

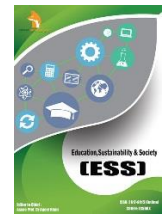


ZIBELINE INTERNATIONAL™

ISSN: 2617-9415 (Online)

CODEN: ESSDAX

# Education, Sustainability & Society (ESS)

DOI: <http://doi.org/10.26480/ess.02.2022.61.66>

## REVIEW ARTICLE

# TEST-BASED ACCOUNTABILITY: CAN TESTING CHILDREN TELL US ANYTHING SIGNIFICANT ABOUT THE PERFORMANCE OF SCHOOLS?

Kexin Wang\*

*International Education School, Guangzhou University of Technology and Business, Guangdong 510850, China**\*Corresponding author email: [kexinwang@gzgs.edu.cn](mailto:kexinwang@gzgs.edu.cn)*

*This is an open access journal distributed under the Creative Commons Attribution License CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited*

## ARTICLE DETAILS

### Article History:

Received 18 March 2022

Accepted 26 April 2022

Available online 27 April 2022

## ABSTRACT

This article would argue that test-based accountability has its rationality and is necessary for policy makers to control and improve public education. It could demonstrate to the public that public funds and resources are used reasonably and effectively. However, it indeed led to unintended issues among system-level, school-level and affected individuals in many ways. The main focus is on criticizing the effect and consequences of test-based accountability system by analyzing an act incorporating the system — “No Child Left Behind”. Besides, some background information as well as main legislations would be included in this article.

### KEYWORDS

Test-based Accountability, Childhood Education, Educational Testing and Assessment, No Child Left Behind

## 1. INTRODUCTION

Testing is a commonly used method to evaluate children's performance. It is also adopted by policy makers as an accountability tool to control education quality and improve public education (Supovitz, 2009). Through test-based methods and severe punishments, policy makers of various countries have incorporated the test-based accountability into their national education systems, arousing attentions and widespread debates of researchers all around the world. Some researchers argued that the implementation of test-based accountability system is beneficial to monitor and stimulate schools and teachers to make progress in children's academic performance (Dwyer, 2004). Others pointed out that testing for accountability, especially high stakes testing indeed improved pupils' performance and even helped with closing the achievement gap among disadvantaged subgroups, but the contribution of high stakes testing to the improvement was difficult to measure.

On the contrary, some studies showed that the performance of students has not improved after the accountability system introduced and even decreased due to mixed factors (Hanushek and Raymond, 2004). Furthermore, some studies found that testing for accountability purpose has brought about unintended consequences on individuals, especially on children (Hamilton et al., 2002; Hutching, 2005; Wenner and Kahan, 2009). Numerous studies have criticized the validity and effect of testing children to hold educators and schools accountable. In fact, considering its severe and invalid punishment standards as well as excessive pressure it imposed on individuals, some areas have suggested to amend some accountability measures or even resisted the enforcement of test-based accountability system.

I argue that test-based accountability has its rationality and is necessary for policy makers to control and improve public education. It could demonstrate to the public that public funds and resources are used reasonably and effectively. However, it indeed led to unintended issues among system-level, school-level and affected individuals in many ways. In this essay, I am going to focus on the effects and impacts of test-based accountability system on children. Firstly, four potential theories would be

reviewed to explain the rationality of testing for accountability. Then, I would review and focus on criticizing the effect and consequences of test-based accountability system by analyzing an act incorporating the system. The act is called No Child Left Behind and some background information as well as main legislations would be included in this part. Finally, I would summarize the impacts of test-based accountability measures through a theoretical framework mentioned in the first part and suggest that testing children for accountability purpose is doubtful and policy makers should be aware of the adverse effects as well as the applicability in different contexts when taking test-based accountability measures.

## 2. THE RATIONALITY OF TEST-BASED ACCOUNTABILITY

Test have different purposes and uses. If a test intends to hold individuals or institutions accountable for their performances and has risks attached to it, then it is commonly described as a test-based accountability system (Supovitz, 2009). As a strategy to improve public education, this system has gained widespread support by policy makers and observers in many countries like the U.S. (Hart and Teeter, 2004). In a report published in 2000, the National Alliance of Business even urged to impose more pressure on schools to improve their performances. They convinced that the introduction of such penalties and rewards in public education is essential to improve student performance and promote continuous school progress (National Business Coalition, 2000). Similarly, some researchers further noted that the test-based accountability system is based on the assumption that student performance could be improved by the inclusion of incentives in standardized tests (Hamilton et al., 2002). What exactly is the connection between test-based accountability system and student performance improvement?

