

1-1-2006

# The Utilization of Gender, Retention, SES and STEEP Scores to Predict Reading Mastery

Lisa Lee Adams  
adams66@marshall.edu

Follow this and additional works at: <http://mds.marshall.edu/etd>

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Elementary and Middle and Secondary Education Administration Commons](#), and the [School Psychology Commons](#)

---

## Recommended Citation

Adams, Lisa Lee, "The Utilization of Gender, Retention, SES and STEEP Scores to Predict Reading Mastery" (2006). *Theses, Dissertations and Capstones*. Paper 433.

This Thesis is brought to you for free and open access by Marshall Digital Scholar. It has been accepted for inclusion in Theses, Dissertations and Capstones by an authorized administrator of Marshall Digital Scholar. For more information, please contact [zhangj@marshall.edu](mailto:zhangj@marshall.edu).

Running head: UTILIZING GENDER, RETENTION, SES & STEEP

The Utilization of Gender, Retention, SES and STEEP Scores to Predict Reading Mastery

Thesis Submitted to  
The Graduate College of  
Marshall University

In partial fulfillment of the  
Requirements for the degree of  
Education Specialist in  
School Psychology

by

Lisa Lee Adams

Stephen O'Keefe, Ph.D. Committee Chairperson  
Fred Jay Krieg, Ph.D.  
Sandra Stroebel, Ph.D.

Marshall University Graduate College

May 2006

## Abstract

The purpose of this study was to examine the relationship between gender, retention, SES and STEEP scores to predict reading mastery. Second grade students' records were reviewed to obtain demographic information. SAT 9 scores were used to measure students' reading mastery. SES was measured and defined by approved applications for free or reduced lunch for the 2004-2005 school year. Retention was defined as the number of students previously retained. Logistic regression was used to determine the predictive validity of gender, retention, SES and STEEP scores dependent on SAT 9 reading mastery. Gender, retention, SES and STEEP scores were entered as covariates in SPSS and analyzed for significance in a regression model. Results revealed the model was not as efficient as the use of STEEP by itself. STEEP predicted correctly 90 percent of students who achieved reading mastery on the SAT 9. The model was less accurate in predicting no or partial reading mastery.

### Acknowledgements

I would like to acknowledge Dr. O'Keefe, Dr. Krieg and Dr. Stroebel for their knowledge, guidance and assistance throughout my thesis project. I would also like to thank Thom, my practicum mentor and friend. Thom provided me with data to make this study possible. Finally, I would like to thank all of the people in my life who have supported me throughout my graduate school career. I am dedicating this paper to my family; Benny, Cheryl and Jason Adams; and to my boyfriend, Vince. Without your unconditional love, constant encouragement and support, I would not be where I am today. Thank you for believing in me. I love each of you very much!

## Table of Contents

Acknowledgements.....	iii
Literature Review.....	1
Method.....	9
Participants.....	9
Instruments.....	10
Procedures.....	12
Results.....	13
Discussion.....	14
Limitations.....	15
References.....	17

## The Utilization of Gender, Retention, SES and STEEP Scores to Predict Reading Mastery

A major achievement gap exists in America's school systems, (USDE, n.d.) and, as a result, federal law is enforcing accountability standards. Accountability was implemented by the government to aid in addressing these achievement gaps in the nation and to provide students with the greatest teachers to deliver a valuable education (USDE, n.d.). In 1965, the Elementary Secondary Education Act (ESEA) was enacted to finance and promote public education (USDE, n.d.). In attempts to achieve these goals, the Improving America's Schools Act of 1994 (IASA) required states to develop assessments matching state standards in reading and math (Goertz & Duffy, 2003). These assessments were used to monitor student performance and identify low performing schools.

