Average (UGPA), and first year law school grade point average (FLGPA) to determine whether a given LSAT score received under normal or accommodated testing conditions accurately predicted the student's FLGPA.[212] Because of the different grading scales used for first year grades in different law schools, first year averages were standardized within each entering class to have a mean of 50.[213] For all participants, the mean LSAT score was 156.23, and the mean UGPA was 3.23.[214] These numbers were then used as the standard for creating an index of expected achievement.[215] That is, an LSAT score and UGPA of 156.23, and 3.23 respectively predicted a FLGPA of 50.[216] The study results showed that students with disabilities who received extra time on the exam had a mean LSAT score of 157.57 and an UGPA of 3.1, which correlated with a predicted first year grade index of 49.93. The students' actual indexed first year grades were 44.85.[217] For those non-disabled students taking the test under standard conditions, the mean LSAT score was 156.22 and the mean UGPA was 3.23, which translated into predicted first year grade index score of 50.00. The students' actual first year index grade was 50.06.[218]

The results of the Thornton study suggest that LSAT scores earned under accommodated testing conditions tend to over predict FLGPA, i.e., a disabled student's predicted first year law school average tended to be higher than their actual

---

[210] ANDREA E. THORNTON ET AL., PREDICTIVE VALIDITY OF ACCOMMODATED LSAT SCORES (Law Sch. Admission Council, Technical Report No. 01-01, 2001), *available at* http://members.lsacnet.org/ (follow "LSAT and legal education research publications" hyperlink; then follow "Full Research Report" hyperlink).

[211] *Id.* at 1.

[212] *Id.*

[213] *Id.* at 7.

[214] *Id.* at 8.

[215] *Id.*

[216] *Id.*

[217] *Id.*

[218] *Id.*

FLGPA.[219] The results also showed that disabled students who obtained an accommodation that did not include additional time achieved scores that over predicted their first year success, but the over prediction was slightly less than those students whose accommodation included additional time.[220] Based on the data, the authors concluded that, "LSAT scores obtained under accommodated conditions that include extra testing time are not comparable to LSAT scores obtained under standard time conditions."[221]

The LSAC uses the findings of the Thornton study to support the continued flagging of test results obtained under nonstandard time conditions. Score reports obtained on tests taken under nonstandard time conditions that are sent to law schools are accompanied by a letter that provides in part:

> [T]his applicant took the . . . LSAT under nonstandard time conditions . . . . LSAC research indicates that scores earned under nonstandard time conditions do not have the same meaning as scores earned under standard time conditions . . . .
>
> This applicant's score should be interpreted with great sensitivity and flexibility . . . . [222]

There are multiple problems with utilizing the conclusions drawn from the Thornton study to support the practice of flagging. Most importantly, the authors noted that "no information [was] available concerning possible accommodations test takers . . . received at their undergraduate institution or [were] receiving (or may have received) at their law school."[223] This is a critical omission that renders the conclusions invalid, because other studies have demonstrated that time also plays a factor in first year law school exams.[224] In the absence of any correlative data regarding accommodations provided during first year law school exams, there is no way to accurately determine whether the over prediction observed in FLGPA results from the provision of extra time on the LSAT or the failure to provide adequate time accommodations during the first year law school exams. If a student is provided with a time accommodation on the LSAT that compensates him or her for a disability, but is not provided a similar accommodation during first year exams, one should expect that the LSAT will over predict FLGPA to some degree. Indeed, the authors warned that, "[c]aution should always be exercised when drawing conclusions from the type of data analyzed in this study."[225]

---

[219] *Id.*

[220] *Id.*

[221] *Id.* at 32.

[222] Letter from Accommodated Testing, Law School Admissions Council, to Admissions Office, Barry University School of Law (form letter rev. June 2003) (on file with author) (obtained from the admissions file of a student with a disability who took the Oct. 2003 LSAT).

[223] THORNTON ET AL., *supra* note 210, at 2.

[224] HENDERSON, *supra* note 208, at 22.

[225] THORNTON ET AL., *supra* note 210, at 32.