By reviewing and summarizing the literature about test-based accountability and high stakes testing, Supovitz first proposed a theoretical framework underlying test-based accountability system (Supovitz, 2009). In brief, he argued that there are mainly four theories that support and encourage the employment of system: motivational theory, alignment theory, informational theory and the theory of symbolism. I'm going to review each of them below as these four theories

### Quick Response Code



### Access this article online

#### Website:

[www.educationsustainability.com](http://www.educationsustainability.com)

#### DOI:

[10.26480/ess.02.2022.61.66](https://doi.org/10.26480/ess.02.2022.61.66)

construct an essential framework which could be used to examine the effect and performance of test-based accountability programs.

The first and predominant theory is the theory of motivation, or the view that test-based accountability system could motivate progress. This theory argues that external rewards and sanctions related to high-risk testing could stimulate educators to improve their performance. A basic belief of this is that educators need external motivation and pressure to improve their teaching performance. McDonnell declared that for policy makers, external motivation is not only beneficial, but also necessary for changing behavior (McDonnell, 2005). If educators lack motivations to improve, incentives will solve the problem. However, for those practitioners who already have a strong sense of intrinsic responsibility for their own profession, the impact of extrinsic incentives on them is currently inconclusive. By conducting a meta-analysis of 94 experimental studies to explore the impact of external incentives on internal motivation, Cameron and Pierce found that rewards or sanctions do not reduce the intrinsic motivation of educators (Cameron and Pierce, 1994). By contrast, other researchers argued that specific external rewards often undermine internal motivation to a certain extent (Deci et al., 1999).

Alignment theory stated that if educators align the major components of the educational system surrounding schools including standards, curriculum, and assessments, the entire system will be generally improved with the mutual reinforcement of these components (Smith and O'Day, 1991; Chatterji, 2002). Fuhrman noted that high stakes test-based accountability has been used as a major component of alignment to promote coherence in the education system by policymakers (Fuhrman, 1993). It is worth noting that alignment is often considered as a concept of synchronization with surrounding systems, but some researchers believe that it could also be considered as the consistency between external accountability surrounding schools and internal accountability in schools (Ablemann and Elmore, 2004). They argued that the external responsibility system cannot completely affect the school's behavior, because the school has an internal accountability system composed of values, norms and expectations, which affects the school's work subtly. They also found that the concepts contained in the internal accountability system, such as who is responsible and responsible for what, greatly influenced educators' responses to external accountability. Therefore, when considering the alignment of external and internal accountability, the responses and influences of internal accountability also needs to be noted.

The third theory underlying the implement of test-based accountability system is information theory. It holds that student performance information such as test data is useful for school leaders and teachers to make decisions. More particular, providing such information to local educators and imposing incentives to them would guide improvement of decision making within classrooms and organizations. Generally, data has been regarded as an important tool for making decisions, addressing problems and enquiry-based learning (Torres and Preskill, 1999). In particular, test data is considered as valuable information for educators when making decisions about student and programs (Holcomb, 1999; Boudett et al., 2005). A group researchers stated that the rapid development of technology tools has made it easier to collect, summarize and manage a series of quantitative data (Stringfield et al., 2005). This undoubtedly simplifies the operation of test data and makes it more feasible for educators to acquire and use it.

The last potential theory contributing to the prevalence of high stakes test-based accountability is symbolism theory, which is relevant to public answerability. Supovitz pointed out that in this theory, the test-based accountability system is considered to convey important information and values to stakeholders, especially the public (Supovitz, 2009). The theory of symbolism is embodied in the concept of "public answerability"- the public has the right to know the resources they invested in are being exploited accountably, and public institutions have the obligation to maintain public trust. In public answerability theory, high risk assessment is a way to test the credibility of the current education system. Some researchers have observed the high stakes testing programs "symbolize order and control, desired educational outcomes, and traditional moral values" (Airasian, 1988).

