

FASP

Position Paper on the Use of the Florida Comprehensive Assessment Test (FCAT) In High Stakes Decision Making

The Florida Association of School Psychologists (FASP) is a professional organization representing approximately 1200 school psychologists in the State of Florida. The mission of FASP is to promote and advocate for the mental health and educational development of Florida's children and youth; to support families, educational systems and communities; and to advance the interests of the profession of school psychology. FASP supports public policies that are educationally equitable, empirically sound, and are likely to increase positive educational outcomes.

Through their training and experience in education and psychology, school psychologists are in a unique position to share their expertise in educational measurement, variables crucial for students' academic success, and the impact of high stakes testing decisions on educational systems and student outcomes. It is, therefore, the purpose of this paper to provide educational stakeholders in Florida (school administrators, legislators, policy makers, parents, teachers, and students) with an understanding of both the positive and negative implications of the current student accountability system and to offer appropriate and educationally sound recommendations for change.

History of the Florida Comprehensive Assessment Test (FCAT) in Florida

The evolution of statewide assessment in Florida began in 1972 with an assessment program that measured students' acquisition of minimum competency skills. In October 1977, the state implemented, upon legislative authorization, the nation's first required high school graduation test. This controversial exit exam became the foundation of a landmark federal case known as *Debra P. v. Turlington*. The state's position, which was upheld in federal court, was that such tests can improve students' educational performance, identify students who need remedial assistance, and evaluate the attainment of state educational objectives (Debra P. v. Turlington, 1979). Students in the graduating classes of 1983 to present have been required to pass the state mandated competency test in order to obtain their high school diploma (FCAT Briefing Book, 2001).

In 1995, the Florida Commission on Education Reform and Accountability recommended procedures for assessing student learning in Florida with the goal of raising educational expectations and ensuring that students had the necessary skills to compete for jobs in the global marketplace. Upon adopting the recommendation for assessment (called the Comprehensive Assessment Design), the State Board of Education specified for the development of new statewide assessments to address broad standards of Goal 3 (i.e. Student Performance) of Blueprint 2000. Subsequently, the Florida curriculum standards, referred to as the Sunshine State Standards, were developed and adopted by the State Board of Education. The FCAT was designed to meet the

requirements of the Comprehensive Assessment Design and the content of the Sunshine State Standards (FCAT Briefing Book, 2001).

Governor Bush's A+ plan was adopted with legislative approval in 1999. The plan called for the expansion of the FCAT to include Grades 3 through 10. In addition, results of student performance on the FCAT were utilized in assigning school grades beginning in 1999. In August of 2001, the State Board of Education established the FCAT passing scores that students must earn as one of the requirements for receiving a regular high school diploma. Students anticipating graduation in 2003 will be required to pass the FCAT (FCAT Briefing Book, 2001).

At the federal level, the 2002 reauthorization of the Elementary and Secondary Education Act (ESEA), the *No Child Left Behind Act*, requires annual assessment of students in grades 3 through 8. It further requires states and schools to meet "adequate yearly progress" by increasing test scores (NASP, 2002).

Positive Aspects of Testing Programs

When used properly, testing programs such as the FCAT can be a means of fostering growth toward high expectations and improving the links among curricular areas. When provided with meaningful information from other sources, such as student grades and teacher evaluations, such testing programs can provide important insights into a student's knowledge and abilities. Test results can also provide teachers with feedback on the effectiveness of the curriculum and their teaching methods. On a system level, the use of tests can help identify those schools in need of support and interventions to improve educational outcomes.

Appropriate Test Use in Education

When used properly, tests are among the soundest and most objective performance measures available. However, tests are frequently used inappropriately. A common violation of ethical testing practices is to use a test for a purpose for which it was not developed. For example, if the test's scores were *only* validated as an accurate indication of individual students' strengths and weaknesses, it should not also be used to evaluate the educational quality of a school (American Psychological Association, 2001).