Another problem with the conclusion reached by Thornton is that no effort was made to standardize undergraduate grade point averages beyond that done by the LSAC.[226] The LSAC does not attempt to adjust the UGPA to reflect course difficulty, number of courses taken, length of study, potential grade inflation, etc.[227] Instead, it requires students to submit their transcripts to the Law School Data Assembly Service (LSDAS),[228] which uses a formula to recalculate the student's GPA that is then used by admissions personnel in the decision process.[229] The problem with this practice is that schools grade differently, and the recalculated GPA depends in part on the school attended. For example, some schools award pluses and minuses with letter grades, while others use one or neither designation. Problems result when grades from different institutions are converted using the LSDAS scale. For example, an A+ is converted to a 4.33 on the LSDAS scale, while an A is converted to a 4.0.[230] The student attending the school that awards pluses is provided an advantage over the student attending the school that does not use pluses, without any evidence that the grades received reflect actual differences in achievement. Other than being fundamentally unfair to some students, the LSDAS scaled scores may not accurately reflect the student's UGPA. As a result, any attempt to correlate the recalculated UGPA with the student's LSAT score to predict FLGPA adds additional error into the analysis and renders any conclusions reached suspect at best.

Another problem with using Thornton's conclusions to support the practice of flagging is that law schools often apply different standards in evaluating students than undergraduate programs. The Council of the Section of Legal Education and Admissions to the Bar (Council) of the American Bar Association (ABA) is charged with accrediting law schools, and has promulgated certain standards to serve that objective.[231] Standard 301 requires law schools to "maintain an educational program that prepares its students for admission to the bar and effective and responsible participation in the legal profession."[232] In evaluating whether a law school complies with standard 301, the council considers "the rigor of [the law school's] academic program, including its assessment of student performance, and the bar passage rates of its graduates."[233] In an effort to assure compliance with accreditation standards,

---

[226]*Id.* at 7.

[227]*Id.*

[228]Law Sch. Admission Council, About the LSDAS, http://www.lsac.org/LSAC.asp?url=/lsac/lsdas-general-information.asp (last visited Mar. 1, 2007).

[229]Law Sch, Admission Council, Support Center: Frequently Asked Questions – LSDAS, http://www.lsac.org/LSAC.asp?url=/lsac/faqs-and-support-lsdas.asp#9 (last visited Mar. 1, 2007).

[230]*See* Pre-Law Handbook, Improving LSDAS and LSAC, http://www.prelawhandbook.com/improving_lsdas_and_lsac (last visited Mar. 1, 2007).

[231]Section of Legal Educ. & Admissions to the Bar, American Bar Association, Standards for Approval of Law Schools Foreword, http://www.abanet.org/legaled/standards/foreword.html (last visited Mar. 1, 2007).

[232]Section of Legal Educ. & Admissions to the Bar, American Bar Association, Standards for Approval of Law Schools Chapter 3: Program of Legal Education, http://www.abanet.org/legaled/standards/chapter3.html (last visited Mar. 1, 2007).

[233]*Id.*

many law schools impose a mandatory grade curve on faculty calculating first year grades.[234] Because of the grade curve, a student may receive a numerical score indicative of superior work, yet actually receive a grade that reflects lower quality work if too many students receive similar grades. In some cases, a single point can make the difference between letter grades and have a big impact on the student's FLGPA. For example, if one student scores a 94 on a test, and another scores a 93, the higher scoring student may receive an A, or a 4.0 GPA while the other receives an A-, or a 3.7 G.P.A.[235] Thus, unlike other institutions where a certain numerical score typically guarantees a certain letter grade, a score on a law school exam carries no such guarantee. According to the logic employed by Thornton if an LSAT score obtained with an accommodation predicts a FLGPA of a 4.0, but the student's actual FLGPA is a 3.7, the accommodated test score over predicts future success and should be flagged. The answer is not that simple. There are other factors, including the effect of the grade curve and the type of accommodation provided, at work that may contribute to the differential predictive capacity provided by the LSAT. Yet, to address the problem the LSAC continues to take the path of least resistance by flagging test scores taken under nonstandard conditions.

Of the 123,065 participants in the Thornton study only 1,249 received a time accommodation, yet the index scores were standardized for the entire group of participants.[236] As a result, the predictions reached are heavily skewed in favor of what is expected from the general student population, most of whom are not disabled. Ironically, despite recognizing that individuals have unique disabilities that require case-by-case evaluations of requests for testing accommodations, LSAC utilizes tests that are not designed to make judgments about individuals then flags the results because they do not know what the score means.[237]

B. MCAT and Score Comparability

Like the LSAT, the MCAT is a speeded exam.[238] Responding to criticism regarding the use of flags, Julian conducted a study to compare scores of MCAT

---

[234]*See e.g*, Seton Hall Law School, Present Grading Curves, http://law.shu.edu/administration/registrar_bursar/examinations_and_grading/grading_curves.htm (last visited Mar. 1, 2007) (discussing mandatory grade curves for required and elective courses); William & Mary Marshall-Wythe School of Law, Academic Regulations Grading Policy, http://www.wm.edu/law/academicprograms/regulations/grading.shtml (last visited Mar. 1, 2007) (discussing mandatory grade curves for classes of 30 or more students, which includes first year classes).