The most recent amendment, The No Child Left Behind Act of 2001 (NCLB), has created an uproar in many school systems across the country with the two major components being; to set standards for grade level achievement; and to develop a system to measure progress of all groups and subgroups of students (USDE, n.d.). The goal of NCLB is to improve the quality of education for all students, including students with disabilities, by holding all students and teachers accountable (Albrecht & Joles, 2003). As stated in Moe (2001), given that students are held accountable for learning certain material, it is logical to hold those responsible for instructing them accountable as well. NCLB requires schools to test each child, grades three through eight in reading and math by the 2005-2006 school year (Goertz & Duffy, 2003). In accordance with NCLB, state accountability systems were put into effect to focus on student achievement and

continuous progress (Goertz & Duffy, 2003). The majority of states have adopted high-stakes assessments, tests that determine opportunities/outcomes for students (Albrecht & Joles, 2003) to demonstrate accountability. The utilization of achievement tests also known as high-stakes tests, is to measure student performance, how well schools are preparing students for subsequent education and their future (USDE, n.d.).

### *Importance of Literacy*

Research has shown that literacy is a vital skill to becoming a successful student as it impacts achievement in all other subjects. For many children, learning to read is challenging and requires years of effort (Lyon & Chhabra, 2004). President Bush has declared that every child shall read at grade level by the third grade (USDE, n.d.). As a result, educators all across the country are striving to guarantee their students achieve reading success as mandated by NCLB. As a society, we continue to advance by evolving the definition of literacy to encompass more than reading alone. Literacy has come to include, "an individual's ability to read, write, speak in English, compute and solve problems at levels of proficiency necessary to function on the job, in the family of the individual and in society" (NIFL). Children who begin pre-reading skills early are at a greater advantage to succeed at reading, as it is never too early to begin teaching children pre-reading skills (USDE, n.d.). Currently, as reported by the National Assessment of Educational Progress (NAEP), 68 percent of fourth graders are not reading at a proficient level (USDE, n.d.). For at least 70 percent of struggling readers, failure to read by age nine portends to a lifetime of illiteracy (Lyon & Chhabra, 2004).

*Socioeconomic Status (SES) and Literacy*

SES is a strong predictor of reading achievement. Research supports reading failure is most prevalent among children from disadvantaged environments (Lyon & Chhabra, 2004). Evidence suggests the lower the socioeconomic status, the less words and pre-reading skills children are exposed to; which affects vocabulary development (Torgesen, 2002). Reading is dependent on student's listening and speaking vocabularies (Garcia & Stafford, 2000). The more words a child hears at home, the more prepared he/she will be to learn to read as "young children who are not read to will, on average, enter school with a handicap" (Barton, 2004). There is a significant difference between exposure to the amount of words children living in professional homes, with parents who have degrees, than children living in homes where parents have limited education or lower socioeconomic status. Many children from lower SES backgrounds lack the vocabulary used in advanced textbooks (Hart & Risley, 2003)

A longitudinal study of spoken words over a one -year period of time revealed a significant difference between the amount of words children were exposed to and parental education (Hart & Risley, 2003). In the homes of professional families', 11 million words were reported; six million words were reported in working class families and three million words were reported for families receiving welfare. The number of words a good-reading middle school student reads is ten times the amount of words read by a poor-reading middle school student (Lyon, 2002).

Students have been shown to have better reading scores, as measured by Letter Word Identification on the WJ-R, if their mothers were more educated and extroverted,

and if their parents reported using more positive parenting practices (Burchinal et.al., 2002).

SES is one of the strongest predictors of composite scores on standardized tests (Baker, Spencer, and McGee, 2000). Eighth grade students were used to demonstrate the impact of academic achievement. Predictors such as SES were correlated with composite scores on The Iowa Test of Basic Skills (ITBS). In addition to SES, family poverty level in the community, educational level of the community, dropout rate, attendance and family median income were compared as indicators for academic achievement. Students' socioeconomic status, as measured by the percentage of students receiving free or reduced lunches, demonstrated the strongest relationship with the outcome variable of the ITBS composite score (Baker, Spencer, and McGee, 2000).

Lower income students' achievement is affected by many factors including: child rearing practices, health needs and mobility rates (Rothstein, 2004). Lower income students are more likely to receive more negative attention (scolding) than positive (encouragement for their actions), (Rothstein, 2004). During the first four years of life, a child in a professional family hears two encouragements to every discouragement, while the average child, living in a welfare family, hears one encouragement to every two discouragements (Hart & Risely, 2003). In sum after the first four years of life, the average child living in a welfare family is likely to have had 144,000 fewer encouragements and 84,000 more discouragements for his/her behavior than the average child from a middle class or upper-class family (Hart & Risley, 2003).