The *Standards for Educational and Psychological Testing*, created by the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education, presents principles designed to foster fairness in testing and to prevent unintended consequences. These principles include:

- When educational testing programs are mandated by school, district, state, or other authorities, the ways in which test results are intended to be used should be clearly described. It is the responsibility of those who mandate the use of tests to monitor their impact and to identify and minimize potential negative consequences. Consequences resulting from the uses of the test, both intended and unintended, should also be examined by the test user.

- In educational settings, when a test is designed or used to serve multiple purposes, evidence of the test’s technical quality should be provided for each purpose.
- When test results substantially contribute to making decisions about student promotion or graduation, there should be evidence that the test adequately covers only the content and skills that students have had an opportunity to learn.
- In educational settings, a decision or characterization that will have major impact on a student should not be made on the basis of a single test score. Other relevant information should be taken into account if it will enhance the overall validity of the decision.
- When test scores are intended to be used as part of the process for making decisions for educational placement, promotion, or implementation of prescribed educational plans, empirical evidence documenting the relationship among particular test scores, the instructional programs, and desired student outcomes should be provided. When adequate empirical evidence is not available, users should be cautioned to weigh the test results accordingly in light of other relevant information about the student.
- For tests that will determine a student’s eligibility for promotion to the next grade or high school graduation, students should be granted, if needed, multiple opportunities to demonstrate mastery of materials through equivalent testing procedures.
- In educational settings, score reports should be accompanied by a clear statement of the degree of measurement error associated with each score or classification level and information on how to interpret the scores.
- In testing individuals with disabilities, test developers, test administrators, and test users should take steps to ensure that the test score inferences accurately reflect the intended construct rather than any disabilities and their associated characteristics extraneous to the intent of the measurement.
- If a test is mandated for persons of a given age or all students in a particular grade, users should identify individuals whose disabilities or linguistic backgrounds indicate the need for special accommodations in test administration and ensure that these accommodations are employed.

The *Code of Fair Testing Practices in Education* was developed by the Joint Committee on Testing Practices in order to “represent the spirit of a selected portion of the *Standards* in a way that is meaningful to test takers and/or their parents or guardians.” The following is offered as further support of the *Standards* when utilizing educational tests in high stakes decision-making:

- *Test developers should:* Provide information that will help users follow reasonable procedures for setting passing scores when it is appropriate to use such scores with the test.
- *Test users should:* Explain how any passing scores were set and gather evidence to support the appropriateness of the scores.

High Stakes Testing

Increasingly, states are mandating the use of testing programs to monitor student achievement and to hold schools and students accountable. These tests are referred to as “high stakes” when the scores from a single test carry heavy consequences for school systems or for students. As currently implemented, the FCAT is being utilized as a high stakes test. As an illustration, in Florida, schools whose students perform well on the FCAT are financially rewarded. Conversely, schools whose students perform poorly on the FCAT are penalized. Research has indicated a correlation between low performing schools and students from a low socio-economic background (Fischer & Dougherty, 1999; Dorman, 2001). Therefore, as a result of current legislative policy, those schools and students *most* in need of financial resources are those that are deprived of much needed funds for educational improvement while their high performing counterparts are receiving financial incentives for their superior performance.

Of further concern is the impact of high stakes testing on individual student tracking and retention. Current policy mandates that third grade students who score at Level 1 on the FCAT reading section be retained, unless there is evidence suggesting exemption from mandatory retention for good cause. Grade retention based on a single measure of performance such as the FCAT is not a practice supported by sound educational research. In fact, grade retention based on any method of evaluation is contraindicated (NASP, 1998).

