

## Standardized Tests as a Means of Showing Individual Projects

---

Suad Faisal Hassan, Babylon University, English Department, College of Education, Iraq,  
[suad7149@gmail.com](mailto:suad7149@gmail.com)

Prof. Hussain Hameed Mayuuf Humaidy (PhD), husm56@gmail.com, Babylon University, English  
Department, College of Education, Iraq

---

Article Received: 25/11/2023

Article Accepted: 22/11/2023

Published Online: 25/12/2023

DOI:10.47311/IJOES.2023.5.03.113

---

### Abstract

Standardized tests are commonly used to evaluate individuals' knowledge and performance. However, their efficacy as a tool for demonstrating individual projects is debated. Standardized tests measure general knowledge and skills across various subjects, whereas individual projects showcase specific abilities, creativity, and problem-solving skills. While standardized tests provide a standardized and objective way of assessing individuals, they may need to capture the full extent of a person's capabilities and unique qualities demonstrated through individual projects. Therefore, it is essential to consider alternative assessment methods that can effectively evaluate and acknowledge the value of individual projects.

**Keywords:** Tests, Projects, Objective, Assessment, Skills.

### Introduction

Standardized tests have been employed in public education as the principal method of evaluating learners. *Standardized tests* are commercially produced and consist of a fixed number of questions. These tests follow a consistent protocol for administering and scoring (Anderson et al., 1985; Popham, 1999). The sales of standardized examinations to public schools increased by almost 100% between 1960 and 1989, reaching annual revenue of \$100 million (Sacks, 2000). Supporters of standardized tests argue that they provide a fairer and more effective evaluation method. Gay (1990) states that standardized tests consist of predetermined protocols for administering and evaluating, and these exams establish benchmarks against which test takers' scores can be evaluated. The test items are based on empirical evidence rather than theoretical assumptions. They adhere to a standardized framework and utilize a consistent collection of materials. Additionally, all test takers are presented with identical tasks and must respond in the same manner (Gay, 1990; Standardized Tests, 1999).

The issue of test-driven education encapsulates the primary dispute surrounding standardized testing. According to Ratcliff (1995), these tests are not reliable indicators of one's competence and acquisition of knowledge. James and Tanner (1993) argue that the curriculum is restricted to emphasize just the skills assessed in the examinations. The limited

curriculum compromises the creativity, self-confidence, and love of school for independent thinkers. School transitions from a setting for play and organic exploration and learning to a formal occupation (Brown, 1993; James & Tanner, 1993; Kohn, 2001). Geocariss and Ross (1999) reported that certain administrators and instructors in public school districts have opted to discontinue standardized testing for young students. Instead, they have employed alternate assessment methods that align more well with children's learning styles. As of autumn 1996, 36 states were implementing various forms of alternative assessment. Ratcliff (1995) argues that alternative assessment should prioritize the learner.

The evaluation should chronicle developmental milestones, track progress, improve students' educational experience, and continue monitoring their goals and skills throughout their academic journey (Black & William, 1998; Ratcliff, 1995; Wadlington & Partridge, 2000). Educators require assessments that facilitate effective education planning and accurately reflect young children's evolving knowledge and skills (Farr & Greene, 1993; Ratcliff, 1995). Standardized testing is presently employed in public schools to yield comparable scores for individual students and identify their areas of proficiency and deficiency. Daniels (1999) states that standardized examinations identify specific content areas where a student may want assistance, provide overall measurements of achievement, and enable comparisons of students' talents and capabilities. Standardized examinations are employed to hold instructors, students, and even entire school districts accountable for their performance and evaluate their efficacy (Bowers, 1989; Popham, 1999). High-stakes testing yields numerous superfluous consequences that impact classroom learning.

As a result of these tests, Teachers are modifying their teaching methods and curriculum in response to their obligation to comply with legislation and meet parents' expectations for improved test scores (Brown, 1993; Hess & Brigham, 2000). The issue with the rise in high-stakes testing is that these assessments pose numerous formidable hurdles for both students and teachers. The message is that the sole determinant of significance in their educational journey is their exam results (Kohn, 2001). Instead of addressing the fundamental cause of academic underachievement, the emphasis is placed on comparing scores across different educational institutions. The schools that achieve high scores are regarded as exemplars, whereas those that achieve poor scores are perceived as failures (Harris & Longstreet, 1990; Sacks, 2000). Given the significance of test scores in making crucial educational determinations, it is necessary to challenge the credibility of these standardized examinations (Farr & Greene, 1993; Hurwitz & Hurwitz, 2000). Popham (1999) asserts that educators should dedicate time to carefully examining the individual assessments and the specific questions they include.

They must determine the precise variables being measured in each test. Furthermore, educators must enlighten the public that schools should be evaluated on something other than standardized test results. While accountability is essential, more reliable methods exist to assess student accomplishment. Daniels (1999) argues that standardized examinations presuppose uniform knowledge among all children, restricting their effectiveness in accommodating individual students' learning styles or requirements. Educators should

provide alternate options for standardized testing. Teachers must provide evaluation instruments that accurately evaluate valuable skills and knowledge (Popham, 1999).

This research aimed to provide a comprehensive examination of relevant literature regarding the function of standardized tests in evaluating young children. An analysis was conducted to evaluate the advantages and disadvantages of standardized examinations for administrators and teachers, encompassing a total of cons. Recommendations were made for alternative assessment measures that improved the recognition of individual variations and learning styles. Informal consultations were conducted with elementary administrators and teachers to provide up-to-date examples of approaches employed in contemporary classrooms. These approaches were also beneficial for campus plans in rural schools in Texas. Depending on their purpose, goals, and desired learning outcomes or proficiencies, these possibilities could also be used in other states.

### **Research Question**

1. What are standardized tests' potential biases and limitations in capturing the nuances of individual learning and growth?
2. How do standardized tests impact students' motivation, self-esteem, and overall learning experience?
3. How do standardized tests influence educational policies and practices, including curriculum design, teacher performance evaluation, and school funding allocation?

These research questions aim to delve into standardized tests' potential strengths and weaknesses as a tool for showcasing individual projects and determining academic progress.

### **Literature Review**

Standardized assessments have predominantly been employed in the field of education. Popham (1998) states that the primary objective of a standardized test is to assess and compare students' knowledge and skills on a national level. A standardized examination is conducted on a representative subset of students. The sample group's scores are utilized to compare the scores of prospective test takers (Eggen & Kauchak, 1992; Popham, 1998). The intelligence tests, the initial standardized assessments, were created during the 1900s to identify children who required additional assistance and segregate immigrant children into specialized courses (Standardized Tests and Our Children: A Guide to Testing Reform, 1990). In the 1920s, multiple-choice examinations were devised and employed to categorize pupils for educational objectives. Following the year 1950, there was a growing utilization of standardized exams for the specific aim of retaining and selecting individuals (James & Tanner, 1993).

Based on the research conducted by Farr and Greene (1993), testing did not assume a prominent position until 1957, coinciding with the launch of Sputnik. Education was perceived as the cause of the inability to achieve objectives. Anticipating the desired results, additional programmers were

created to assess the effectiveness of our educational initiatives, resulting in a rise in the amount of federal funding allocated to education (Farr & Greene, 1993; Hacker & Hathaway, 1991). As no other evaluation methods existed, norm-referenced standardized tests were mandated as the sole measure of progress. During this period, there was a progression in test development that involved the incorporation of multiple sub-sections to assess specific skills and objectives (Farr & Greene, 1993; Valencia, 1997). Farr and Greene (1993) argued in the 1970s that clearly defined goals and objectives were necessary for teachers to concentrate on their teaching priorities. This emphasis compelled test developers to incorporate additional sub-skills and objectives into their examinations, which then served as the framework for training. The popularity of Ditto masters, workbook pages, and programmed teaching increased significantly (Farr & Greene, 1993). In the 1960s, public leaders sought methods to enhance education and establish accountability. With the advent of computers and norm-referenced assessments, standardized testing emerged as a cost-effective method to evaluate student progress (Stiggins, 1999). By 1970, three states had initiated comprehensive testing programs across their entire state. Currently, there are 50 states, as reported by Stiggins in 1999.