In addition, Airasian pointed out that the political interest of politicians who promote high stakes testing programs is that these programs already have "sufficient motivation despite unexamined implementation" (Airasian, 1988). Therefore, as a symbolic of accountability system, high stakes testing could serve as a proof that public education is responsible and rigorous. These four potential theories are important as being a theoretical framework to analyze and examine the effect of test-based accountability system. In the following section, I am going to use this framework to review and criticize an influential education reform

incorporating test-based accountability system in the United States, which is called No Child Left Behind Act (NCLB).

### 3. NO CHILD LEFT BEHIND ACT AND THE IMPACT OF TEST-BASED ACCOUNTABILITY

In this section, I will start with reviewing the historical and legal background of test-based accountability programs in the United States to give a picture of the macro context of No Child Left Behind (NCLB) Act. Standardized testing has always been one of the characteristics of public education system in the United States. Nevertheless, such tests were basically limited to certain grades in elementary and middle schools until 1994. Two incidents during this period led to the large-scale promotion of the heavily test-focused policies. First, the large-scale implementation of testing is related to competition with other western countries. In 1983, an influential report entitled *A Nation at Risk: The Imperative for Educational Reform* was published by the US education department, warning that American school were failing as American graduates were lagging far behind in knowledge and skills necessary to promote economic development compared to students in other western countries (National Commission on Excellence in Education, 1983).

The publication of this report was viewed as a landmark and many reformers argued that it is necessary to strengthen the intensity of learning and testing to reduce the gap with the educational level of western nations. Second, the large-scale implementation of testing is related to the shift of political atmosphere. In particular, testing is likely to become a means of political gaming for political purpose instead of for promoting education. According to Dwyer, in the 1980s, the American national political temperament was extremely shift to the right (Dwyer, 2004). Since that time, an important part of the conservative agenda has been transferring education investment and expenditures from public school system to private schools and charter schools through voucher programs and tax credits. Under this circumstance, legal provisions to strengthen heavily testing of students are either to stimulate public schools make determined efforts to improve public education quality by imposing great pressure on them, or to embarrass educators within public education system by giving unrealistic goals on their schools, leading to public schools show inadequate and poor performance and thus policies could provide more public funding to private education with high quality.

In this context, demands for greater accountability in public education increased and the federal government came to achieve it by revising of the Elementary and Secondary Education Act (ESEA). As a mechanism used by federal legislators to adjust education policy, ESEA was firstly adopted during Johnson's administration in 1965. It initially provided disadvantaged children with financial support in education without additional conditions such as assessment. In 1994, Clinton Administration and congress revised the Act in making learning and assessment standards in reading and mathematics for those who received federal funding and requiring them to take three standardized tests in these two subjects at the K-12 stage. Although the standard was introduced, this revision in 1994 did not have a clear mandate for the specific aggregate outcomes reached by the states in the tests. In order to ensure the quality of students' learning, policy makers have made requirements for the annual progress of various school districts and individual schools. However, this mandate was too ambiguous so that the Education Department in the United States has hardly implemented it (Dwyer, 2004).

It is worth mentioning that the amend of the Act did affect schools and students with inadequate performance in some states. For example, at the school level, the amended Act increased supervision of state government for underperforming schools and provided more vouchers for students to transfer to other school. Besides, it stimulated students since they were not allowed to advance to upper grades or graduate with poor performances. In 2001, the ESEA was further amended in several important aspects when President Bush took office and signed the No Child Left Behind (NCLB) Act. This Act is viewed as a landmark event in American education history, and it has been viewed as 'the most sweeping plan to shake up public education in a generation, as well as the most intrusive federal intervention in local schools' (Dillon, 2004; Manna, 2006). Test-based accountability was integrated into the NCLB because increasing accountability in public education is the top priority of Bush's agenda (McBeath et al., 2007). The act aims to gain extensive improvements in student performance and reduce educational inequity among ethnic and disadvantaged groups (Supovitz, 2009).

The main measures could be summarized in three aspects. Firstly, the NCLB mandated states to introduce the test-based accountability system and promote it statewide. Students would be tested annually in reading and mathematics from grades 3 to 8 and a test would be administrated at one year in grades 10 to 12. In addition, science was first added to the

2007-2008 school year's assessment targets. Secondly, NCLB scheduled school annual progress standards including severe sanctions and required districts to report annual progress of different student subgroups. Schools were required to demonstrate "adequate yearly progress" (AYP) in proficiency targets of English literacy and mathematics and bring all students to the targets in by 2014. If a school failed to complete AYP for two consecutive years, it would be considered as an underperforming school, leading to severe penalties. Students in these underperforming schools could transfer to other public schools and the states where these schools were located would not receive federal funding.