[235]To meet the grade curve imposed, the author has on many occasions been required to issue letter grades wherein the difference in letter grades was based on a small difference in numerical scores obtained among students.

[236]THORNTON ET AL., *supra* note 210, at 6.

[237]SUSAN C. ANTHONY, UNDERSTANDING STANDARDIZED TESTS (1997), *available at* http://www.susancanthony.com/pdfhandouts/Testhandout.pdf (noting that standardized tests are more useful for making judgment about large groups of people than individuals).

[238]*See generally* SCOTT H. OPPLER ET AL., THE EFFECT OF SPEEDEDNESS ON MCAT SCORES: AN INITIAL EXAMINATION (Am. Insts. for Research, prepared for the 19th annual conference of the Society for Industrial and Organizational Psychology, Inc., 2004), *available at* http://www.air.org/news_events/Documents/ SIOP2004MCATSpeed.pdf.

examinees who took the test under standard conditions to those of examinees who took the test under nonstandard conditions.[239] The study examined scores received on the MCAT between 1994 and 2000, broken down by standard and nonstandard administrations. Flagged scores were classified by disability, e.g., learning disability (LD), attention deficit hyperactivity disorder, (ADHD), etc. The results showed that examinees taking the MCAT with extended time achieved higher mean scores on all sections of the test.[240] Students who took and then retook the MCAT under standard conditions achieved an average increase of 1.5, whereas students with disabilities who took the test first under standard conditions and then retook it with additional time gained an average of 6.5 points.[241] Interestingly, students with ADHD scored higher than those with LD when both were provided with the same additional amount of time.[242] The authors noted that the higher scores may reflect an overcompensation for some of the students with disabilities, but added:

> A variety of factors might explain the higher mean scores of examinees granted special accommodations . . . . One explanation could be that accommodations properly compensated for the test-takers' disabilities, and that the flagged population had slightly more academic ability than the standard examinees.
>
> One reason that might happen . . . is that students with learning disabilities, having succeeded for years in meeting academic challenges, are more aware of their strengths and weaknesses, and may be more realistic in assessing their educational potential, so fewer of them take the MCAT when unprepared.[243]

The authors noted that one way to determine whether a certain time accommodation adequately or overcompensates an individual for a disability is to carefully assess the processing speed of each person requesting extended time.[244] The authors concluded that it remained unclear whether allowing disabled students additional time fundamentally alters what the MCAT measures, and noted that the validity of the flagged MCAT is unknown.[245] Interestingly, the American Medical Student Association, (AMSA), opposes the practice of flagging, commenting:

> AMSA does not support flagging; we support the right of individual students to choose whether or not to disclose to admissions committees

---

[239]Ellen R. Julian et al., *The Impact of Testing Accommodations on MCAT Scores: Descriptive Results*, 79 ACADEMIC MED. 360 (2004).

[240]*Id.*

[241]*Id.* at 363.

[242]*Id.* at 362.

[243]*See* Suria Santana, '*Flagging' Debate Continues; Data Show Score Fluctuations*, AAMC REP., Jan. 2003, *available at* http://www.aamc.org/newsroom/reporter/jan03/flagging.htm (internal quotation marks omitted) (quoting Ellen R. Julian, Director of the MCAT).

[244]*Id.*

[245]Julian et al., *supra* note 239, at 363.

> that they received testing accommodations because of a [documented] disability . . . . Generally, students receive testing accommodations because they have a diagnosed learning disability, and this is a matter of privacy. There haven't been any studies that show that students with learning disabilities make poor clinicians in the long run, and because of it, there's no basis to place a flag on their score.[246]

Collectively, the results from the studies conducted on the LSAT and MCAT are too inconclusive to support the continued use of flags. Despite this fact, no lawsuit has directly challenged the use of flags on either test. In the only case to directly address the propriety of flagging scores received on a professional exam, the court reached a result that is inconsistent with the spirit and scope of disability laws.