Drop-out rates and high school completion rates are vital indicators of the health of any educational system. The 1996 *Comprehensive Biennial Report on Texas Public Schools: A Report to the 75th Texas Legislature* indicated that being over-age for grade is a better predictor of dropping-out than are below average test scores (Clarke, Haney, Madaus, Lynch & Lynch, 2000)

G. Orfield, Co-Director of the Civil Rights Project at Harvard, indicates that high-stakes tests penalizes low income and ethnic minority students and is linked to high drop-out rates among these groups. Furthermore, Orfield indicates that African-Americans and Hispanics are three to four times as likely to be retained as whites. He warns that high-stakes tests are not standards, but rather can be “the punishment of innocent victims of unequal education” (The Impact of High Stakes Testing Policies on Minority and Disadvantaged Students, 2000).

The Ineffectiveness of Retention

One of the most controversial policies of the FCAT is its use in determining which children will be retained at a grade level. The rationale for this policy is purportedly to put an end to the practice of “social promotion.” Empirical studies

demonstrate the long- term negative effects of retention and further indicate that social promotion, sans remediation, does nothing to promote learning. The National Association of School Psychologists offers an empirically based *Position Statement on Student Grade Retention and Social Promotion*, which includes the following research regarding the ineffectiveness of retention:

In elementary school:

- Some groups of children are more likely to be retained than others. Those at highest risk for retention are male, Black or Hispanic; have a late birthday, delayed development and/or attention problems; live in poverty, in a single-parent household or have parents with low educational attainment; or have changed schools frequently.
- While delayed entry and readiness classes may not hurt children in the short run, there is no evidence of a positive effect on either school achievement or adjustment. Furthermore, by adolescence, these early retention practices are associated with numerous health and emotional risk factors.
- Retention is generally associated with poorer academic achievement when groups of retained children are compared to groups of similar children who are promoted. The most notable deficit for retained students is in reading, the primary academic problem for which students are retained.
- Initial achievement gains may occur during the retention year, but the consistent trend across many research studies is that achievement gains decline within two to three years of retention such that retained children either do no better or perform more poorly than similar groups of promoted children. This is true whether children are compared to same-age or same-grade students who were promoted.
- Retention appears to have no significant impact on overall school adjustment as measured by self-esteem inventories; however, retention is associated with significant increases in behavior problems as measured by behavior rating scales, with problems becoming more pronounced as the child reaches adolescence.

At the secondary level:

- Students who were retained or had delayed kindergarten entry are more likely to drop out of school compared to students who were never retained, even when controlling for achievement levels. The probability of dropping out increases with multiple retentions. Even for single retentions, the most consistent finding from decades of retention research is the high correlation between retention and dropping out.

- Retained students have increased risk of health-compromising behaviors such as emotional distress, cigarette use, alcohol use, drug abuse, driving while drinking, use of alcohol during sexual activity, early onset of sexual activity, suicidal intentions, and violent behaviors. Furthermore, students who themselves were not retained but who attend schools with a high proportion of old-for-grade students are also at risk for increased substance abuse.

In adulthood:

- Adults who were grade repeaters are more likely to be unemployed, living on public assistance or in prison than adults who did not repeat a grade.

The Cost of Retention

In public policy decisions, one should also consider the fiscal impact of a practice that is demonstrably ineffective in remediating educational shortcomings. During the administration of the year 2000 FCAT, 65,000 fourth graders in Florida scored at a Level 1. Retaining those students as mandated by suggested legislative policy would cost taxpayers approximately \$200 million. This figure does not include capital outlay costs for housing these students for an additional year (Midgett, 2002). With the abundance of research indicating the ineffectiveness of retention coupled with the financial crisis facing Florida's schools, policy makers should be encouraged to promote and financially support empirically based interventions and programs for Florida's students.

Accountability for Students with Disabilities

Florida's students who are identified as having a disability under federal and state guidelines are not held to the same accountability standards in regards to FCAT testing as their non-disabled peers. As currently implemented, the A+ plan punishes schools for having low performing students but provides an avenue for removing these students from the school's accountability numbers through Exceptional Student Education (ESE). While ESE has a strong accountability system, the accountability system is based on monitoring for compliance (i.e. signatures, notification of parental rights, paperwork) as opposed to accountability via academic progress. The report of the *President's Commission on Special Education* (2002) illustrates this point, indicating that "the current system often places process above results, and bureaucratic compliance above student achievement, excellence, and outcomes."