An increase in concern for educational quality in the 1980s led to yet another increase in the use of standardized tests (Brown, 1993; Standardized Tests and Our Children: A Guide to Testing Reform, 1990). Stiggins (1999) reported that our nation's math and science results made many worry that our academic standing was too low and that America was headed for decay. Therefore, there was an increase in the use of standardized tests due to public concern. School personnel began to rely on test scores to make many educational decisions that affected a student's instruction (Perrone, 1991). According to Perrone (1991), a high school graduate of 1950 may have taken three standardized tests during his or her school career; however, a graduate of 1989 would have taken as many as 21 standardized tests. Neil and Medina (1989) stated that 105 million tests were used for 40 million students for the 1986-1987 school years. Stiggins (1999) concluded that the history of standardized tests revealed an increase each year in the use of standardized tests, with no objective evidence of a definite impact on classroom instruction. According to the National Assessment of Educational Progress, there has been no improvement in student learning since 1974, despite the increase in standardized testing (Sacks, 2000; Standardized Tests and Our Children: A Guide to Testing Reform, 1990). However, standardized tests have also been a significant means of student assessment in the last decade. Scores from these tests have been used to place students in gifted and talented programs, remedial classes, or special education programs. Results from standardized tests were used to determine eligibility in enrichment programs and a student's academic level and even became the basis for tracking (Facts, 1999; Perrone, 1991).

### **Conflicting Attitudes: Means of Assessment in Schools:**

The facts of standardized tests in the classroom often contradict public opinion. Based on the third edition of the Gallup Poll on the Public's Attitudes towards Public Schools, 43% of the respondents believe that the current level of testing in schools is appropriate. This number decreases by 5% compared to its value in 1997. The percentage of those who believe that there is an excessive focus on testing rose from 20% in 1997 to 30% in 2000. A study conducted by Rose and Gallup in 2000 found that 65% of the public believe that standardized examinations should be primarily used to

select instruction. However, 68% of the participants believe that the best measure of student progress should be based on the work done in the classroom or at home. Standardized exams allow for assessing students' performance by comparing it to a representative sample of the state or the entire nation (Elliott et al., 1998).

Standardized examinations are specifically created to assess and compare pupils nationally. These tests allow for directly comparing individual student's strengths and weaknesses in a reference group. Students are compared based on their familiarity with limited information (Popham, 1998). Commercial test publishers market the overwhelming majority of standardized examinations. These test developers strive to provide a series of standardized assessments that can be universally applied. However, achieving this 98% completion rate is logically impossible because these examinations will always include elements that may not be fully linked with all school curriculums (Bushweller, 1997; Popham, 1999).

Standardized multiple-choice tests are called objective due to the automated scoring process, eliminating any subjective influence on a child's score. Human beings remain intimately engaged when they use their agency in selecting the inquiries to pose and determining the precise language to employ in their formulation. Test creators determine correct or erroneous answers and establish passing grades (Facts, 1999; Popham, 1999; Standardized Tests and Our Children: A Guide to Testing Reform, 1990). Skinner (1994) asserts that standardized test publishers lack knowledge about the pupils for whom they create the test. Test authors are not immune to errors. Occasionally, questions may have a pair of valid solutions or no solution. If a child chooses two correct answers or leaves a question unanswered, their test will not be scored due to the machine's counting mechanism, resulting in a loss of points. Inaccurate markings or inadequate erasure might disrupt the functioning of a machine, resulting in an incorrect score (Popham, 1999; Standardized Tests and Our Children: A Guide to Testing Reform, 1990).

The standardized examinations yield test results that are used to generate statistics on specific classrooms, schools, and districts (Fair Test Examiner, 1996). Standardized test scores are subject to the influence of several factors. According to Popham (1998), kids' intellectual aptitude and the level of stimulation in their environments significantly impact their performance on standardized examinations. In schools with high socioeconomic status, kids generally achieve high grades because they can access various enriching activities in stimulating environments. Standardized tests correlate with socioeconomic status, as evidenced by Sacks' research in 2000. According to Sacks (2000), compelling evidence suggests that a student's test result may be accurately predicted by considering factors such as the family's income, educational qualifications, and the type of vehicle they own. Given these evaluation facts, it is unsuitable to evaluate a school's staff and efficacy solely based on test scores. Although teachers and administrators may perform exceptionally well in education, their scores may not accurately reflect their performance (Popham, 1998). Perrone (1991) argues that standardized examinations pose challenges for all ages but must be more credible for young children.

Children's growth during the formative years can frequently be misread due to varying developmental patterns, so the consequences of failing during this period can be detrimental. Ratcliff (1995) states that a single test administration can assess a child's abilities on a particular day

but cannot forecast their performance in any program. Ratcliff (1995) argues that tests are unreliable indicators of a young child's aptitude and knowledge acquisition. Standardized tests, for instance, are not explicitly crafted to captivate the attention of young children. These measurements aim to assess a child's attention to a task rather than their performance in a specific area or skill (Ratcliff, 1995). Test administration, anxiety, uncomfortable seating, severe temperatures, inadequate lighting, and noisy environments can all impact test performance. The impact of these alterations is particularly pronounced in young children. Any deviation from the norm can significantly impact a child (Herman, 1998; Standardized Tests and Our Children: A Guide to Testing Reform, 1990).

As Fair Test Examiner (1996), standardized tests lack perfect reliability as an individual's result can fluctuate from day to day due to factors such as testing settings and the mental state of the test taker. James and Tanner (1993) found that young infants frequently experience fear in response to all testing processes. The environment may appear daunting, fellow youngsters may find their concentration disrupted, and the exam supervisor may need to be more familiar. According to James and Tanner (1993), young children may fail to apply adequate pressure when using a pencil, resulting in light markings. They may also create unintentional marks outside the designated areas on their test booklet.

Additionally, young children may be less interested in the test content than in their surroundings. Test scores of young children are significantly less accurate than adult ones, as stated by the Fair Test Examiner (1996). Standardized test scores remain a fundamental measure and forecaster of academic achievements and shortcomings (Birrell & Ross, 1996; Sacks, 2000). The scores serve the purpose of identifying students with low performance for placement in special education classes and identifying academically bright students for enrollment in gifted and talented classes (Popham, 1999).

The primary purposes of standardized test results include making placement decisions for individual students, designing a personalized curriculum for a child's education, evaluating programmers, and ensuring responsibility for school performance (Bowers, 1989; Fair Test Examiner, 1996; Popham, 1999). According to a survey conducted by Brown (1993), around 25% of lawmakers believed it suitable to utilize standardized test scores for program evaluation. Seventy percent of the legislators concurred that the collected data from the exam was deemed suitable for informing legislative decision-making on education.

### **Accountability: Legislative Demands and Decisions**

Several of these interconnected legislative actions have had an impact on accountability. Harris and Longstreet (1990) state that before the 1970s, standardized exams were predominantly employed for individual placement, diagnosis, and monitoring. The influence on classroom behavior was negligible. In modern times, these assessments oversee not just individual students but the entire educational system. Teachers are currently under pressure not only to choose the content and methods of their instruction but also to ensure that their pupils attain good marks. Since the inception of the educational reform movement in the 1980s, legislative engagement has continually been associated with various educational directions and decisions (Brown, 1993). In the past five years, instructors have faced increased accountability for student learning, unprecedented in history (Eisner, 2001; Gay, 1990; Hess & Brigham, 2000). Legislators seek to ascertain the efficacy of

students' classroom performance and the extent of their mastery of the curriculum, as stated by Farr and Greene (1993).