C. Raising the Flag: Judicial Response to Flagging Professional Exams

In *Doe v. National Board of Medical Examiners*, a disabled medical student challenged the National Board of Medical Examiners' (NBME), practice of flagging scores obtained on the United States Medical Licensing Examination (USMLE).[247] The lawsuit sought to enjoin the practice of flagging, and alleged that as applied to the student, flagging violated Title III of the ADA. The District Court granted the motion, holding, *inter alia*, that Doe demonstrated a reasonable likelihood of success on his claim that flagging his test score violated section 302 of the ADA, and demonstrated that absent an injunction he would be irreparably harmed.[248] The court declined to make a finding regarding the comparability between time-accommodated scores and scores achieved under standard conditions, but noted that all expert witnesses agreed that additional research was needed to properly evaluate the issue.[249]

Interestingly, the court found that the NBME failed to establish that flagging was necessary or that flagging test scores would fundamentally alter its services, and opined:

> What is clear from the Willingham study, Dr. Mehrens' paper, and the Testing Standards, is that: (1) the research on comparability of standardized scores and scores where time-related accommodations are given is too sparse to support any definitive conclusions; (2) the available data on handicapped examinees necessary to allow such research is severely limited; and (3) there is no simple psychometric solution to the question of interpreting scores for persons with disabilities, given competing legal, psychometric, ethical and practical concerns.[250]

---

[246]Santana, *supra* note 68 (first alteration in original) (internal quotation marks omitted) (quoting Sophie Jan, Chair, AMSA Advocacy Action Committee).

[247]Doe v. Nat'l Bd. of Med. Exam'rs, No. CIV. A. 99-4532, 1999 WL 997141, at *1 (E.D.Pa. Nov. 1, 1999), *vacated*, 199 F.3d 146 (3d Cir. 1999).

[248]*Id.* at *3.

[249]*Id.* at *10.

[250]*Id.* at *11 (footnote omitted).

The court noted that the practice of compelling a disabled individual to take an exam and then allowing the test sponsor to flag the score received because the sponsor is not sure what the score means represents "precisely the type of discrimination that is prohibited by Title III of the ADA."[251] Finding that NBME's annotation policy ran afoul of the general prohibition against discrimination set forth in section 302 of the ADA, the court enjoined NBME from flagging Doe's test scores, and ordered NBME to report the scores as though Doe took the test without accommodation for his disability.[252]

The Third Circuit Court of Appeals agreed that NBME's annotation of Doe's test score injured Doe, but found that the practice of flagging "does not constitute an ipso facto violation of Title III of the ADA."[253] The court noted that the trial court erred in analyzing the flagging issue under section 302 of the ADA, because section 309, the more specific statute governing discrimination by providers of examinations, controlled the issue.[254] The court also found that Doe had the burden of proving that his test scores were comparable to non-accommodated test scores.[255] Finding that Doe had not met his burden, the court vacated the order granting the preliminary injunction.[256] In reaching its decision, the court noted that Doe failed to demonstrate that he was likely to "suffer discrimination" as a "result of an annotation to his scores," and noted that if Doe were subject to discrimination by employers, that discrimination would not be attributable to the NBME.[257]

The appellate court's decision in *Doe* is problematic because it recognizes that section 309 requires testing entities to modify examinations to eliminate features of standardized tests that disadvantage disabled test takers, without recognizing that testing entities continue to ignore the mandate. The court failed to recognize that flagging is utilized by testing entities to mask their own failure to take appropriate steps to ensure that the tests used are applicable to all test takers. Providing an accommodation does not solve the problem if score results are flagged to indicate that they may not mean the same thing as scores received without an accommodation.

---

[251]*Id.* at *14.

[252]*Id.* at *9, *14 (basing decision, in large part, on 42 U.S.C. § 12182(b)(1)(A)(ii) (2006)). 42 U.S.C. § 12182(b)(1)(A)(ii) prohibits discrimination, on the basis of a disability, that denies an individual "the opportunity to participate in or benefit from a good, service, facility, privilege, advantage, or accommodation that is not equal to that afforded to other individuals." 42 U.S.C. § 12182(b)(1)(A)(ii).

[253]*See* Doe v. Nat'l Bd. of Med. Exam'rs, 199 F.3d 146, 149 (3d Cir. 1999) (citing 42 U.S.C. § 12189 (2006)). 42 U.S.C. § 12189 provides: "Any person that offers examinations or courses related to applications, licensing, certification, or credentialing for secondary or post-secondary education, professional, or trade purposes shall offer such examinations or courses in a place and manner accessible to persons with disabilities or offer alternative accessible arrangements for such individuals."

[254]*Id.* at 149.

[255]*Id.* at 156-57.

[256]*Id.* at 156-57, 158.

[257] *Id.* at 158.