Besides, the act also requires school districts to collect and report annual progress information of some student subgroups, including underprivileged students, students with disabilities, major ethnic and racial groups, and students with limited English proficiencies. Schools were required to take actions if any group of students fails to meet the AYP required goals. Thirdly, NCLB mandated states to certify teachers in core subjects through measurable goals for different groups of students. If teachers demonstrated high levels of teaching ability and were fully accredited in subjects such as mathematics and English literacy, then they would be viewed as highly qualified teachers. NCLB have aroused a widespread debate among the public and many researchers conducted research on its effects and multiple consequences. Some people indicated that the enforcement of the act did bring about broad improvements in student performance and reduce inequities between different groups (Supovitz, 2004).

For example, schools used to be identified as lowest category on test scores scale in Florida, North California, and Texas showed subsequent improvement on their performance. It suggested that test-based accountability system did get desired effect, especially on underperforming schools. On the contrary, some argued that NCLB was problematic as it brought a series of unintended issues such as leading to a narrowing curriculum, imposing too much pressure on school personnel, causing mental health problems on children, spending too much on the administration of testing etc. Some states and school districts, in fact, even refused to implement the act. According to the three major regulations contained in the act, I am going to analyze the advantages and disadvantages of test-based accountability measures incorporated into NCLB separately and then put forward my views on the impacts of these measures, especially the impacts on children. As I see it, the first mandate of NCLB promoted the learning of fundamental subjects like reading and math as well as expanded the coverage of students taking tests.

However, the frequency of testing was too high, and the curriculum would be narrowed due to neglect of learning other subjects. Learning subjects like reading, math and science are beneficial to the development of children's ability to acquire and analyze information, understanding as well as logical thinking. More specifically, as one of the most powerful learning methods, reading could not only help children acquire knowledge, skills and valuable experience, but also inspire them to get new ideas based on what they read (Stumm et al., 2011). Subjects like math and science are important tools to help children with developing logic suites (observation, questioning, experimentation, argumentation and evaluation) and stimulating creativity. Moreover, the emphasis on studying these basic subjects could construct a solid foundation for children to understand more complicated knowledge and concepts, which is helpful for their future study and life-long learning.

Therefore, The NCLB's focus on testing basic subjects is beneficial to the development of children's basic learning abilities and to preparing them better for more advanced study in the future. Meanwhile, the increased frequency of testing these courses brought about some undesirable consequences. First, the excessive testing is likely to damage the physical and mental health of children. Childhood is a vital period for children's physical growth, and it is necessary for them to have enough time playing and doing activities outdoors. Studies have shown that outdoor activities could make children close to nature and build a good connection with the natural environment from an early age. Playing and teaching activities in the outdoor space not only help with children's physical growth, but also stimulate children's curiosity and encourage problem-solving and negotiation skills, which are good for developing learning interest and inspiring teamwork awareness (Acar, 2013).

Roger argued that children feel more engaged and satisfied with their learning practice than studying within the four walls of a classroom (Roger, 2019). However, frequent testing makes it difficult to play and study outside as plenty of time and energy needs to be spent on preparing for written exams, making children tired and unhealthy. In addition, primary and secondary school periods are a crucial stage for students to establish self-esteem and self-awareness, and they are sensitive to external incentives. Growing in a learning environment focusing too much on tests,

children would suffer from failure' at younger ages and loose interest to attend school if they did not achieve expected test scores. More seriously, children's mental health is also greatly threatened by the huge pressure of testing. Wenner and Kahan claimed that the number of students with test anxiety in the United States is rising along with the increased attention and frequency of testing in schools (Wenner and Kahan, 2009). The same thing happened in the UK, where World Health Organization found that British pupils were more anxious about school assignments and tests compared with their counterparts in other European countries (World Health Organization, 2012).

This might be related to the increasingly fierce competition in Britain. Numerous studies have shown that stressed out and overanxious modes damage well-being of students, leading to worse study performance and even some academic avoidance behaviors (Wenner and Kahan, 2009). Research also suggested that the increase in the number of children diagnosed with ADHD (attention deficit hyperactivity disorder: a behavioral disorder that includes symptoms such as inattentiveness, hyperactivity and impulsiveness) in recent years is closely related to the high frequency of high stakes testing (Hutchings, 2015). It noted that children stay much longer in the classroom to take lessons or prepare for the tests. They are under tremendous psychological pressure, and this cannot be released through entertainment or outdoor activities as they have rare chance to move or play. Gradually, children feel uncomfortable with the school environment both physically and mentally, causing them to suffer from ADHD and other diseases.