The Individuals with Disabilities Education Act (IDEA) mandates that students with disabilities be included in school-based testing (IDEA, 1997). Research conducted by The National Center for Educational Outcomes indicates that there are a number of positive consequences of including students with disabilities within large scale assessments (when large scale assessments are used appropriately). These positive outcomes have included increased levels of student performance, higher expectations for student achievement, increased access to the general education curriculum, and improved teaching and instruction (NASP, 2002). However, since the performance of ESE students is not included in the school grades as generated by the FCAT, it is highly reinforcing to

identify low performing students as having a disability so that these students do not count “against” the school when the school is assigned a grade.

Test Scores and Socio-economic Issues

Data gleaned from the *Kids Count Data Book* (Annie E. Casey Foundation, 2001) indicate that approximately 1 out of every 5 children in Florida lives in poverty. Poor families do not have the resources to pay for quality childcare, stimulating toys and books, and other factors linked to enhancing educational readiness during early childhood. Research has indicated that children from financially advanced homes arrive at first grade with their verbal and math skills at a higher level than do children from financially disadvantaged homes. As a result, the strong relationship between family economic backgrounds and children’s school outcomes begins early (Park, Turnbull & Turnbull, 2002).

Research has consistently revealed the impact of poverty on poor educational performance (Park, et al., 2002). While not all children living in poverty perform poorly academically, poverty appears to be one of the variables most predictive of low academic achievement. As a further illustration, a recent research study conducted in a large urban school district has yielded statistically significant relationships between the percent of students on Free/Reduced Lunch (an indicator of low socio-economic status) and fourth and eighth grade reading and math scores on the FCAT (Dorman, 2001). The impact of low socio-economic status and other relevant variables on student achievement and FCAT scores cannot be ignored.

CONCLUSIONS AND RECOMMENDATIONS:

FASP supports public policies that are educationally equitable, empirically sound, and are likely to increase positive educational outcomes. Therefore, the Florida Association of School Psychologists recommends the following:

With regard to sound testing practices:

1. Multiple sources of assessment information should be used when making high stakes decisions. No single test should be used alone for making decisions about the tracking, promotion, or graduation of individual children.
2. There should be a clear validation for each separate intended use of the FCAT. If the FCAT is only validated for individual progress monitoring, there must be separate validation for using scores to monitor school progress.
3. The contracted testing company has an ethical responsibility to disclose fully the likely negative consequences of high stakes programs.
4. There should be an ongoing monitoring of alignment between the test and the curriculum.
5. An ongoing evaluation of intended and unintended effects of high stakes tests, of which the FCAT is one example, is recommended.
6. The standard error of measurement of the FCAT for different score levels should be available and taken into consideration when making high stakes decisions about students.

7. The *Standards for Psychological and Educational Testing* should be employed as a best practice model in test design, administration, and ethical interpretation.

With regard to educational equity for all students:

8. FASP supports the provision of adequate resources and opportunities to learn for all students.
9. FASP supports practices that are linked to positive educational outcomes, such as providing resources to teachers to improve reading levels of poor readers and to strengthen the assessment and intervention roles of Student Services personnel.

With regard to the use of the FCAT as a component of the A+ plan for rewarding schools:

10. FASP does not support punitive measures for low performing schools through diversion of needed funds to reward high performing schools.

With regard to using the FCAT to determine retention of students:

11. FASP supports opportunities for meaningful remediation for examinees that fail the FCAT. Retention is **NOT** considered a meaningful remediation. Retention is leaving children behind.