Interestingly, even though Doe requested only time and one-half to take the exam, the NBME provided Doe with double time because the computerized version of the exam he took allowed only an accommodation of double time. Studies that have evaluated the comparability of scores obtained with additional time have suggested that in some cases the amount of time provided may overcompensate for the individual's disability and provide an advantage to the student. If true, then the provision of more time than deemed medically necessary by a test taker's health care provider may actually cause the alleged comparability problem. The *Doe* court failed to recognize that NBME's inability to provide the exact amount of additional testing time deemed medically appropriate contributed to the problem NBME argued it was trying to address.

The appellate court's conclusion that any discrimination Doe experienced in applying to residency and internship programs would not be attributable to the NBME is inconsistent with the court's conclusion that flagging harmed Doe. Flagging harms because it improperly informs recipients of test scores that the test taker has a disability. The court's argument places the proverbial cart before the horse by assuming that any discrimination Doe might experience would result from his interaction with representatives from the residency or internship programs during the interview process. By placing a flag on Doe's test score, NBME allowed representatives to factor Doe's disability into their decision to grant an initial interview. The court failed to consider the fact that there is no way to know how those same representatives used that information. Given the prejudice exhibited against disabled individuals, it is not unreasonable to assume that for some representatives Doe's disability would be a factor, conscious or unconscious, considered in the decision.

VI. RECOMMENDATIONS

Testing entities must meet a dual obligation to provide reasonable accommodations to disabled test takers while assuring that the scores reported represent valid estimates of an individual's likelihood of success. To meet this obligation, testing entities have elected to travel the path of least resistance by providing accommodations then annotating the test scores received to indicate that the scores do not mean the same thing as scores received without an accommodation.

The major problem with flagging test scores is that there is simply no evidence to support the assumptions used to support the practice. The studies cited by proponents of flagging all share the same common flaws. First, the data only weakly supports an inference that tests taken under nonstandard time conditions are incomparable to tests taken under standard time conditions. When supporting evidence is considered, the distinction becomes even less clear. Moreover, the available data provides evidence of score comparability and non-comparability. In every study conducted to date, the authors cautioned against interpreting the results too broadly due to inherent limitations and gaps in the data available. Such weak empirical evidence is insufficient to support the assumption and inferences made by testing entities. Yet, the practice of flagging continues, exposing countless students to the possibility of discrimination and unequal access to post-secondary education. To address the inequity of flagging, the following recommendations are advanced.

A. Amend Titles II and III of the Americans with Disabilities Act

The practice of flagging tests scores unquestionably violates the spirit of the Rehabilitation Act of 1973 and the ADA by segregating disabled test takers from non-disabled test takers, by circumventing the prohibition against preadmissions inquiries, and by placing disabled students at a competitive disadvantage as a direct consequence of having a documented disability. Regulations implementing section 504 of the Rehabilitation Act prohibit post-secondary institutions from using tests for admissions that have an adverse effect on disabled persons unless the test has been validated as a predictor of success. By utilizing flagged test scores to indicate that the scores may not be valid, post-secondary institutions run afoul of this mandate. Further, Title III of the ADA requires testing entities to develop and utilize tests that measure the abilities of disabled and non-disabled students equally.

The use of flags provides clear evidence that testing entities continue to violate Title III by utilizing tests, the results of which may not offer valid predictions of the student's potential for future success. Indeed, the express purpose of flagging is to inform score recipients that a score may not accurately reflect the student's abilities, but may be influenced by other factors including the student's disability. Flags are improperly utilized by testing entities to mask their own failure to develop and utilize tests that accurately and equally measure the abilities of disabled and non-disabled students. As a result, disabled students who take tests with accommodations may be subjected to discrimination within the admissions process. This is exactly the type of potential for harm the ADA and the Rehabilitation Act sought to address.

To fulfill the goals of the ADA and the Rehabilitation Act, testing entities should be required to develop and administer standardized tests that allow disabled students to demonstrate their true abilities. Perhaps the simplest way of ensuring this is to amend Title II and Title III of the ADA to expressly prohibit the practice of flagging. Such a rule would comport with the ADA's general mandate that tests equally measure all candidates. As a practical matter, it may be impossible to develop a single test that accurately and fairly measures all students' abilities. Given the need to maintain test integrity and validity, another possible solution might be to amend the ADA to expressly prohibit the practice of flagging unless the testing entity can establish, based on sound scientific data, that test scores received with an *appropriate* accommodation are fundamentally different than test scores received without an accommodation. No study to date has shown that a test result obtained from a test taken with an *appropriate* accommodation is incomparable to a test taken under standard conditions. Such a rule would force testing entities to conduct the detailed studies needed to demonstrate the difference they perceive exist and force them to more carefully evaluate the type of accommodation provided. This, in turn, would force testing entities to comply with the professional standards that they currently misinterpret to support the practice of flagging.