Primarily, they desire to ascertain the relative performance of their kids in comparison to their peers around the nation, as the prevailing belief is that education catalyzes economic success. Therefore, standardized testing has been primarily justified over the past many years based on the need for accountability in student accomplishment (Engle, 1980; Fair Test Examiner, 1996; Tuch, 1996). Accountability is a valid and significant concern for parents, citizens, and politicians (Hess & Brigham, 2000; James & Tanner, 1993; Standardized Tests and Our Children: A Guide to Testing Reform, 1990). Accountability demonstrates pupils' learning and growth (Elliot et al., 1998). Bushweller (1997) asserts that numerous states hold themselves responsible for their educational outcomes solely based on test scores provided by commercial test publishers, without considering whether these tests accurately assess the standards the state has approved. In Texas, administrators can face termination, and school boards can be disbanded if test scores fail to meet the required standards.

If a school in Texas has a low rating for three consecutive years, state officials have the authority to designate a monitor who assumes the responsibilities of the school board. If a school achieves high scores on the Texas Assessment of Academic Skills (TAAS), it can substantially benefit the school. In Maryland, monetary funds are confiscated due to subpar test results, and in the event of persistent low scores, the state has the authority to assume control of the school. This scenario poses a predicament for administrators, teachers, and students, particularly when their occupations and the future of the school hinge on the outcomes of the standardized tests. Tuch (1996) is a reference to a specific publication.

Furthermore, it has been argued that accountability has placed significant pressure on teachers and administrators within school districts, resulting in an environment characterized by intimidation and burden. This is exacerbated by the government's imposing various mandates addressing societal issues. Stiggins (1999) suggests we may enhance teacher effort and student learning by using the fear of public humiliation.

The prevailing belief has been to enhance exertion through coercion by imposing severe repercussions for low test scores. Hence, educators must reassess their dependence on high-stakes exams as a means of public accountability and attaining educational superiority (Stiggins, 1999). Attaining responsibility is a complex and costly endeavor. High-stakes testing is a more convenient and cost-effective option when compared to alternatives such as recruiting and training skilled teachers, decreasing class sizes, or renovating dilapidated school facilities (Eisner, 2001; Hurwitz & Hurwitz, 2000; Popham, 1999). Politicians at various levels have advocated for the greater utilization of standardized testing to achieve accountability (James & Tanner, 1993; Moore, 1992; Sacks, 2000). According to Sacks (2000), Americans had allocated about 200 million dollars for testing in public schools by 1997. High-stakes testing grants the state authority to determine the most suitable academic content for children, allowing them to make educational decisions (Hess & Brigham, 2000).

Based on Brown's (1993) findings, state and federal policymakers show minimal regard for the testing rules they impose on school districts. Hurwitz and Hurwitz (2000) assert that educators at

the local level need more confidence in judgments made by lawmakers, mainly those persons or groups needing more educational expertise. Politicians create the perception of being leaders in education by criticizing underperforming schools or highlighting schools that have made significant improvements. There have been minimal improvements, primarily limited to enhancing students' test-taking abilities (Harris & Longstreet, 1990; Hurwitz & Hurwitz, 2000; Kohn, 2001; Popham, 1999). Schools that receive low ratings are subjected to intense public scrutiny (Hurwitz & Hurwitz, 2000). Undoubtedly, imposing stringent expectations on educational professionals within districts is imperative. Both teachers and students should be responsible for students' academic progress (Glickman, 2001; Kauthold, 1998). It is imperative that children receive the highest quality education available, and instructors must be held responsible for ensuring that children acquire knowledge. James and Tanner (1993) agree that teachers should deliver superior learning experiences and adjust their teaching methods when pupils struggle to learn.

Hurwitz & Hurwitz (2000) ascertain the actual underlying issue. Some argue that standardized tests are the sole means to achieve more excellent standards and enforce more demanding responsibility. Conversely, some individuals argue that these assessments exert control over teaching methods and disproportionately penalize pupils from impoverished and minority backgrounds. The question arises regarding how standardized assessments might be optimally utilized in public schools. Popham (1999) asserts that standardized examinations have significant instructional value but cautions against using them to assess the quality of education. "Using standardized achievement tests to determine educational quality is comparable to measuring temperature with a tablespoon" (Popham, 1999).

The increasing use of test scores in significant educational determinations brings the inquiry into the authenticity of standardized tests (Farr & Greene, 1993; Kohn, 2001; Popham, 1999). School districts allocate significant resources in terms of time and finances towards administering standardized examinations to facilitate effective instructional planning inside classrooms. As per Brown's (1993) findings, teachers do not prioritize test scores to the same extent as politicians and the general public. Moreover, standardized assessments need more data for teachers to adapt their instruction to meet the needs of individual students. Educators assert that their assessments are more dependable than the results of standardized tests (Brown, 1993; Hurwitz & Hurwitz, 2000; Kohn, 2001). Mitchell (1997) argues that traditional accountability differs from contemporary teaching practices. This issue can be ascribed to the material on the tests needing memorization and regurgitation, often resulting in a dull and uninteresting curriculum (Bushweller, 1997). Standardized testing continues to serve as a convenient method for the general public to comprehend the focus on fundamental principles. By exclusively assessing reading, math, and writing, the general public may readily discern comparisons among schools and specific classrooms (Jones et al., 1999; Eisner, 2001).

Nevertheless, curriculums sometimes prioritize the repetitive memorizing of fundamental core areas while neglecting significant courses like physical education, art, and music. According to Jones et al. (1999), evaluating critical thinking, artistic skills, student enthusiasm, and creativity is considerably more challenging. Regrettably, the challenges, duration, and supplementary expenses frequently surpass the overall advantages of various evaluations. Standardized tests typically

provide advantages regarding administration convenience and cost-effectiveness compared to other assessment methods discussed above.

### **Pros of Standardized Tests**

Harris and Longstreet (1990) and Coleman (2000) argue that standardized assessments provide significant benefits. These tests are meticulously designed by professionals, evaluated by machines, relatively simple to do, cost-effective, and impartial. Moreover, standardized examinations offer the general public a comprehensible evaluation of their child's or adolescents educational institution (Eisner, 2001; Harris & Longstreet, 1990; Popham, 1998; Sacks, 2000). Popham (1999) states that creating standardized tests aims to establish an assessment instrument encompassing a significant quantity of knowledge relevant to a particular grade or age group. Tests can only encompass some knowledge and abilities due to their excessive length. Therefore, test producers must devise an instrument that, with only a few items, can accurately assess a student's proficiency in a wide range of subjects. Items correctly answered by only 50% of the pupils are retained on the examinations.

Developers refrain from including items that elicit excessively high or low rates of accurate student responses. If the national norm group is a representative sample of the entire nation, educators and parents can utilize these test findings to draw meaningful conclusions about the pupils. Popham (1999) asserts that the creators of standardized tests develop assessment instruments that enable individuals to conclude the knowledge and abilities of a particular student in a specific subject area. The inference is norm-referenced, allowing for comparing knowledge and skills with other students of the same age or grade level. The 1817 data acquired from examination outcomes reveals the aptitudes and deficiencies of the pupils (Daniels, 1999; Popham, 1999). For instance, the assessment can determine that a student in the 4th grade is achieving at the 84th percentile in reading while only reaching the 39th percentile in science.

Determining a student's aptitudes and deficiencies in a specific academic domain is feasible. For example, suppose a mathematics test consists of 30 things, with ten items dedicated to computation, 10 to geometry, and 10 to algebra. In that case, it is possible to identify specific areas of concern. Nevertheless, these assessments frequently need more questions to provide significant comparisons (Popham, 1999; Thompson, 2001). Daniels (1999) states that standardized examinations identify specific content areas where a student may want assistance, provide overall measurements of achievement, and enable comparisons of students' talents and capabilities. These regions are applicable for communicating with parents regarding their child's aptitudes and capacities. Johnson (1981) conducted research including 298 school superintendents, 12 educational service agency superintendents, six state school superintendents, six state directors of statewide evaluation, and selected legislators. Two hundred eighty out of individuals, accounting for 84.34% of the sample, responded. Analyzed data pertained to the perceptions of school district superintendents regarding the effects of compulsory standardized assessment.