12. FASP, in conjunction with the National Association of School Psychologists, supports effective alternatives to retention and social promotion including:

- Adopting age-appropriate and culturally sensitive instructional strategies that accelerate progress in all classrooms
- Providing effective early reading programs such as Success for All, Reading Recovery and Direct Instruction
- Offering extended year and extended day programs
- Implementing tutoring programs

With regard to the use of the FCAT for students with disabilities:

13. FASP supports the inclusion of ESE students in school grading and evaluation.
14. FASP encourages advocating for ESE monitoring of student progress, not clerical compliance.

With regard to academic monitoring:

15. FASP supports a continuous progress monitoring of student achievement utilizing empirically sound assessment strategies such as curriculum-based measurement.

With regard to future research on the FCAT:

16. FASP encourages research by state and local district officials to examine the link between FCAT scores and poverty as well as other relevant factors that may contribute to underachievement. Results of such research should be shared with pertinent stakeholders so that all of the causes of underachievement are appropriately addressed.

References

- American Psychological Association (1999). *Standards for educational and psychological testing*. Washington, DC: Author
- American Psychological Association (2001). *Appropriate use of high-stakes testing in our nation's school*. Washington, D.C.: Author. Retrieved April 18,2002, from www.apa.org/pubinfo/testing
- American Youth Policy Forum. *The impact of high stakes testing policies on minority and disadvantaged students*. Capital Hill: A Forum Brief, January 7, 2000. Retrieved March 11, 2002, from <http://www.aypf.org/forumbriefs/2000/f6010700.htm>
- Annie E. Casey Foundation. (2001). *Kids count data book: State profiles of child well-being*. Baltimore, MD: Author.
- Clarke, M., Haney, W., Madaus, G., Lynch, P., Lynch, C. (2000). High stakes testing and high school completion. *The National Board on Educational Testing and Public Policy*,1,1-14. Retrieved March 19,2002, from www.nbetpp.bc.edu
- Code of fair testing practices in education*. (1988). Washington, D.C.: Joint Committee on Testing Practices.
- Dorman, C. (2001, November). *Everything you always wanted to know about the FCAT and some things you didn't*. Paper presented at the Annual Conference of the Florida Association of School Psychologists, Orlando, FL.
- Fischer, K. & Dougherty, G. (1999, March 21). A lesson in grading schools. *St. Petersburg Times*.
- Florida Department of Education (2001). *FCAT briefing book*. Retrieved May 7, 2002, from <http://www.firn.edu/doe/sas/>
- Heubert, J.P. & Hauser, R.M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington: National Academy Press.
- Midgett, J. (2002, February 20). Political arrogance in the A+ plan. *The Daytona News Journal*, p. A5.
- National Association of School Psychologists (NASP, 1998). *Position statement: Student grade retention and social promotion*. Bethesda, MD: Author. Retrieved March 19,2002, from www.nasponline.org
- National Association of School Psychologists (NASP, 2002). *Large scale assessments and high stakes decisions: Facts, cautions and guidelines*. Bethesda, MD: Author. Retrieved June 17, 2002, from www.nasponline.org

Park, J., Turnbull, A., Turnbull, H. (2002). Impacts of poverty on quality of life in families of children with disabilities. *Exceptional Children*, 68, 151-170.

President's Commission on Excellence in Special Education (2002). *A new era: Revitalizing special education for children and their families*. Washington, D.C.: Author. Retrieved July 15, 2002, from www.ed.gov/inits/commissionboards/whspecialeducation/

Texas Education Agency (1996). *Comprehensive biennial report on Texas public schools: A report to the 75th Texas legislature*. Austin: Author.

U. S. Department of Education (1997). *Nineteenth annual report to Congress on the implementation of the Individuals with Disabilities Education Act*. Washington, D.C.: Author.

Legal References

Debra P. v. Turlington, 474 F. Supp. 244 (M.D.Fla. 1979); *aff'd in part and rec'd in part*, 644 F.2d 397 (5th Cir. 1981); *rem'd*, 564 F. Supp. 177 (M.D. Fla. 1983); *aff'd*, 730 F.2d 1405 (11th Cir. 1984).