B. Require Testing Entities to Adhere to Their Own Agreements

If testing entities are going to make the final decision regarding a requested accommodation, they should not be allowed to waiver in their decision by flagging test scores received with that accommodation. This practice is inconsistent and unfair and represents a complete misunderstanding of the function of an accommodation.

Testing entities are required to take those steps necessary to ensure that a particular accommodation compensates an individual for their unique disability, without providing them with an actual advantage over other non-disabled

individuals. Notwithstanding the Standard's admonition against the use of speeded tests, professional entrance exams continue to be flagged, such as the LSAT and MCAT. Speeded tests are inherently discriminatory because they are based on what the average, non-disabled individual would be able to do within a certain period of time. However, such tests are discriminatory in practice only if reasonable accommodations are not provided for disabled individuals who require additional time to complete the exam. Although extended time accommodations may change the construct measured in speeded tests, the added time alters the construct measured only if it places a student at a competitive advantage by providing more time for analysis than needed to compensate for the individual's disability.

Studies show that testing entities typically provide some multiple of additional time to complete the exam, e.g., time and one-half, double time, without any evidence to show that the specific amount of time provided is appropriate. This practice violates Standards 10.6 and 10.10, which collectively require test entities to provide a specific amount of time, based on empirical evidence that is appropriate to address that student's unique disability. The failure to address this problem is exemplified in *Doe*, where the testing entity gave more time than Doe requested then flagged his test score.[258]

With speeded tests, the differences observed in scores received under standard and nonstandard time conditions do not lie in the provision of additional time. Rather, the studies show that the difference results from the provision of too much time. The problem, therefore, lies in the failure of testing entities to accurately quantify the amount of time each individual needs to place them on a level playing field with the average test taker. Learning disabled students, for example, require more time than non-disabled students to answer the same number of questions.

The amount of time needed to adequately compensate for a learning disability is likely different for each individual, influenced in part by the severity of the disability. If too little time is provided, the individual is under-compensated and placed at a competitive disadvantage. This may occur when a learning disabled student is required to take a test under standard conditions that has been standardized for the average population of non-disabled test takers. Conversely, if too much time is provided, the disabled test taker is overcompensated and placed at a competitive advantage. Such under-compensation or overcompensation may result in deflated or inflated scores that do not accurately reflect the student's actual ability. The key then is to find the amount of time that is just right to properly address the student's disability. Unlike Goldilocks, who kept sampling from bowls until she found the porridge that was "just right,"[259] testing entities have been lax in evaluating new strategies to reach a solution that satisfies their dual obligations.

Testing entities should be required to accept the recommendation for accommodation provided by a student's treating physician, unless there are justifiable reasons for not doing so. Once the decision is made to grant an accommodation, there should be no further consideration of the effect of the accommodation because, by granting the accommodation, the testing entity is

---

[258]Doe v. Nat'l Bd. of Med. Exam'rs, No. CIV. A. 99-4532, 1999 WL 997141, at *2 (E.D.Pa. Nov. 1, 1999), *vacated*, 199 F.3d 146 (3d Cir. 1999).

[259]DLTK's Sites, The Story of Goldilocks and the Three Bears, http://www.dltk-kids.com/rhymes/goldilocks_story.htm (last visited Mar. 1, 2007).

effectively agreeing that the accommodation is appropriate. An appropriate accommodation does no more than eliminate artificial disability-related barriers to place all test takers on an equal level.

If a testing entity believes an accommodation is inappropriate, it should be required to submit the medical documentation provided by the student to a neutral third party with qualifications comparable to the medical professional recommending the accommodation. This would ensure that the decision to grant or deny an accommodation is based on sound medical understanding of the student's disability and is not made by an agent of the testing entity who lacks the medical understanding to make such a decision. If no form of accommodation will place the student on an equal level with other students taking the same test, then a different form of assessment should be utilized. Such an approach comports with the existing disability laws and professional standards.

C. Reevaluate OCR's Interim Policy

The Office of Civil Rights (OCR) should be required to reevaluate its interim policy on flagging in view of existing empirical evidence on the practice. The existing policy appears to be based on an assumption that flagging is not discriminatory because the flag reflects a real difference in scores obtained under standard and nonstandard conditions. Because the data does not support such an assumption, OCR should be inclined to reverse the policy and prohibit flagging until testing entities offer proof that there is a measurable difference in test scores obtained under standard and nonstandard conditions.