The study's findings indicated that mandatory testing programmers were seen as enhancing the capacity to collect and evaluate information about the needs of all tested students, promoting the utilization of test results to enhance instructional approaches in areas with low achievement rates in 1918, and fostering the utilization of student assessment and record keeping. The study also determined that obligatory testing programs led to increased contact with parents regarding kids'

strengths and limitations. Utilizing the tests led to an augmentation in communication with the community. These assessments offer a platform for educators and careers to converse about several facets of a child's education and advancement. Standardized assessments measure students' abilities to perform specific tasks within a specified timeframe (Harris & Longstreet, 1990; Mitchell, 1997). Proponents of standardized testing argue that these examinations identify a student's strengths and limitations, allowing for informed decisions regarding classroom training (Bowers, 1989; Popham, 1999).

The main benefit of standardized tests is their ability to mitigate biases in evaluating individual students and generate data that enables comparisons of different groups against a predetermined standard (Standardized Tests, 1999). Popham (1999) suggests that standardized examinations can provide valuable data on a student's progress over some time. Each year, it is possible to compare a child's results in several subject areas to determine if there has been considerable improvement or regression. According to Sacks (2000), employing these standardized examinations is an economical method for assessing students. Procurement agents can only choose resources directly relevant to specific objectives on the statewide assessment test (Hess & Brigham, 2000). Standardized examinations, designed for broad application, such as statewide assessments, are straightforward and comparatively simple.

The assessments are cost-effective, meticulously designed by professionals, and unbiased, making them suitable for automated scoring (Goodwin & Goodwin, 1982; Harris & Longstreet, 1990). Standardized tests provide an impartial method of evaluating the performance of instructors and individual schools (Coleman, 2000; Daniels, 1999; Harris & Longstreet, 1990). Skinner (1994) posits that confident educators perceive standardized examinations as the sole means of conducting an unbiased assessment for readers who depend on these results. This benefit enables the administration of a single exam by established norms without any subjective interpretation of answers. Standardized tests consist of questions with a single definitive solution, which results in quantifying findings based on the number of correct and erroneous responses. Excluded are test items that are either too easy or too challenging. Another benefit of standardized examinations is that they provide the test taker with knowledge about their national ranking among other students.

The manuals and scoring procedures that come with standardized examinations provide an additional benefit, with the most significant advantage being their accessibility (Skinner, 1994). Standardized assessments also have a positive impact on the curriculum. Hess and Brigham (2000) argue that standardized testing enhances the curriculum by providing teachers and students with a clear and precise grasp of the requirements for successful learning. Teachers determine the teaching objectives most likely to be assessed on the test and concentrate on those, enhancing their test scores. Standardized testing significantly mitigates the disparities arising from the varying curricula offered in different schools or even different classes within the same school (Harris & Longstreet, 1990; Hess & Brigham, 2000). Hess and Brigham (2000) argue that the absence of a standardized curriculum challenges our society. A child can significantly improve their grade point average by driving a short distance up the road and enrolling in a program different from their current one. Finally, certain educational districts perceive standardized examinations as beneficial in assessing

teacher performance. Teacher evaluations can prioritize the assessment of students' performance on standardized examinations (Hess & Brigham, 2000).

### **Cons of Standardized Tests**

While standardized examinations offer certain benefits, it is essential to acknowledge their numerous significant drawbacks. Hacker and Hathaway (1991) argue that while standardized examinations offer valuable insights, they lack authenticity and harm the educational system. The United States is the sole country that depends on standardized testing for extensive evaluation. Europe and Asia employ many assessment methods, including essays, oral exams, and exhibitions of students' work. These evaluation methods evaluate students' abilities and understanding more significantly, encompassing advanced cognitive capabilities such as higher-order thinking and problem-solving.

In contrast, standardized examinations primarily concentrate on specific, isolated skills. James and Tanner (1993) and Popham (1999) have stated that it is challenging to estimate academic performance for a child aged 5 or 6 due to their developmental stage. Young children are inherently unstable; their abilities and skills transform as they progress.

The outcome of a test administered on one day may vary significantly when administered on another day. Standardized multiple-choice tests are often called objective due to their machine-based scoring system, eliminating any subjective influence on a child's score. Nevertheless, humans play a significant role as they actively select the questions to pose and determine the specific wording to employ while posing them. Thompson (2001) asserts that test-makers are responsible for determining the accuracy of answers and the passing mark for exams. Standardized exams provide a misleading perception of impartiality as humans construct the test questions (Facts, 1999). Another premise of standardized exams is that if a student can demonstrate a skill in the examination, they can also demonstrate the same proficiency in their work. For instance, if a student can edit written material without considering its context, as demonstrated in a standardized examination, they will also be capable of editing their work. Hacker and Hathaway (1991) argue that the context of a particular ability is significant and cannot be separated.

Valencia (1997) asserts that standardized examinations emphasize specific skills, promote shallow understanding, exclusively rely on multiple-choice forms, and generate ratings that often need to be more helpful for instructional preparation. Students need more active participation in their assessment. Due to their infrequent administration, standardized exams cannot effectively capture a student's learning progression over time (James & Tanner, 1993; Valencia, 1997). Multiple-choice assessments do not evaluate cognitive skills at a higher level, such as writing, mathematical proficiency, or the capacity to comprehend and interpret written material. Additionally, they do not gauge the practical abilities of young children in real-life situations (Eisner, 2001; Fair Test, 1992; Geocaris & Ross, 1999). Multiple-choice tests have limitations in assessing complicated thinking and do not evaluate the affective domain, which includes feelings, interests, and attitudes (Davies & Wavering, 1999; Wadlington & Partridge, 2000). Since most standardized tests are in the form of multiple-choice questions, originality and expression are discouraged and result in negative consequences.

Therefore, it is likely that the skills being assessed are primarily focused on the ability to memorize and repeat specific pieces of information in the short term. The assessed quantity is convenient to quantify but may not necessarily be the most important (Harris & Longstreet, 1990). Bigelow (1999) and Johnson (1981) argue that standardized test questions prioritize isolated facts and fail to address the broader and more complex implications of those facts. Teachers experience a sense of pressure to engage in rote instruction, focusing on memorizing information rather than fostering a deep knowledge of the presented event or situation. Gilman and McDermott (1994) and Kohn (2001) assert that testing has grown so widespread that it has led to the fragmentation of learning into discrete skills. Hacker and Hathaway (1991) state that early psychological theorists believed that the mind consisted of discrete knowledge that could be deconstructed into more minor elements. It was commonly believed that to assess a person's reading ability, it was sufficient to evaluate simply the specific subtask that constitutes the skill of reading.

This strategy has faced criticism over the years. Although significant advancements have been made in our understanding of the brain and the learning process, our perspectives on testing have remained unchanged. Most standardized examinations rely on memorizing individual facts and limited abilities (Fair Test, 1992; Standardized Examinations and Our Children: A Guide to Testing Reform, 1990). Another issue about the improper use of standardized examinations is that these assessments have frequently been employed to control classroom activities (Herman, 1998; Kohn, 2001). Kohn (2001) argues that principals have reduced extracurricular activities, arts programs, recess, electives, and other activities to prioritize the content covered in tests. Jones et al. (1999) conducted a study with 470 certified teachers from 16 elementary schools throughout five school districts in North Carolina. The schools were chosen randomly, ensuring a proportional representation of rural, urban, and suburban school systems. Out of the 236 teachers who provided feedback, 89% identified as Caucasian, 10% as African American, and 1% as Hispanic.

The teachers provided information regarding the annual duration of students' preparation for standardized end-of-grade exams. Eighty percent of the teachers said that pupils spend over 20% of their instructional time engaging in test preparation. Twenty-eight out of the eighty percent reported that sixty percent of instructional time was dedicated to exam preparation. This period was allocated from standard teaching, resulting in a reduction of the curriculum to only the principles that were evaluated by the state. The same educators were requested to specify the influence of testing on their pupils. 61% of the teachers perceived increased worry among their pupils about learning, whereas 24% observed decreased student confidence. Remarkably, 45.8% of educators expressed that standardized assessments hurt pupils' enthusiasm for acquiring knowledge. 76% of polled instructors reported feeling increased job stress due to testing. Teachers conveyed emotions of remorse, unease, and stress. An essential score on a standardized exam can lead to implementing a limited curriculum for several youngsters (Popham, 1999). Standardized tests frequently restrict and divert instruction and fail to enhance teaching and learning. Some instructors express concern that examinations influence the curriculum, leading it in undesirable ways. Based on research by Sacks (2000), 85% of Texas educators assert that pupils are solely acquiring test-taking abilities in the classroom. Farr and Greene (1993) propose that exams have evolved into a comprehensive plan for guiding education (p. 23).