Happier and healthier students are likely to have better performance in school. They should not be deprived of happiness and health in their childhood by the massive tests and excessive test-related pressure. Additionally, Hutchings argued that over-intensive tests may mislead children into thinking that the purpose of schooling is to pass exams and get a diploma at the end of study, since this is what schools tend to focus on (Hutchings, 2015). The test-centered approach was widely blamed by public, especially universities and employers. They criticized that the existing testing system is not qualified enough to prepare students with necessary life skills and other desired capabilities beyond school. Schools should develop students with ability such as innovative thinking, sense of responsibility and cooperation, independent problem-solving ability etc. Schools, educators and policy makers should be aware of these impacts on children's cognition as well as physical and mental health and administrate testing frequency to a moderate level so as to give more time and space for children to grow.

Furthermore, frequent testing in core subjects affected teachers in different aspects, which also had indirect impacts on children. For one thing, as the subject of accountability, teachers were under great pressure from test-based accountability measures. Public concerns about whether teachers make adequate efforts on their jobs have triggered demands for accountability (Linn, 2003). The great accountability pressure faced by schools passed down to principals and school leaders, but eventually imposed the substantial pressure on teachers. Due to accountability strategies school used, they had experienced considerable workload and unreasonable stress to achieve goals related to their students' performances (Hutchings, 2015). Some teachers noted that NCLB made them spend a lot of time on researching tests, helping pupils with the preparation of the massive tests and writing feedbacks, but this did not bring about expected benefits for students. By contrast, the increased workload and stress to improve test scores had damaged their relationships with pupils and this inevitably affected children's moods and performances.

Even worse, a group researchers found that some teachers felt too anxious and exhausted with their excessive workload after the enforcement of NCLB so that they required to transfer their positions to a grade with few tests or intended to leave the profession instead (Pedulla et al., 2003). In fact, most states were facing serious teacher shortages before NCLB enacted. The Act made it more difficult to retain and attract qualified teachers (Mcbeath et al., 2007). For another, the emphasis on core subjects testing and the associated accountability pressure have encouraged teachers to teach to the test and led to a narrowing curriculum. A group researchers studied the impact of test-based accountability on teachers and found that teachers were particularly concerned about information related to external high stakes tests and how that information was used to evaluate performance (Ingram et al., 2004). In order to improve performance, teachers have aligned the curriculum and teaching content with state testing standards, leading to teaching to the test and reducing teaching time for other subjects that do not take the test.

Accountability supporters argued that the reallocation of teaching practice is effective because teachers spend their time on teaching the more important and practical content (Hamilton et al., 2002). However, this

reallocation also takes up the instruction time of other disciplines such as history and art, and studying these disciplines is necessary for children to establish a complete knowledge system as well as comprehensively understand the world and themselves. In my view, this kind of reallocation, emphasizing the instruction of tested content while neglecting the need to learn non-tested knowledge, is a huge hidden educational cost. Without offering a comprehensive knowledge system, the narrowed curriculum is likely to 'narrow' children's future academic development and even life choices. Some teachers have noticed this but still argued that most of the schooling time was spent on preparing for tests such as practice and explain exam questions and they had no choice but to narrow the curriculum (Pedulla et al., 2003; Zastrow, 2004). Besides, teachers modelled their pedagogy and instruction to align with contents and skills in tests (Berry et al., 2004). I would argue that this test-oriented teaching concept defines good teaching as simply conveying tested content to children, which constrains teachers' practice in creative teaching as well as the development of children's multiple abilities beyond exam skills.

Schools are not exam factories, and children should not be taught to be the same narrow-minded testing tools according to teaching and testing standards specified by the factory. The second accountability regulation quantified assessment standards for schools and paid attention to the education quality of various student groups. However, penalties for failing to meet stringent standards are problematic, and special attention to a particular student group may lead to reverse effect. Compared with previous policies, the NCLB's requirements for the improvement of school performance are more specific, promoting schools to take measures to improve performance under annual accountability pressure. The punishment for failing to meet the progress target in two consecutive years demonstrates that underperforming schools still have opportunities to adjust their strategies. These new quantitative regulations also strengthen the implementation and supervision of NCLB Act. Additionally, it is worth to mention that NCLB also pays attention to the academic performance and education quality of student groups such as ethnic groups and underprivileged students.