Lower income students are more likely to have vision problems as result from watching too much television too close to the screen than middle or upper-class students. In addition, lower income students are less likely to receive corrective lenses which can greatly affect school performance (Rothstein, 2004). Many lower income students live in housing developments which still have lead paint. As a result many lower income students have high traces of lead in their blood affecting their cognitive abilities and academic achievement (Rothstein, 2004). As a result of unemployment, lack of funds to pay rent, many low income families are forced to relocate more often than middle or upper-class families (Rothstein, 2004). Frequent moving also affects students' academic achievement. Rothstein (2004) reported poor students and students from single-parent homes move/change schools the most. Data from the US General Accounting Office showed that 41 percent of movers were below grade level in reading (Barton, 2004).

### *Retention*

Retention occurs when a student is held back, or forced to repeat a grade before permitted to proceed to the next grade. Theories supporting retention convey that repeating a grade will improve a student's school performance by providing more time for the student to develop adequate academic skills (Owings & Magliaro, 1998), and or the assumption that retention motivates students to try harder the second time around (Black, 2004).

Retention originated in the nineteenth century as a solution for students who were not prepared for material in the next grade (Rafoth, 2002). Even though negative effects of retention have been articulated since the 1930s, it is still a common practice revealing

a steady increase over the past 25 years (Rafoth, 2002). Research suggests retention is still on the rise (Owings & Magliaro, 1998), ignoring conflicting evidence that for most children, retention fails to improve achievement in reading and math (Black, 2004). Nearly 2.5 million students are retained each year at about \$6,000 cost per child (Rafoth, 2002). National evidence indicates 30 to 50 percent of students will be retained at least once before ninth grade (Jimerson, 2003). Few studies have been successful in supporting the benefits of retention (Rafoth, 2002; Jimerson, 2003).

Research reveals retained students typically come from lower SES backgrounds than non-retained students. Ten percent of retained students come from the lowest SES quartile while only 8.5 percent come from the highest SES quartile (Owings & Magliaro, 1998). About two-thirds of retentions occur between kindergarten and third grade (Owings & Magliaro, 1998). Retaining students in kindergarten and the early years as a preventive measure proves no better for students than retaining them in the upper grades (Black, 2004).

Majority of students are retained at the kindergarten and elementary level as a result of poor readiness skills and or poor reading skills (Rafoth, 2002). At the secondary level, students are often retained for either insufficient credits or failure to pass mandated high-stakes exams (Rafoth, 2002). Research on retained kindergarteners indicates that retained students have significantly lower scores on standardized achievement tests than do non-retained students (Owings & Magliaro, 1998). Retained students tend to decline in reading achievement over time compared to non-retained students (Owings & Magliaro, 1998). Rafoth makes the point that profiles of retained children reveal multiple

deficits and special needs, while making the assumption that repetition of the same curriculum for an extra year would be ineffective.

### *Gender Differences*

A gender gap between male and female school performance has existed in America's schools for sometime. Girls reportedly receive higher grades in reading and spelling as evidenced by girls' higher performance on achievement tests in reading (Pomerantz, 2002). Boys are reported to outperform girls in math and sciences, while girls outperform boys in reading and language arts (Gurian, 2004 & Pomerantz, 2002). One theory explaining these gender gaps is that of brain structure. This theory is supported by male/female brain structural differences detected by positron emission tomography (PET) and MRI technologies (Gurian, 2004).

Girls, on average, are able to multitask better than boys; girls have a larger corpus callosum, enabling more "cross talk" between the left and right hemispheres of their brain. Girls also tend to have stronger neural connectors in their temporal lobes providing them with better listening skills and more detail when completing writing assignments (Gurian, 2004). Finally, girls tend to use more cortical areas of their brain for more verbal functioning. These "girl" brain tendencies support why girls generally outperform boys in reading and writing (Gurian, 2004). Characteristics associated with learning to read such as memory, listening, sound discrimination, etc., are innately easier for girls due to their brain structure (Gurian, 2004).