OCR would have considerable difficulty justifying continuing the policy in view of its own finding that schools have misused the flag to discriminate against applicants with disabilities. Given the long history of prejudice exhibited against individuals with disabilities, one must assume at least some admissions personnel will utilize knowledge of a person's disability in an inappropriate manner. A flagged score effectively indicates to admissions personnel that the examinee has a disability of some sort, because only disabled people receive testing accommodation. The annotation often provides no information regarding the type or severity of the disability. Due to the inherent secrecy surrounding admissions decisions, the only way to avoid the possibility for bias in the admissions process is to ensure that information regarding an individual's disability is withheld. The best way to do this is to require the OCR to change its policy to acknowledge that flagging violates the prohibition against preadmission inquiry into an applicant's disability.

D. Eliminate the Element of Speed from Standardized Tests

Notwithstanding the Standard's admonition against the use of speeded tests, parts of the LSAT, MCAT and other Standardized tests are speeded. Speeded tests are inherently discriminatory because they are based on what the average, non-disabled individual would be able to do within a certain period of time.[260] However, such

---

[260]NAT'L CTR. FOR LEARNING DISABILITIES, *supra* note 99 (noting that students with disabilities are usually not included in the sample population used in test development nor are students with disabilities, when included, given appropriate accommodations, which results in a lack of test validity). Given that the LSAT was formulated in 1974, well before the ADA, it is highly unlikely that the developers considered how long it would take students with particular disabilities to complete a set number of questions.

tests are only discriminatory in practice if reasonable accommodations are not given to disabled individuals who require additional time to complete the exam. That is, if the time allowed is inadequate to fully compensate for a student's disability, the test is unfair and places non-disabled students at an advantage. Conversely, if the amount of time provided overcompensates the individual for the disability, the test is unfair because it places the disabled individual at a competitive advantage. The problem, therefore, lies in the failure to accurately quantify the amount of time each individual needs to place them on a level playing field with the average test taker. Because learning disabled students require additional time to complete the same number of questions, the goal of providing an accommodation should be to identify that quantum of time that accurately compensates the individual for his or her particular disability. Testing entities should be required to perform the tests needed to identify the amount of time appropriate for a particular disability, or eliminate the element of speed from tests.

Although extended time accommodations may change the construct measured in speeded tests,[261] the added time alters the construct measured only if it places a student at a competitive advantage by providing more time for analysis than needed to compensate for the individual's disability. For tests such as the LSAT and MCAT, speed is not necessary. Although it is true that in some cases law students and medical students are required to think and react quickly, the practice of law or medicine should not be restricted from those individuals unable to think or react at the same rate as non-disabled students. For example, for transactional lawyers the speed at which information is processed or assignments completed may be less important than it would be for attorneys who are required to make split second decisions during a trial. Similarly, some medical students will conduct research in lieu of serving in the emergency room. Using speeded tests that may have the effect of eliminating from consideration otherwise qualified students that have the potential to become successful professionals is inappropriate.

E. Amend the Standards to Comport with Existing Disability Laws

Notwithstanding the Standards acknowledgement that the practice of flagging "may conflict with legal and social policy goals promoting fairness in the treatment of individuals with disabilities," testing entities that rely on the Standards for guidance continue to flag certain test scores. [262]

The Standards are problematic for several reasons. They encourage the practice of flagging where there is no evidence that scores obtained from tests taken under nonstandard conditions are incomparable to those obtained from standard administrations. Aside from any discriminatory impact that may result through application of the Standards, the Standards also discourage testing entities from performing the detailed, and likely costly, studies that would be required to accurately demonstrate differences in test results. Under the ADA, testing entities are required to use tests that fairly and equally measure the actual abilities of the disabled, regardless of the burden imposed.[263] Providing a student with an

---

[261]Sireci, *supra* note 70, at 10.

[262]*See* AM. EDUC. RESEARCH ASS'N ET AL., *supra* note 70, at 108.

[263]*See generally* 28 C.F.R. § 36.309(b)(1)(i) (2006).

accommodation and then flagging the student's test scores runs afoul of the ADA's mandate.