This is conducive to closing the achievement gap among different groups and promote educational equality. However, it seems to me that the punishments of test-based accountability are too severe. First, the education level of each state is different due to historical, geographical, and economic factors: some states have good social environments for developing education while others might be trapped in massive education issues left over from history. Meanwhile, testing forms and standards vary from state to state. It is too difficult to make all states to adopt an interchangeable proficiency standard like NAEP, not to mention to require all children to achieve the defined proficiency standard. Therefore, I would argue that it is unrealistic to require long-term underperforming schools to achieve an over ambitious goal and imposing same punishments on them is unfair.

As a result, some states like Pennsylvania have resisted the implementation of NCLB (Dwyer, 2004). Policy makers should be aware of the diversity and complexity of education levels in different regions and schools when it comes to sanctions. Further, McGuinn indicated that a few states have requested to amend some accountability regulations under NCLB like adding value-added approaches (McGuinn, 2006). These states argued that NCLB should recognize the progress they made instead of labelling them "low performing" schools and punishing them continuously. While some people opposed it for the amend would undermine the integrity of the law, I convinced that the amend is necessary. It would be better to adopt value-added approach to assess school performance and incentives like granting awards or education funding could be used to stimulate school improvement rather than taking too much punishment measures. I suggest that punishment is not the purpose of accountability. Instead, it makes more sense to identify poorly performing schools and then provide them with appropriate guidance and assistance.

Moreover, research shows that school education is merely accountable for a small part of children's academic performance and the improvement cannot be separated from the careful instruction and warm company from their families (Hutchings, 2015). It is unfair to blame schools for the poor performance of children and parents should take responsibility for their children's education. According to Sunderman NCLB's extra attention to subgroups is conducive to studying the reasons behind the performance inequality among them, and further close the performance gap (Sunderman, 2008). Nevertheless, research shows that after expanding test-based accountability in states, the change in achievement gap between majority and minority student groups is mixed. For example, while there is a shrink in Hispanic-white gap, the gap between black and white did not reduce (Hanushek and Raymond, 2004). Moreover, a researcher argued that NCLB unintentionally increased the dropout rate of minority students (Ryan, 2004).

This attributed to the exclusion of long-term underperforming students in each subgroup since states and schools prefer to report better performances, leading to an increase in the number of dropouts, especially among high school students. Besides, just as I mentioned earlier, education outside school is equally important for children's academic performance. Pupils from wealthy families could enjoy additional resources such as extra-curriculum classes to assist them to improve attainment but disadvantaged pupils lack such opportunities. Therefore, it is insufficient to take accountability measures on schools or school districts in order to narrow the performance gap among subgroup students, and policy makers should supervise the performance report of subgroups strictly. In the third legislation, the certification and assessment of teachers stimulates teachers' enthusiasm for improving the teaching level, which helps to ensure the quality of teaching and further promotes the improvement of student learning quality.

However, it is unreasonable to use student achievement to evaluate the teachers' teaching level and their donation in teaching. Firstly, it's too absolute to certify teachers' competence and accountability by setting quantitative goals and testing their students. As there are too many factors affecting students' achievements, such as the contribution of family education or whether students have taken after-school tutoring classes. Besides, it seems to be a long-term process for students to improve achievement and their efforts as well as teachers' dedication may not be paid off in a short time. It is therefore inappropriate to hold teachers accountable only by using this way. In addition, it's hard to assure that the teacher would continue to be responsible after being certified. What if they become satisfied with what they already owned in their professions and start to neglect the research of new advanced teaching contents or methods? Policy makers should consider regular assessments of the teaching quality of certified teacher. More importantly, test results could neither fully reflect the teacher's teaching level nor reveal students' understanding and cognition of the learning content. Hutchings argued that the test results could not truly reflect the children's real understanding of what they learned (Hutchings, 2015).