According to Clay (1993), standardized test scores are considered simple approximations and frequently distort an individual's actual progress in learning. They need to convey the necessary information to develop or assess effective instruction. Popham (1998) suggests that test items answered correctly by at least half of the students are most effective in distributing students' scores. However, test items that receive a correct response rate of 80% or above from students are initially excluded from the exams and are likely to be removed during the test revision process. However, if teachers effectively foster proficiency in crucial skills and information, it is improbable that such proficiency will be assessed using standardized assessments. Brown (1993) and Hurwitz and Hurwitz (2000) assert that while others argue that compulsory testing is in children's best interest, numerous studies have shown that classroom education has become inflexible and devoid of purpose for pupils. Students frequently experience boredom and develop unfavorable views toward learning (James & Tanner, 1993; Kohn, 2001; Sacks, 2000).

In a study by Mitchell (1997), 20 school administrators provided insights on the impact of standardized testing on school restructuring endeavors. Principals asserted that standardized assessments served as an obstacle to school reorganization. According to them, the tests were causing their curriculum to deviate from the reform efforts. Increasing emphasis on improving standardized test scores leads to a skewed curriculum (Tuch, 1996). The specific objectives of individual classrooms often need to align more precisely with the objectives of standardized assessments (Standardized Assessments, 1999). Often, the tests themselves end up serving as the curriculum. Perrone (1991) states that individuals spend approximately two to three hours daily engaging in exam practice and related exercises. Ratcliff (1995) states that numerous classrooms have adopted formal instruction and incorporated activities resembling tests. The curriculum has experienced a decline in its relevance to students and needs to be more efficient in appropriately equipping them for the future. Textbooks are frequently simplified because the test-making firms produce them to boost test results (Popham, 1998; Standardized Tests and Our Children: A Guide to Testing Reform, 1990).

In 1983, Freeman, Kuhs, Porter, Floden, Schmidt, and Schwille conducted a study examining the content of 5 nationally standardized mathematics tests administered to pupils in grades 4-6. Presuming that the content taught in the classroom aligns with the information presented in the textbooks, they also engaged with textbooks designed for students in grades 4-6. Their findings, consisting of 2726 data points, substantiated that a significant proportion, ranging from 50 to 80%, of the content assessed in the standardized examinations was not included in the textbooks. Hence, the substance of the standardized assessments needed to align with the curriculum taught in the classroom. Standardized assessments restrict the curriculum by prompting teachers to focus primarily on the content assessed in the examination. The test effectively supplants the curriculum, deterring effective pedagogy and meaningful knowledge acquisition (Facts, 1999). According to Dounay (2000), standardized assessments have simplified the public school curriculum. Students are exposed to the practice of memorizing information mechanically rather than developing genuine problem-solving abilities. Teachers often prioritize specific subject areas when designing tests. The importance of art, music, and physical education is reduced (Kaufhold, 1998; James & Tanner, 1993; Perrone, 1991).

According to James and Tanner (1993), the curriculum is restricted to emphasize only the skills assessed in the assessments. The demand for improved test scores frequently results in a restricted curriculum that is harmful, especially to the development of young children. Children are discouraged from cultivating autonomous thinking. It compromises their inventiveness, self-assurance, and pleasure derived from education. School transitions from a space for play, organic exploration, and learning to a formalized occupation (Dounay, 2000; James & Tanner, 1993; Kohn, 2001; Moore, 1992). Popham (1998) suggests that teachers can acquire knowledge of the exam topic, which can then influence their instructional choices in the classroom. The instructions often require teachers to take the test to acquaint themselves with it before administering it. Hence, increases in test results over time may be attributed to the teachers' growing familiarity with the material rather than the actual academic progress of the students.

### **Types of Tests**

Objectivity in evaluation pertains to the administration of the evaluation procedure and the scoring of its outcomes. Both should be standardized to minimize measurement errors, ensure reliable comparability, and mitigate bias. By itself, this needs to be improved. We can objectively evaluate responses to a nonsensical series of inquiries, which will not be beneficial. Put, practical tests must be unbiased and accurately represent the knowledge we are trying to evaluate. However, it should be noted that identical scores do not necessarily have to be interpreted in the same manner or have the same implications. These factors may also be influenced by how society collectively assigns value to the scores, considering additional societal factors such as equity, fairness, and potentially others. Maintaining a clear distinction between these other factors is crucial, explicitly weighing them against the meritocratic exam score. If we fail to take action, we may compromise the impartial outcomes of the merit-based system and contaminate the evaluation process with apparent prejudices. This, in turn, could result in the erosion of public confidence in its inherent equity and neutrality. Educational assessments can be categorized based on their intended use (diagnostic, formative, summative), their nature (achievement, aptitude), the types of questions they include (chosen response, constructed response), and the way the results are reported (standard-based, norm-referenced). Diagnostic tests aim to ascertain learning problems and will not be discussed in this context.

Formative exams, delivered by classroom teachers to assess students' progress and adapt instruction, and aptitude tests, designed to measure students' potential, will not be effective. This article will primarily examine summative achievement tests, which have increasingly become standards-based. These tests are often referred to as criterion-referenced examinations, as they measure accomplishment against a predetermined set of educational grade-level expectations, also known as standards. This approach is different, unlike prior norm-referenced examinations that evaluated achievement by comparing it to representative reference groups, usually at a national level. Another significant classification arises about the utilization of testing outcomes. In order to provide information to individual kids and parents, the test must be administered in a census manner, meaning it must be given to each individual. (Bishop,1997).

The public frequently associates objective and standardized assessment with the structure of selected response tests, commonly known as "multiple choice" tests. Although the association is

understandable, it needs to be more accurate. The objectivity of assessment relies on the explicit and unambiguous presentation of genuine test items, the consistent administration of the test, and a standardized method of scoring and scale. It is not contingent on the specific format of the test items. Considering these criteria, it is evident that creating an unambiguous multiple-choice question is more straightforward than formulating an open-ended question in the same manner.

Furthermore, it is even more apparent that ensuring objectivity in grading a multiple-choice question is more accessible than grading an open-ended question. Scoring a multiple-choice item is far more cost-effective than scoring an open-ended one. As a result, there is a common misconception that objective evaluation is limited to the multiple-choice format, even though objective examinations can also incorporate open-ended constructed-response questions. In summary, it is widely acknowledged that rigorous and objective assessments are an essential element of our educational systems. Deprived of the knowledge they impart, we would all be disoriented: Parents would need more awareness of how the school system was providing education to the children, but policymakers needed more means to assess the effectiveness of their programs or identify areas for improvement.

Similarly, teachers could not track their students' performance in previous grades or compare it to other classes and schools. Furthermore, students needed to be made aware of their proper understanding of the material. Without objective assessments, our education system would lack direction, like a ship without a navigation system. It would need to be made aware of its location and destination and would have no means to rectify its path. The only action it could take would be to accelerate the engines and rely on optimism for a favorable outcome.

Nevertheless, it is essential to acknowledge that numerous educators harbor a strong aversion towards testing and firmly feel that testing, especially high-stakes summative testing is detrimental to educational progress. This is expected because testing essentially serves as a possible criticism of educators and the teaching methods they advocate. However, these concerns would be more valid if the results of our education system, especially in Western countries, had been more robust and effective. The general public acknowledges the self-interested aspects of educators' protests and strongly favors educational testing. An instance of this is a survey carried out by Mark Holmes in Ontario in 1998, which revealed that only 11% of directors of education were in favor of annual achievement testing for elementary students, while 59% of a comparable group of no educators with similar levels of education supported the proposal (Holmes, 1998). According to a recent survey conducted in the United States, over 68 percent of Americans expressed their support for conducting student testing every year.