According to the 2000 National Assessment of Educational Progress (NAEP), boys are one to one-and-one-half years behind girls in reading and writing (NAEP, 2000).

Boys tend to use more cortical areas of the brain for spatial-mechanical functioning and significantly less space is dedicated to verbal functioning (Gurian, 2004). Boys experience more attention problems than girls as a result of their innate desire to move objects through space due to their brain structure. Boys also operate with less blood flow to the brain; girls operate with about fifteen percent more blood flow (Gurian, 2004). Boys tend to earn lower grades, account for 60 percent of learning disability diagnoses, and are more likely to dropout of high school and less likely to attend college (Gurian, 2004).

### *High-Stakes Tests*

A high-stakes test is a test that is used to make decisions about student outcomes based on the student's performance on a specific test (Albrecht & Joles, 2003). Examples of high-stakes tests are the Stanford Achievement Test - Ninth Edition (SAT 9), the Iowa Test of Basic Skills (ITBS), and the Florida Comprehensive Assessment Test (FCAT). Schools throughout the United States are adopting these and other high-stakes tests for accountability purposes as mandated by NCLB (USDE, n.d.). Public reporting is the most common practice in ensuring accountability, currently 48 states use public reporting via "report cards" displaying schools' status, performance and progress (Abrams & Madaus, 2003). High-stakes tests are used to hold school districts and teachers accountable for the progress or lack of student progress (Goertz & Duffy, 2003). Common uses of high-stakes tests besides accountability include: promotion/retention, student placement, and high school graduation exit criteria (Goertz & Duffy, 2003). Currently Massachusetts, New York, Texas and Virginia require high school seniors to

pass a state-wide assessment before students are awarded a high school diploma (Abrams & Madaus, 2003).

### *Purpose of the Study*

The purpose of this study is to determine the strongest predictor of reading mastery by measuring total reading performance on the SAT 9 using SES, gender, retention and STEEP scores. The results of the study may provide educators reliable information they can use to intervene and identify students “at risk” for low reading performance.

### *Hypothesis*

It is hypothesized that reading mastery will be dependent on SES, gender, retention and STEEP scores, and can be used as a valid inventory of student’s reading skills when compared to their reading mastery outcomes measured by the SAT 9.

### Operational Definitions

The following terms were designed for use in this study:

Reading mastery refers to a SAT 9 performance standard of 3 or 4, demonstrating solid or superior reading performance on the SAT 9.

Retention refers to having repeated one or more grades.

### Method

#### *Subjects*

Subjects included one-hundred sixteen second grade students, 58 female and 56 male, from a rural elementary school in southeastern Ohio. Three of these students were retained and 34 received free or reduced lunch. The current research study was reviewed

and approved by the Marshall University Institutional Review Board.

*Instruments*

*Screening to Enhance Equitable Placement (STEEP)*

Screening to Enhance Equitable Placement (STEEP) is a screening instrument used to identify students at risk for reading and math problems. STEEP was developed by Louisiana State University professor Joe Witt. STEEP has been identified as highly accurate in identifying children who should and should not qualify for special education (Witt, VanDerheyden, & Ardoin, 2004 NASP). Despite evidence of early identification and remediation, poor trends in reading continue (Lyon, 2002). Dr. Witt created STEEP, which uses curriculum-based assessments for both math and reading development, in attempt to decrease the number of special education referrals. STEEP also increases early identification, develops appropriate interventions and guides instruction via benchmark goals (Witt, 2002). With the increasing demand for accountability and success on high-stakes tests, STEEP can be used to successfully predict which students will be at risk for failure on high-stakes tests while planning appropriate interventions to ensure they meet proficiency and help schools achieve adequate yearly progress (AYP). Within the next five years, half of the U.S. will require students to pass state-mandated tests before being awarded a high school diploma (Abrams & Madaus, 2003).