As I see it, those children who do poorly on the test do not necessarily fail to understand the relevant learning content, instead, they are just unfamiliar with the test questions or the format of test paper. Although there is evidence that high stakes tests did correlate positively with test scores, the results of other low stakes tests indicated that student performance has not improved in these subjects (Amrein and Berliner, 2002). Similarly, group researchers research found that students' performance in the GCSE test increased by 15% between 2006 and 2012, but their performance in other tests during this period, such as PISA, did not increase (Wheater et al., 2014). This shows that good test results merely indicate the sufficient preparation of students for the test but does not mean that they have had a deeper understanding and thinking about the knowledge tested. In addition, Hutchings also pointed out that pupils' test results do not represent their learning ability (Hutchings, 2015).

Because these grades are based on English literacy or math tests, which cannot be used to evaluate children's studying potential in art, music or other subjects. Thus, since the test itself is defective and it cannot fully reflect the real situation of student learning, it is invalid to use it to assess teachers' accountability. Moreover, according to many teachers in low performing school districts could not be certified by NCLB due to their inadequate teaching and insufficient local teacher training, and this made it more difficult to retain teachers or attract high-quality talents to teach in those areas (Mcbeath et al., 2007). To meet NCLB's requirements for teachers' quality, the government need to invest a lot of resources in teacher welfare and career development. Other issues, such as rising test costs, false reporting of data by schools and states in response to pressures of accountability, would not be discussed here because they are not relevant to the impacts of test-based accountability on children.

#### 4. WHAT WE LEARNED FROM NCLB

By using the theoretical framework mentioned previously, I would summarize the impacts and effects of the test-based accountability measures used in NCLB. From the perspective of motivation theory: Test-based accountability measures have motivated the improvement of schools and teachers, which is beneficial to the students' academic progress. However, this improvement is superficial and incomplete: The strategies adopted to cope with high pressure and severe punishment of accountability are test-centered, which means what schools and teachers did is to make children acquire satisfied' test scores rather than help them develop deeper understanding of knowledge and skills. In this process, the learning demands of other subjects were also ignored, leading to a narrowing curriculum. In addition, the motivation of the improvement is based on severe external sanctions, which is harmful to the mental health of educators and students as well as their relationships.

From the perspective of alignment theory: Test-based accountability encouraged schools and teachers to align standards, curriculum and assessments, promoting a more integrated education system. However, it is difficult to say whether tests are aligned to standards and curriculums or vice versa. Further, research shows that teachers are more affected by testing rather than standards. Additionally, it is unrealistic to align academic contents and standards of all states to national requirements regardless of the variety and reality of contexts. From the perspective of information theory: The information acquired from tests could be used by policy makers as a tool to monitor system-level academic performance, but it is not sufficient to judge the performances of individuals like teachers and students. The test scores merely reflect students' preparation of tests but may not represent they are 'qualified' in understanding and mastering the knowledge and skills. Thus, only using this information to evaluate the performance and contributions of teachers is doubtful. From the perspective of symbolism theory: Test-based accountability measures help policy makers monitor the quality of education and show the public that taxes and public resources are being used reasonably and effectively. At the same time, policy makers should develop plans to address rising test costs and strengthen the supervision of the authenticity of testing process and reports on districts and schools.

## 5. CONCLUSION

Testing children could provide information about school performance but it is not an appropriate and accurate approach. There are numerous factors influencing children's performance such as family education and extra-curriculum classes, and it is difficult to figure out how much schools or teachers have contributed. Besides, it may need a long time for the efforts schools and educators made to pay off. Moreover, children's testing results do not effectively reflect their ability to understand knowledge and master skills, instead, these are only evidences that children have prepared for testing. Therefore, it is doubtful to merely use children's testing result to evaluate responsibility and performance of schools and teacher.

Used as an accountability method, testing children stimulated school districts and schools to make progress in order to improve children's performance, but it also had serious unintended impacts on educators and children: the narrowing curriculum, teaching to the test, the excessive workload and high pressure on teachers, the damaging physical and mental health of children, the increasingly lower performance of disadvantaged students and so on. Testing children, especially using high stakes testing, is not beneficial to the long-term development of children's learning and well-being. Schools should not become 'exam factories and children are not exam tools developed according to an identical standard. Schools should pay attention to teaching other necessary life and social skills apart from the testing subjects, which could help children to adapt to society and live a happy life when they graduate from school.