In contrast, only 48 percent of teachers shared the same viewpoint. In addition, a lower percentage of teachers who belonged to teacher unions supported testing than non-members, as Education Next reported in 2018. A survey conducted in 2022 among Canadian parents revealed that a significant majority of 84 percent expressed support for standardized testing despite the opposition from various teachers' unions (Mac Pherson in 2022).

**Standardized Tests are a Vital Source of Information for Parents and Policymakers:**

Previously, the educational level achieved by a student was considered a satisfactory measure of their degree of education. Due to many historical factors, the situation has changed, and there is now a greater demand for more accurate evaluation of students' education. For instance, if we consider two Maryland graduates with comparable GPAs, one hailing from Baltimore and the other from Bethesda, they would not likely have achieved equal proficiency in reading and mathematics. School Grade Point Averages (GPAs) and other teacher assessments lack reliability as markers of student achievement. Furthermore, the quality of educational achievement cannot be determined solely by the titles of courses completed or grade advancements. For instance, the newly published US 2019 National Assessment of Education Progress (NAEP) High School Transcript Study reveals that although students enroll in more challenging courses and achieve higher marks, their understanding of the subject matter has remained the same (Hess, 2022). In order to establish a meaningful comparison between schools, it is essential to have a dependable standardized metric.

The primary benefit of standardized test results is that they offer parents and policymakers an impartial and comparable assessment of student achievement. This allows them to evaluate schools and districts in comparison to one another using a standardized metric. Standardized tests serve as the sole genuine external assessment of the caliber of our educational institutions. Richard Phelps, a former researcher at the US Government Accounting Office (GAO), succinctly said that testing the curriculum is necessary to determine if it has been effectively taught (Phelps, 1999, p. 25). Standardized examinations are necessary to accurately assess student development for individuals not directly involved in the classroom. There are no education officials at the district or provincial level. A political leader is absent. No individual pays taxes. Each pupil must unquestioningly accept the teacher's instructions. Without standardized assessments, teachers lack external benchmarks for comparison outside their experiences. The foundation for any coherent and logical conversation regarding funding priorities or program development and execution relies on objective and comparative data on school performance.

Furthermore, objective measurements establish and secure discussions regarding education policy, preventing them from becoming disconnected from the actual issues at hand. This is particularly important as such discussions can be influenced by demagogues advocating for unproductive reforms or interest groups aiming to maintain the existing state of affairs. Teacher evaluations need to be more comprehensive in this aspect. Although teacher assessments can contribute to our comprehension of student performance, they cannot substitute standardized exams in providing information about school quality to parents and policymakers. There are multiple factors contributing to this phenomenon. One issue is that individual teachers can limit the curriculum to their personal preferences. Grades can be subject to inflation when regular teachers are involved, as pupils become more familiar with a teacher and understand their unique characteristics. According to Phelps (1999), a teacher's grades and test results are more likely to be unique and not applicable to a larger population than standardized tests.

A summative objective standard-based test can be a highly successful instrument for assessing whether the intended curricular content has been taught. This is because such a test is designed to

sample that content comprehensively. Indeed, the critique that it promotes "teaching to the test" actually becomes a favorable characteristic when it comes to standard-based testing, as "teaching the curriculum" and "teaching to the test" become indiscernible. It is a frequent occurrence for grades in schools, and even within specific classrooms, to converge towards a state of equilibrium, where most grades fall within the A and B range without noticeable gains in academic performance (Evers, 2001). This reveals significant disparities in the caliber of teaching and the pace of academic progress among various schools and educators. Teachers' in-class assessments frequently fail to gauge their pupils' mastery of the subject matter accurately. A comprehensive evaluation is conducted while determining a grade. There are likely two factors contributing to this. Educators must thoroughly contemplate all pertinent facets of a student's classroom experience before issuing a grade. Simultaneously, there is no agreement regarding the variables to consider when determining a grade (Cizek, 1996, emphasis in original). For numerous educators in education schools, expertise in the subject matter is merely one of the considerations, and in fact, may not even be the most crucial one, as expressed by an elementary teacher: "Ensuring the child's progress with a positive mindset and pleasant recollections is more significant than a numerical grade..." The prioritization of shaping children's cognitive development through group interaction, exertion, and active engagement holds greater significance than calculating average scores from examinations and quizzes (Cizek, 1996).

### **Testing Makes Our Schools More Accurate and More Responsive**

Instead of discouraging kids, formative and diagnostic testing helps teachers detect issues early, allowing them to immediately address these problems before the student falls significantly behind and gets genuinely disheartened. Early reading intervention prevents struggling pupils from experiencing long-term reading difficulties. According to Barbara Foorman and her colleagues, there is less evidence to suggest that pupils who are considered late bloomers in reading skills can make up, despite the common belief among educators in developmental delay. Thankfully, studies demonstrate that taking action early can have positive results, and it is crucial to identify the issue promptly through objective diagnostic tests to achieve this goal. Foorman et al. reference a study that identified kindergarten children with inadequate phonological awareness, meaning they struggled to blend and segment sounds in speech. By the time they reached second grade, individualized tutoring led to a significant improvement in reading skills for 75 percent of the students, bringing them up to the expected level for their grade. A further study discovered middle-class students who exhibited meager word recognition abilities at the onset of first grade. Following a single semester of individualized instruction, 70 percent of the participants achieved proficiency in reading at the expected grade level. After completing two semesters, almost 90 percent of the students achieved proficiency in their respective grade levels. (Foorman, Fletcher, and Francis, 2019).

- < UNK > For the intervention to be successful, it must commence before the pupil reaches third grade, thus emphasizing the crucial need for early diagnosis. Consequently, it is necessary to conduct comprehensive testing of every student methodically.

- According to Foorman et al., performance-based assessments that require students to generate responses and demonstrate their procedural knowledge are considered insufficient for the task. The assessments mentioned in the quote by Foorman, Fletcher, and Francis (2019) lack proof of reliability and validity.
- Furthermore, they do not measure knowledge transfer without being influenced by a particular curriculum. The authors assert that utilizing multiple-choice formats to evaluate declarative knowledge carefully and thoughtfully may be the most accurate, consistent, and beneficial approach (Foorman et al., 2019).

### **Standardized Tests Effectively Capture What Students Know**

Despite teachers' good intentions, it is a fact that their assessments are unlikely to be as precise in evaluating students' understanding compared to standardized examinations. The issue is that teachers, even those who try to grade their students based on academic performance, are unlikely to have received extensive training in testing and measuring. Those who critique standardized tests for their purported flaws in structure and content often fail to acknowledge that these tests are developed, evaluated, and refined by extensive teams of Ph.D. holders with specialized expertise in testing and measuring (Phelps, 2003). Therefore, standardized examinations are expected to possess a higher degree of accuracy and dependability than evaluations conducted by teachers. The endeavor dedicated to the creation of a significant standardized test is vast. A typical extensive examination undergoes numerous research, development, and pilot testing cycles to ensure its reliability. In other words, the scores it produces are consistent and stable across multiple test administrations. A comprehensive examination is also considered valid, indicating that it accurately assesses the intended subject matter and enables users to make precise and significant inferences regarding students' knowledge and abilities" (Kober, 2002). The idea that standardized tests, especially those with multiple-choice questions, are fundamentally basic and only promote low-level thinking skills is equally erroneous. Richard Phelps states that test items can range from mundane and uncomplicated to very intricate and can be presented in a multiple-choice or open-ended fashion. There is no inherent relationship between the complexity of a problem and the format in which an answer is expected. Even complex and comprehensive assignments that take fifty minutes to complete, involving classification, assembly, organization, calculation, and analysis, ultimately provide the test taker with a multiple-choice format for their response. Although the answer to the question is included in the options given, it may mean something other than it is straightforward or apparent how to get at the answer.