An initial reading probe, the class wide assessment, is used to assess Oral Reading Fluency (ORF), which is the number of correct words read in one minute. Based on their ORF, students are classified in one of three levels: Mastery, Instructional and Frustrational. STEEP offers many positive outcomes besides decreasing special

education referrals. STEEP resolves the problem of the wait to fail model by identifying at risk students early. STEEP data provides which students need interventions and of what type. In addition, STEEP enables teachers to begin ruling out contributing factors such as poor instruction, and/or educational disadvantage (Witt, VanDerheyden, & Ardoin, 2004 NASP).

*Stanford Achievement Test, Ninth Edition (SAT 9)*

The SAT 9 is a group administered standardized achievement test, for grades K.0 to 13.0, available to school districts and states. It is one the most popular and well developed standardized achievement tests on the market for monitoring student performance and progress (Haladyna, 1997 & Berk, 1997). The SAT 9 measures language skills, reading, spelling, listening, mathematics, social science, science and study skills. The SAT 9 was standardized based on stratified random samples totaling 450,000 students. The test sampled 250,000 students from 1,000 school districts in the spring of 1995, and another 200,000 students in the fall of 1995. The stratification variables used were socioeconomic status, urbanicity and ethnicity. The final standardization represented 49 states and the District of Columbia (Berk, 1997).

The SAT 9's reliability, or the extent to which the SAT 9 yields consistent results measured by the Kuder-Richardson Formula 20 (K-R20) coefficients are within the acceptable range of .80s to .90s for majority of multiple choice tests, and in .80s for most of the Total Reading. The listening, language, science and social science subtests coefficients were reported in the .70s and low .80s (Berk, 1997; Haladyna, 1997). The validity of the SAT 9, and the degree to which it measures what it is intended to measure,

was evidenced by content validity, criterion-related validity and construct validity. For both forms of most tests, except mathematics and science, these rates were all in the .90s (Berk, 1997; Haladyna, 1997). The entire SAT 9 Battery was scrutinized by an advisory panel of minority educators to control for bias/stereotyping related to gender, ethnicity, culture, socioeconomic status and geographic region to ensure items on the SAT 9 are valid for all examinees (Berk, 1997). Criterion-related validity was supported through correlations between the eighth and ninth editions; as well as correlations between open-ended and multiple-choice assessments. Finally, construct validity is supported by comparing the multiple-choice and subtests of the Otis-Lennon School Ability Test, demonstrating the interrelationships between school achievement and ability. Haladyna (1997) reported the construct validity between the Otis-Lennon and SAT 9 to be “fairly high,” while correlations between the SAT 8 and SAT 9 to be “about as high as reliability permitted.” Berk (1997) reports that comparisons between other achievement batteries, such as the California Achievement Tests and Iowa Tests of Basic Skills, would have been more useful and informative in establishing construct validity.

### *Procedure*

Permission was given by the principal in order to review the subject’s demographics, SAT 9 scores, STEEP data, student lunch status and to determine retention rates from the school’s student records. As all archival data was used for analysis, students’ identities were not needed, nor used for the purposes of this research study. Subjects’ socioeconomic status was derived by approved applications for free or reduced lunch during the 2004-2005 school year.

In the spring, students were administered the SAT 9. For the purposes of this study, January STEEP scores, the most recent probes administered prior to taking the SAT 9, were used. For the purpose of this study, only the SAT 9 Total Reading scores were analyzed. Logistic regression analysis was used to predict reading mastery on the SAT 9, dependent on gender, retention, SES and STEEP scores. SAT 9 served as the dependent variable coded as 1 for scored 3 or 4 (solid or superior reading performance on the SAT 9 Total Reading Performance Standards) and 2 for scored 1 or 2 (little/no mastery or partial mastery on the SAT 9 Total Reading Performance Standards). Independent variables included: sex, SES, retention and STEEP scores. Each student's final STEEP score, prior to taking the SAT 9, was entered; sex was coded 1 for male and 2 for female; SES was coded 1 for paid and 2 for received free/reduced lunch; retention was coded 1 for promoted and 2 for retained. The Statistical Procedure for Social Sciences (SPSS) version 11.0 was then used to compute and analyze the significance of the predictor variables. The alpha level was designated as  $p < .05$  indicating a significant result.