The Standards are also problematic because they encourage testing entities to provide accommodations but fail to address the misuse of those accommodations by testing entities. In theory, an accommodation compensates a student for his or her disability and places that student on the same level as others taking the same test.[264] Properly used, accommodations eliminate testing barriers without providing the disabled student with a competitive advantage over other non-disabled test takers. Testing entities make the final determination regarding whether a requested accommodation will be granted. When a testing entity demands detailed and costly documentation of a disability; and then, based on that documentation, grants an accommodation, the entity is in effect agreeing that the accommodation adequately compensates the student for his or her disability. Thus, for example, by agreeing to provide a student with double time to take an examination, the entity acknowledges that that student requires that precise amount of additional time to be placed on the same level with other non-disabled students taking the same test. It is inconsistent to provide such an accommodation and then flag the test score received based on a perception that the accommodation somehow rendered the test score inaccurate. The practice is particularly egregious when the person making the decision to grant or deny the requested accommodation has little or no medical expertise in diagnosing or evaluating disabilities.[265] Yet, this is exactly what continues to occur. Disabled students are forced to expend considerable sums of money to obtain diagnoses and recommendations for accommodations from medical professionals. Testing entities evaluate the information, grant an accommodation deemed appropriate, and subsequently penalize the same students for receiving the accommodation.

The Standards have the effect of sanctioning discrimination by encouraging testing entities to utilize stigmatizing flags to address problems created by the entities' own failure to provide disabled students with equally effective means of demonstrating their abilities. An accommodation is either appropriate or it is not. Under the current practice, accommodations are provided without any empirical evidence of whether the accommodation under or overcompensates the student, or places the student on the same level as other students.

---

[264]*See* Daniel B. Tukel & Katherine Donohue Goudie, *Accommodation in Testing: Is a "Level Playing Field" Unfair?*, MICH. B. J., Aug. 2006, at 28, 31, *available at* http://www.michbar.org/journal/pdf/pdf4article1034.pdf.

[265]For example, in *Badgley v. Law School Admission Council, Inc.*, No. CIV.A. 4:99CV-0103-M, 2000 WL 33225418 (N.D.Tex. Aug. 24, 2000), Badgley took the LSAT without an accommodation and then again with an accommodation. *Id.* at *2-*3. He requested that LSAC report the test score he took with an accommodation, arguing that LSAC improperly refused his request for accommodation on the first test. *Id.* at *1. The facts showed that Kim Dempsey, Disability Specialist/Manager of Accommodated Testing for the LSAC, personally made the determination to refuse Badgley's first request for accommodation. *Id.* at *2. The court ordered LSAC to report Badgley's accommodated test score after finding, *inter alia*, that "Dempsey does not have any medical expertise in the fields of tremors or visual impairments. . . . Dempsey does not have sufficient medical knowledge to justify a refusal to grant reasonable accommodations which were recommended by Badgley's treating physicians and on LSAC's specialist forms." *Id.*

Additionally, the Standards improperly encourage testing entities to disclose a student's disability to those who would otherwise have no right to obtain the same information before preadmission. Disclosing the nature of the accommodation or simply annotating the test result with a "flag" circumvents the prohibition against preadmission inquiry and subjects disabled applicants to the same potential for discrimination that the rule was promulgated to eliminate.[266] Singling out students with disabilities in the admissions process may deny some the equal opportunity to education required under law. Moreover, the practice of flagging may have the unintended and equally damaging effect of making some forego requesting an accommodation to avoid publicly disclosing their disability. While maintaining test integrity is important, that goal does not justify subjecting some students to the potential for preadmission discrimination, particularly in the absence of empirical evidence demonstrating a difference in test scores obtained under standard and nonstandard conditions.

For all of the preceding reasons, the Standards should be amended to urge testing entities to avoid using flags unless scientifically acceptable evidence shows that scores obtained with an accommodation are not accurate predictors of a student's likelihood of future success. Testing entities should not be allowed to disregard federal law designed to protect the rights of disabled individuals, particularly where the testing entities have failed to present sufficient justification for infringing on those rights.

VII. CONCLUSION

The practice of flagging provides strong evidence that testing entities have little interest in ensuring that the tests they administer measure the abilities of disabled and non-disabled students equally. To answer the question of whether scores achieved on standardized tests taken under standard and nonstandard conditions are comparable, further research is required. Because flagged tests scores have the effect of informing individuals who make admissions decisions that an applicant is disabled, the practice of flagging should be prohibited. Students with disabilities should not be subjected to the potential for discrimination in the absence of proof that their test scores do not accurately predict their likelihood for future success. Given the increasing number of disabled students attending college, placing the burden on testing entities to develop fair and accurate test measurements and procedures that are comparable for all students is both reasonable and required.

[266] A preadmissions inquiry may only be made if the school is: (1) "taking remedial action to correct the effects of past discrimination" or (2) "taking voluntary action to overcome the effects of conditions that resulted in limited participation" in the past. *See* 45 C.F.R. § 84.42(c) (2006).