Some critics contend that the validity and reliability of a standardized test hold little significance, as it may not accurately assess an individual's knowledge and abilities. The Importance of Educational Testing Essential aptitudes that hold significant value in future life, such as ingenuity, collaborative proficiency, and the like. This claim is disproven by several studies that connect students' success on reliable, standardized examinations and their economic well-being at the individual and national levels. If the skills assessed by standardized tests were not pertinent, then no correlation would be observed. Stanford's Eric Hanushek provides three recent studies as

examples, all utilizing distinct nationally representative data sets that track students beyond their schooling years and into the workforce. These studies reached a consistent finding, indicating that a one standard deviation gain in mathematics performance at the end of high school corresponds to a 12 percent rise in annual wages when scores are standardized. To clarify, a one standard deviation increase in median earnings in 2001 would result in a \$3,600 increase in earnings for each year of work life (Hanushek, 2006).

Hanushek suggests that these estimates represent the minimum impact of high achievement. The number is 6. Gregory Cizek highlights a crucial aspect sometimes overlooked by critics who argue that tests fail to assess all the knowledge and skills required by students. The correlation between standardized performance exams and the fundamental objectives of education, such as developing responsible and productive citizens, heavily relies on being 'necessary but not sufficient.' Competency in reading, mathematics, writing, and other subjects is crucial for attaining vital educational objectives. Undeniably, attaining the aim also requires other student attributes, such as personal accountability and innovative cognition or analytical reasoning. However, more than an extra degree of personal responsibility or creative thinking is needed to compensate for shortcomings in a student's grasp of language or mathematics or in the student's ability to articulate his or her thoughts. Students may become effective and responsible participants in society if they acquire core academic abilities; they will likely be unable to do without them. (Cizek, 1998)

### **Standardized Tests Are Cost Efficient**

Multiple-choice tests, which are standardized and primarily consist of multiple-choice questions, are cost-effective. They can extract comprehensive information about many students within a short timeframe and at a significantly lower cost compared to other less standardized methods like essays or complex, open-ended "performance tasks." Caroline Hoxby, an economist from Harvard University, analyzed the expenses of 25 evaluation systems in the United States during the 2000-01 year. The prices varied from \$1.79 per student in South Carolina to \$34.02 per student in Delaware. However, Delaware's spending accounts for a mere 0.4 percent of the average amount spent per student in the United States, which was \$8,157 for the 2000-01 school year. Hoxby also compares the expenses associated with the creation and execution of assessments and two other commonly adopted reforms: reductions in class size and salary increases for instructors. Start by implementing class-size reduction (Cizek, 1998); reducing class size by 10 percent in American schools would cost approximately \$615 per student, 12,399 percent higher than the current average assessment cost.

This cost includes teacher compensation, which accounts for 54 percent of school costs, and expenses related to the size of school buildings, such as heating, which make up an additional 22 percent of costs. The increase in teachers' salaries is comparable: "Raising teachers' compensation by 10 percent would require an additional \$437 per student in the average American school," which is 5,011 percent higher than the present average assessment cost (Hoxby, 2002). According to a 1993 research conducted by the US Government Accounting Office [GAO], the mean expense of state testing was approximately \$15 per student, which accounted for staff time required for multiple choice examinations. However, when the testing incorporated specific open-ended topics, the cost

rose to almost \$20 per student. In 1990-91, the amount spent per student was \$5,885. To clarify, the expenditure on testing accounted for approximately 0.3 percent of the total amount spent per student. This amount is negligible when considering the importance of ensuring quality assurance [US GAO, 1993].

### **What Does the Research Show About Testing and Student Achievement?**

When combined with an accountability framework, standardized assessments can significantly impact student achievement. Research conducted by Phelps (2019, 2012) and Bishop (2004, 2005) suggests that accountability systems targeting kids are the most efficacious. However, these systems can still provide substantial positive outcomes even when directed toward schools, as exemplified by the No Child Left Behind (NCLB) exams. Tom Loveless, a researcher from the Brookings Institute, refers to a study conducted in 2003 that analyzed the impact of accountability on state NAEP ratings. In this study, Martin Carnoy and Susanna Loeb assessed the effectiveness of each state's accountability system using a five-level scale (Loveless, 2005; Carnoy & Loeb, 2002). The ranking took into account both student and school accountability. Carnoy and Loeb discovered that, while considering variations in spending and student demographics, there was a positive correlation between the strength of the accountability system and the period between 1996 and 2000.

An increase of two ranks in the accountability index was associated with a statistically meaningful gain of approximately one-half standard deviation. The findings showed a substantial and favorable outcome for black, white, and Hispanic kids. Furthermore, these results were consistent even after accounting for the number of pupils excluded from NAEP testing in each state (Loveless, 2005: 9). John Bishop, a researcher from Cornell University, conducted a study that analyzed educational systems aimed at students and educators. The study focused on the 1996 and 1998 NAEP scores of eighth graders in states with varying accountability regimes. These regimes included requirements for students to meet basic course requirements, pass minimum competency exams, and pass a curriculum-based external exit exam. Schools were also rewarded or sanctioned based on test scores (Loveless, 2005). He discovered that students in states that mandate curriculum-based external exit exams, such as New York and North Carolina, demonstrated the highest levels of academic performance. These students had an advantage of 0.45 grade levels in math and science.

Following closely behind were states that used rewards and sanctions to hold schools accountable, which resulted in gains of 0.20 grade levels. The minimum competency tests had a favorable impact, albeit not statistically significant. The source cited is Loveless (2005), pages 9-10. As noted by Bishop in his regression analyses, accountability systems generate incentive effects. These effects include enhancing knowledge and skills resulting from students dedicating more time to studying, paying greater attention in class, and enrolling in more challenging courses within a high-stakes testing environment (Phelps, 2003). Bishop conducted research on high-stakes exit tests in Canada during the early 1990s, when, similar to the present, only certain provinces had curriculum-based exit assessments. He discovered that the examination systems had widespread impacts on school officials, teachers, and students. In provinces where external exams are

conducted, schools are more inclined to employ specialized mathematics and science teachers and teachers who have pursued these subjects in college. These schools also tend to have well-equipped science laboratories and allocate additional hours for instruction in mathematics and science.

Furthermore, they assign more homework in mathematics, science, and other subjects, engage students in hands-on experiments during science classes, and administer frequent tests in mathematics and science. Similarly, teachers in these provinces are more likely to assign more homework, cover more challenging topics, and schedule a more significant number of quizzes and tests. They also reduce students' time on group problem-solving activities, increase the time students work independently, and incorporate more experiments into science classes. An analysis of curriculum and instruction systems in 30 countries that took part in the 1994-95 TIMSS demonstrated a significant correlation between the number of decision points, specifically high-stakes selection points, which function as quality controls, and the performance of students on the TIMSS 8th-grade mathematics examination. According to Phelps (2003), a country's performance improves as the number of high-stakes selection points increases. The study indicates a strong correlation between quality control and student accomplishment. Once sufficient quality control measures are implemented, student achievement can significantly improve, following an exponential pattern (Phelps, 2001; Bishop, 1997).

### **Conclusion**

Standardized tests have long been used to evaluate individual performance and provide a conclusive measure of knowledge and skills. However, when showcasing a person's true potential and project, the efficacy of these tests could be better. While they can provide a snapshot of a person's academic abilities, they often need to consider the multifaceted talents and strengths that cannot be captured through a simple exam. A person's creativity, critical thinking, and problem-solving abilities, essential in real-life projects, must be accurately reflected in standardized tests. Therefore, relying solely on these tests to measure an individual's capabilities may not provide a comprehensive understanding of their true potential. It is essential to consider alternative assessment methods that allow individuals to demonstrate their skills and achievements holistically and authentically.

Standardized tests are frequently utilized to evaluate personal performance and gauge a person's capabilities. However, relying on these tests to gauge one's abilities can be unrestricted. While standardized tests offer a consistent and unbiased manner of assessing individuals, they fail to account for the complete scope of a person's abilities, talents, and potential. Qualities like creativity, critical thinking, and problem-solving skills may need to be adequately assessed through these tests. Therefore, it is vital to view standardized tests as just one component of the evaluation process, not the sole factor in determining an individual's abilities and potential.