### Results

The current study examines the statistical significance of sex, SES, STEEP scores and retention as predictors of reading mastery. The results indicate STEEP scores as the single significant predictor variable in the regression model ( $r = .60, p < .05$ ). The model used STEEP scores to predict SAT 9 reading mastery. As STEEP scores increase, SAT 9 reading performance increases. Out of 116 cases, the model correctly identified 90 percent of cases obtaining solid or superior performance on the SAT 9 total reading

assessment. The model was less accurate at identifying cases who obtained little or partial mastery on the SAT 9 Total Reading by correctly predicting only 65 percent of the cases who did poorly on the SAT 9 Total Reading. STEEP is more effective at predicting who will achieve mastery. Overall the model predicted SAT 9 reading mastery with 81.6 percent accuracy.

### Discussion

The research study investigated the significance of sex, SES, STEEP scores and retention to predict reading mastery measured by the SAT 9 Total Reading performance standards. The hypothesis anticipated reading mastery will be significantly dependent on SES, gender, retention and STEEP scores. The implications of the study are that gender, SES, retention and STEEP scores can be used as a valid inventory of student's reading skills, when compared to their reading mastery outcomes, measured by the SAT 9.

Although the overall model is not significant, the non-significant predictor variables are worth retaining in future studies. SES and retention were significant alone, but, in the model, STEEP overpowered the other variables. Non-significant correlations were found between SES and SAT 9 reading mastery as well as retention and SAT 9 reading mastery. Although these effects were not found significant, according to the model, students identified as lower SES status along with students who had been retained were less likely to achieve reading mastery on the SAT 9. However, according to this data, SES does not predict reading outcome. The standard error of measurement for SES is too large and not confidently not predict reading outcome. Current literature supports that as years retained increase, reading performance decreases; however, no significant

correlation was found in the current study. The direction of these effects is expected and consistent with the literature.

These results reveal the importance of identifying students “at risk” for obtaining reading mastery. As the results imply, STEEP is a much better predictor of who will achieve reading mastery than who will not. Therefore intensive reading instruction needs to be given to students who perform poorly on STEEP in order to overcompensate for their weakness, as well as, to overcompensate for STEEP’s inability to accurately predict students “at risk” for no or little mastery on the SAT 9. SES does not predict reading outcome, the only significant predictor of reading outcome is STEEP. STEEP can serve as an effective intervention that allows educators to “catch students before they fall.” Educators can confidently use STEEP to identify students who will succeed and identify students who are “at risk.” If we want to ensure “at risk” students achieve reading mastery, those students need to receive intensive reading instruction. As indicated from the results, students who can read and obtain high ORF will achieve reading mastery on the SAT 9 Total Reading. Educators should focus on reading instruction as the explanation for reading failure rather than using socioeconomic status.

### *Limitations*

Due to the homogeneity of the population, several variables with implications for the results were unable to be included in this research. Due to the regional/Appalachia location, the study excluded racial and ethnic minorities and Limited English Proficient students. The current research underestimated SES and retention. There were not enough retained students represented, nor enough variability in SES represented to

generalize for every school system. Further research that includes a larger sample, along with the variables mentioned above may provide additional insight and validity for the current study. Replication is needed to determine if the results will generalize to other school systems.

References

- Abrams, L. & Madaus, G.F., (2003). The lessons of high-stakes testing [Electronic Version]. *Educational Leadership*, 61(3), 31-55.
- Albrecht, S.F., & Joles, C., (2003). Accountability and access to opportunity: Mutually Exclusive tenets under a high-stakes testing mandate [Electronic version]. *Preventing School Failure*, 47(2), 86-91.
- Anderson, Gabrielle E., Whipple, Angela D., & Jimerson, Shane R. *Grade retention: Achievement and mental health outcomes*. (n.d.). Retrieved February 17, 2005, from <http://www.nasponline.org>
- Baker, Spencer R.; McGee, Zina T.; Mitchell, Wanda S.; Stiff, Helen Randolph. (2002) Structural effects on academic achievement of adolescents. Retrieved February 18, 2005, from
- Barton, Paul E. (2004). Why the gap persist [Electronic version]. *Educational Leadership*, 62(3).
- Berk, Ronald A., (1997). Test review of the Stanford achievement test, ninth edition. From J.C. Impara & B.S. Plake (Eds.), *The thirteenth mental measurements yearbook* [Electronic version]. Retrieved January 26, 2005 from the Buros Institute's Test Reviews Online website: <http://www.unl.edu/buros>.
- Black, Susan (2004). Second time around: if repeating a grade doesn't help kids, why Do we make them do it? [Electronic version]. *American School Board Journal*. Nov, pp. 40-42.