In general, test-based accountability system has rationality and its implementation have brought about an overall improvement in children's academic performance. However, some accountability measures are too strict and absolute without concerning the variety of schools and areas as well as characteristics of children. The aim of adopting test-accountability measures is not to punish pupils, teacher and schools. Instead, it aims to identify those who need helps and provide them with appropriate support and assistance. Policy makers who decided the testing standard and content should be aware of that the improvement of public education has a long way to go and they should give more patient to the development of education as well as consider more about the influences of policies they made on individuals, especially on children.

## REFERENCES

- Acar, H., 2014. Learning Environments for Children in Outdoor Spaces. *Procedia - Social and Behavioral Sciences*, 141, Pp. 846-853.
- Airasian, P., 1988. Symbolic Validation: The Case of State-Mandated, High-Stakes Testing. *Educational Evaluation and Policy Analysis*, 10 (4), Pp. 301-313.
- Amrein, A., Berliner, D., 2002. High-Stakes Testing & Student Learning. *education policy analysis archives*, 10, Pp. 18.
- Boudett, K., City, E., Murnane, R., 2005. *Data Wise: A Step-by-Step Guide to Using Assessment Results to Improve Teaching and Learning*. Revised and Expanded Edition. Harvard Education Press, Pp. 116-122.
- Cameron, J., Pierce, W., 1994. Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis. *Review of Educational Research*, 64 (3), Pp. 363-423.
- Chatterji, M., 2002. Models and Methods for Examining Standards-Based Reforms and Accountability Initiatives: Have the Tools of Inquiry Answered Pressing Questions on Improving Schools? *Review of Educational Research*, 72 (3), Pp. 345-386.
- Deci, E., Koestner, R., Ryan, R., 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125 (6), Pp. 627-668.
- Dwyer, J., 2004. *School Accountability and 'High Stakes' Testing*. Faculty Publications. William & Mary Law School.
- Etienne, G. 2016. All children should learn math and science. <https://www.globalpartnership.org/blog/all-children-should-learn-math-and-science>.
- Hamilton, L., Klein, S., Stecher, B., 2002. *Making sense of test-based accountability in education*. Santa Monica, CA: Rand.
- Hanushek, E., Raymond, M., 2004. Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24 (2), Pp. 297-327.
- Hart, P., Teeter, R., 2009. Equity and adequacy: Americans speak on public school funding: A national opinion survey conducted for the educational testing service.
- Holcomb, E., 1999. Getting Excited about Data. [Place of publication not identified]: Distributed by ERIC Clearinghouse.
- Hutchings, M., 2015. *Exam factories: The impact of accountability measures on children and young people* Research commissioned by the National Union of Teachers. London Metropolitan University.
- Improving Performance: Competition in American Public Education. 2000. <http://www.nab.com/PDF/comp.pdf>.
- Ingram, D., Louis, K., Schroeder, R., 2004. Accountability Policies and Teacher Decision Making: Barriers to the Use of Data to Improve Practice. *Teachers College Record*, 106 (6), Pp. 1258-1287.
- Linn, R., 2003. *Accountability*. Los Angeles, Calif.: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.
- McBeath, G., Reyes, M., Ehrlander, M., 2007. *Education reform in the American states*. Charlotte, N.C.: Information Age Pub.
- McDonnell, L., 2005. Assessment and Accountability from the Policymaker's Perspective. *Yearbook of the National Society for the Study of Education*, 104 (2), Pp. 35-54.
- McGuinn, P., 2006. *No Child Left Behind and the transformation of federal education policy*. Lawrence, Kan.: Univ. Press of Kansas.
- National Foundation for Educational Research 2020. *Achievement of 15-Year-Olds in England: PISA 2012 National Report*. OECD Programme for International Student Assessment. Department for Education.
- Preskill, H., Torres, R., 1999. *Evaluative inquiry for learning in organizations*. Thousand Oaks, Calif.: Sage Publications.
- Ryan, J., 2003. The Perverse Incentives of the No Child Left Behind Act. *SSRN Electronic Journal*.
- Simon, C., 2007. School's In: Federalism and the National Education Agendaby Paul Manna. *Political Science Quarterly*, 122 (1), Pp. 177-178.
- Smith, M., O'Day, J., 1991. Systemic school reform. *Journal of Education Policy*, 5 (5), pp. 233-267.
- Sunderman, G., 2008. Holding NCLB accountable: achieving accountability, equity & school reform. *Choice Reviews Online*, 46 (1).