Using contemporary, standardized, curriculum-aligned educational testing is crucial for upholding a meritocratic system characterized by rigorous standards. Previously, commonly utilized norm-referenced tests were not closely related to the curriculum. However, this is now the case for

curriculum-aligned assessments. In addition, contemporary examinations undergo thorough scrutiny to identify and eliminate any group biases that may have been present in previous assessments. Criticisms of standardized testing by interest groups are fueled by obsolete and inaccurate information, a hesitancy to acknowledge and reward excellence, and a resistance to implementing necessary changes based on the results of these exams. Standardized exams associated with the curriculum and including consequences for students encourage increased academic performance and school enhancement by effectively communicating to all parties involved - students, parents, teachers, and administrators - what areas are prosperous and what areas require improvement. It is hardly surprising that the public strongly favors such testing.

### References

- Bishop, John H. (1997). Do Curriculum-Based External Exit Exam Systems Enhance Student Achievement? Working Paper 97-28. Center for Advanced Human Resource Studies, Cornell University. <<https://ecommons.cornell.edu/handle/1813/77025>>, as of May 18, 2022.
- Brown, D.F. (1993). The political influence of state testing reform through the eyes of principals and teachers (Report No. EA-025-190). Atlanta, GA: Conference Paper. (ERIC Document Reproduction Service No. ED 360 737)
- Bushweller, K. (1997). Teaching to the test. *The American School Board Journal*, 184(9), 20-25.
- Bushweller, K. (1997). Teaching to the test. *The American School Board Journal*, 184(9), 20-25.
- Carnoy, Martin, and Susanna Loeb (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Education Evaluation Policy Analysis* 24, 4. Chicago:Information Analyses. (ERIC Document Reproduction Service No. ED 338 445)
- Cizek, Gregory J. (1996). Grades: The Final Frontier in Assessment Reform. *NASSP [National Association of Secondary School Principals] Bulletin* 80, 584 (December): 103-110.
- Clay, M.M. (1993). An observation survey of early literacy achievement. Portsmouth, NH: Heinemann.
- Daniels, V.I. (1999). The assessment maze: Making instructional decisions about alternative assessment for students with disabilities. *Preventing School Failure* 43(4), 171-178.
- Eisner, E.W. (2001). What does it mean to say a school is doing well? *Phi Delta Kappan*, 82(5), 367-372.
- Elliott, J., Ysseldyke, J., Thurlow, M., & Erickson, R. (1998). What about assessment and accountability? *Teaching Exceptional Children*, 30(1), 20-27.
- Evers, William (2001, August 20). What Do Tests Tell Us? Hoover Daily Report. Hoover Institution. <<https://www.hoover.org/research/what-dotests-tell-us>>, as of May 18, 2022.
- Facts. (1999). [Online]. Available: <http://www.heinemann.com/info/08894f10.htm>
- FairTest Examiner. (1996). [Online]. Available: <http://fairtest.org/examarts/summer96/acntbty.htm>
- Farr, R., & Greene, B. (1993). Improving reading assessments: Understanding the social and political agenda for testing. *Educational Horizons*, 72, 20-27.

Foorman, Barbara, Jack M. Fletcher, and David J. Francis (2019). Chapter 3: Early Reading Assessment. Part Two: Constructive Uses of Tests. In Williamson Evers and Herbert Walberg (eds.), *Testing Student Learning, Evaluating Teacher Effectiveness*. Hoover Institution. <[https://www.hoover.org/sites/default/files/uploads/documents/0817929827\\_79.pdf](https://www.hoover.org/sites/default/files/uploads/documents/0817929827_79.pdf)>, as of May 18, 2022.

Freeman, D.J., Kuhs, T.M., Porter, A.C., Floden, R.E., Schmidt, W.H., & Schwillie, J.R. (1983). Do textbooks and tests define a natural curriculum in elementary school mathematics. *Elementary School Journal*, 83(5), 501-513,

Gay, G.H. (1990). Standardized tests: Irregularities in administering of tests affect test results. *Journal of Instructional Psychology*, 17(2), 11-18.

Hanushek, Eric A. (2006). Alternative School Policies and the Benefits of General Cognitive Skills. *Economics of Education Review* 25: 447-462. <<http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%202006%20EEduRev%2025%284%29.pdf>>, as of May 18, 2022.

Harris, K.H., & Longstreet, W.S. (1990). Alternative testing and the national agenda for control. *Clearing House*, 64(2), 90-96.

Herman, J.L. (1998). The state of performance assessments. *The School Administrator*, 55(11), 17-22.

Hess, F.M., & Brigham, F. (2000). None of the above. *American School Board Journal*, 187(1), 26-29.

Hess, Frederick M. (2022, March 24). When ‘Rigorous’ Courses Aren’t. *The Dispatch*. American Enterprise Institute. <<https://www.aei.org/op-eds/when-rigorous-courses-arent>>, as of May 18, 2022.

Holmes, Mark (1998). *The Reformation of Canada’s Schools*. McGill Queen's University Press.

Hughes, S. (1993). What is alternative/authentic assessment and how does it impact

Hurwitz, N., & Hurwitz, S. (2000). Tests that count. *American School Board Journal*, 187(1), 21-25.

Hurwitz, N., & Hurwitz, S. (2000). Tests that count. *American School Board Journal*, 187(1), 21-25.

James, J.C., & Tanner, C.K. (1993). Standardized testing of young children. *Journal of Research and Development in Education*, 26(3), 143-152.

Kober, Nancy (2002). What Tests Can and Cannot Tell Us. *TestTalk for Leaders 2* (October). Centre on Education Policy. <<https://web.archive.org/web/20030821213952/http://www.cep-dc.org/testing/testtalkoctober2002.pdf>>, as of May 18, 2022.

Loveless, Tom, Robert M. Costrell, and Larry Cuban (2005). *Test-Based Accountability: The Promise and the Perils*. Brookings Papers on Education Policy 8. Brookings Institution. <<https://nicspauill.files.wordpress.com/2011/04/loveless-2005-test-based-accountability.pdf>>, as of May 18, 2022.

McDaniel, J.P. (1997). Developing problem-solving skills in primary students (Doctoral dissertation, University of Cincinnati, 1997). *Dissertation Abstracts International*, 58(5A), 1593.

- Mitchell, K.J. (1997). What happens when school reform and accountability testing meet? *Theory Into Practice*, 36, 262-265.
- Neil, M. (1993). A better way to test. *Executive Educator*, 15(9), 24-27.
- Perrone, V. (1991). On standardized testing (Report No. RI-880-620-12). Chicago: Information Analyses. (ERIC Document Reproduction Service No. ED 338 445)
- Perrone, V. (1991). On standardized testing (Report No. RI-880-620-12).
- Phelps, Richard P. (1999). Why Testing Experts Hate Testing. *Fordham Report* 3, 1. Thomas B. Fordham Foundation. <<https://files.eric.ed.gov/fulltext/ED429089.pdf>>, as of May 18, 2022.
- Popham, W.J. (1998). Your school should not be evaluated by standardized test scores! [Online]. Available: <http://www.aasa.org/Issues/assessment8-26-98.htm>
- Popham, W.J. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, 56(6), 8-15.
- Ratcliff, N. (1995). The need for alternative techniques for assessing young children's emerging literacy skills. *Contemporary Education*, 66, 169-171.
- Rose, L.C., & Gallup, A.M. (2000). The 32<sup>nd</sup> annual Phi Delta Kappa/Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan*, 82(1), 41-58.
- Sacks, P. (2000). *Standardized minds: The high price of America's testing culture and what we can do to change it*. Cambridge, MA: Perseus.
- special education? *Educational Horizons*, 72(1), 28-35.
- Stiggins, R.J. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan*, 81(3), 191-198.
- Tuch, W.A. (1996). The impact of high stakes performance assessment on teachers and learning (Doctoral dissertation, Arizona State University, 1996). *Dissertation Abstracts International*, 57(10A), 4220.
- Valencia, S.W. (1997). Authentic classroom assessment of early reading: Alternatives to standardized tests. *Preventing School Failure*, 41(2), 63-70.