- Burchinal, M.; Peisner-Feinberg, E.; Pianta, R.; Howes, C. (2002). Development of Academic skills from preschool through 2<sup>nd</sup> grade: family and classroom Predictors of developmental trajectories. *Journal of School Psychology, 40*(5), 415-436.
- Garcia, Gailyn M.; Stafford, Mary E. (2003). Prediction of reading by ga and gc Specific cognitive abilities for low-ses white and Hispanic English speaking Children [Electronic version]. *Psychology in the schools, 37*(3), 227-235.
- Goertz, Margaret & Duffy, Mark (2003). Mapping the landscape of high-stakes testing and accountability [Electronic version]. *Theory Into Practice, 42*(1).
- Gurian, Michael (2004). With boys and girls in mind [Electronic version]. *Educational Leadership, 62*(3), 21-28.
- Haladyna, T.M., (1997). Test review of the Stanford achievement test, ninth edition, from J.C. Impara & B.S. Plake (Eds.), The thirteenth mental measurements yearbook [Electronic version]. Retrieved January 26, 2005 from the Buros Institute's Test Reviews Online website: <http://www.unl.edu/buros>.
- Hart, B., & Risley, T.R. (2003). The early catastrophe: The 30 million word gap. *American Educator, 27*(1), 4-9.
- Jimerson, Shane R., & Kaufman, Amber M. (2003). Reading writing, and retention: A primer on grade retention research [Electronic version]. *The Reading Teacher, 56*(7) 622-635.
- Lyon, Reid G. & Chhabra, Vinita (2004). The science of reading research. *Educational Leadership, 61*(6), 12-17.

- Lyon, Reid G.(2002). Reading development, reading difficulties, and reading instruction educational and public health issues. *Journal of School Psychology*, 40(1), 3-6.
- Moe, Terry M. (Ed.). (2001). A primer on america's schools: An assessment by the koret task force on k-12 education. Stanford: Hoover Institution Press.
- Owings, William A., & Magliaro, Susan (1998). Grade retention: a history of failure [Electronic version]. *Educational Leadership*, 56(1), 86-88.
- Pomerantz, Eva M. (2002). Making the grade but feeling distressed: gender differences Academic performance and internal distress [Electronic version]. *Journal of Educational Psychology*, 94(2), 396-404.
- Rafoth, Mary Ann (2002 ). Best practices in preventing academic failure and promoting Alternatives to retention. In A. Thomas & J. Grimes (Eds.) *Best Practices in School Psychology*, 789-802.
- Rothstein, Richard (2004). The achievement gap: a broader picture [Electronic version]. *Educational Leadership*, 62(3)
- Torgesen, J.K. (2002). The prevention of reading difficulties. *Journal of School Psychology*, 40(1), 7-26.
- United States Department of Education (n.d.). Introduction: No child left behind. Retrieved January 29, 2005, from <http://www.ed.gov/nclb/overview/intro/index.html>
- Witt, J.(2002). Screening to Enhance Equitable Placement (STEEP). Retrieved on January 26, 2005, from <http://www.joewitt.org>

Witt, J.; VanDerheyden, A.; Ardoin, S. (2004). STEEP. Poster session presented at the

Annual meeting of the National Association of School Psychologists.

Retrieved on February 28, 2005 from,

[http://www.vail.k12.az.us/district/general/Steep/SteepResearchandPresentations  
NASPFinal200v7Revised.pdf](http://www.vail.k12.az.us/district/general/Steep/SteepResearchandPresentationsNASPFinal200v7Revised.pdf)