

Perceptions of Test Score Pollution Stemming From COVID-19 and State Testing: An Exploratory Case Study

Elif Kalemdaroglu-Wheeler¹

Capella Educational Services, LLC., Midlothian, Virginia, USA

ABSTRACT

The purpose of this qualitative exploratory case study was to explore teachers' and administrators' perceptions of test score pollution deriving from COVID-19-related issues that may affect students' test scores on state-mandated standardized tests for grades six through 12 in a state along the Atlantic Coast of the United States. Four research questions were devised to investigate participants' perceptions of factors stemming from COVID-19-related issues that may alter students' performance on state standardized tests, commonly referred to as test score pollution. The conceptual framework centered around Heisenberg's Uncertainty Principle, Campbell's Law, and Kane's Validation Model. A purposeful stratified sampling method was utilized for participant inclusion criteria. The study sample included four middle school teachers in grades six through eight, two high school teachers, two middle school administrators, and two high school administrators. The data collection method included semi-structured interviews. Data analysis was conducted in unison with memoing, member checking, In Vivo, Descriptive, Focused, Axial, and Process Coding methods. The following themes emerged through data analysis: (a) inadequately preparing students for state standardized tests, (b) de-valuing of education, (c) understanding students' emotional well-being, (d) providing data-driven support, (e) questioning test validity, and (f) recognizing magnified disparities among students.

KEYWORDS: Campbell's Law, COVID-19, Heisenberg's Uncertainty Principle, Kane's Validation Model, state standardized tests, test score pollution.

The use of standardized testing to measure student progress has been in place since the Elementary and Secondary Act of 1965. Federal funding is now tied to states' standardized tests through laws such as the No Child Left Behind (NCLB) Act of 2001 and Every Student Succeeds Act (ESSA) of 2015. Each state administers these tests to assess academic progress for accountability measures and provide parents with supplemental educational services options (U.S. Department of Education, 2004). State standardized tests play a significant role in decisions about student placement in certain classes, teacher professional development activities, accountability of teachers, program efficiency, allocation of resources, curricula efficiency, and bringing awareness to the public regarding educational issues (Abrams et al., 2003; Cuban, 1991; Ghaicha, 2016). The scores also impact the real estate market as families decide residency choices based on schools' performance on state standardized tests. Higher-quality school districts and schools, which show

¹ Corresponding Author: She is studying Education Law at St. Mary's University School of Law and pursuing further educational research as a Fellow in Residence with the University of Phoenix.
E-Mail: capellaeducationservices@gmail.com

higher student achievement on state standardized tests, tend to have higher housing values (He & Giuliano, 2018; Mast, 2018).

The COVID-19-related school closures, changes in teaching modality, and widespread impact on education have become an area of concern since March 2020. After the initial school closures, an increased interest in research related to online education, teacher professional development, the economic impact of school closures, and the quality of education emerged. Additionally, concerns regarding the effects of COVID-19-related issues on students' state standardized test scores gained attention (Middleton, 2020). Test score pollution, a term coined by Thomas Haladyna at the start of the accountability measures in the early 1990s, is described as test preparation and test administration practices that may impact test scores and performance (Haladyna et al., 1991). Factors such as the testing environment, student anxiety, and students' family issues prompt further concern during tests, increased distractibility, and negatively impact test scores (Haladyna et al., 1991). Test score pollution is likely to occur indirectly due to pressures placed on standardized tests (Alimorad, 2014; Chalak & Tavakoli, 2010; Neemati et al., 2014; Pearson et al., 2001). Haladyna et al.'s (1991) seminal work has documented 21 sources of test score pollution factors in three different categories: test preparation, situational factors, and external factors, which are beyond the control of schools.

It is uncertain whether the effects of COVID-19 on test-taking have impacted state standardized test scores and whether these tests can be relied upon to measure student progress accurately. The reliability and validity of these tests remain a topic of discussion in education. Popham (2020) identified three factors contributing to validity: the alignment of test content with what was taught, the response process of the test takers, and test construction. Lim (2021) added to this list, noting that test content, response process, internal structure, relation to other variables, and testing consequences all contribute to test validity. However, the U.S. Department of Education has stated that the validity of K-12 standardized tests only considers whether the test content aligns with state standards (Tannenbaum & Kane, 2019). Exposing educators' beliefs about the distortion of test scores based on test score pollution factors deriving from COVID-19-related issues and effects on state standardized tests may prompt state agencies and the federal government to re-examine the current state standardized tests in K-12, the use of standardized test scores and the requirements for student achievement progress.

This study explored the perceptions of ten K-12 teachers and administrations about test score pollution stemming from COVID-19 factors and possible impacts on state standardized tests. This study was guided by questions: What are middle/high school teachers' perceptions of test score pollution stemming from COVID-19? What are middle/high school administrators' perceptions of test score pollution stemming from COVID-19? Two sets of interview questions were prepared for the semi-structured interviews. The questions for teachers were slightly distinct from the questions for school administrators. Teacher questions incorporated subject matter regarding teaching methods, instructional design based on beliefs about state standardized tests, and response to a study on external factors' impact on student test scores. Administrators' questions were oriented towards supporting teachers and students concerning state standardized tests, utilization of test scores, and addressing concerns regarding state standardized tests. All participants were asked about their general perceptions of state standardized tests, their roles in the education field, and additional comments they might have liked to add on the subject.

Previous studies on standardized tests and accountability often refer to Heisenberg's Uncertainty Principle and Campbell's Law (Koretz, 2017, 2019; Madaus, 2009; Sidorkin, 2016; Walt & Madaus, 1986). Kane's Validation Model also provides helpful guidelines for establishing a reliable and valid framework around standardized tests (Longmuir et al., 2021; Tannenbaum & Kane, 2019). The conceptual framework is based on these three fundamental notions and guided

E. KALEMDAROGLU-WHEELER

this study by demonstrating the use of test scores in decision-making, addressing validity concerns related to current test scores, and the need for change in response to pandemic-related challenges. Additionally, these concepts highlight the unpredictability of tests and student performances due to external factors like instructional delivery, test preparation, students' external environment, and test administration conditions.

Conceptual Framework

Heisenberg's Uncertainty Principle states that accurately determining the trajectory of particles is impossible (Heisenberg, 1927). This principle extends to education as well. Successful initiatives in schools and communities face uncertain outcomes due to constant changes, as pointed out by Glickman et al. (2018). The COVID-19 pandemic has further complicated matters, making it difficult to predict the outcomes of state standardized tests. Bridges and Woolcock (2022) have argued that haphazard data collection can lead to unintended consequences, and data-driven policies should be approached with caution. An example of such a policy is the accountability system in K-12 education, as noted by Bai et al. (2020). These policies involve various entities, such as teacher-learner interactions, the taught and tested curriculum, and the social, cultural, and political contexts, which can further complicate matters. McIntyre-Mills (2021) emphasized the importance of examining the cause-effect cycle in all aspects of life, including the relationship between students and their learning environment, curriculum, and health concerns. Any unforeseen changes in these factors can lead to uncertainty and affect student performance in state standardized tests.

Campbell (1979) warns that social indicators heavily relied upon in decision-making can easily become distorted and corrupted. Braganza (2022) provides clear examples of this phenomenon in social programs. Proxy measures are commonly used to evaluate specific objectives in various fields. For example, standardized test scores are used as a proxy for measuring students' knowledge and skills, and teachers' effectiveness is assessed based on their students' performance. Similarly, vote counts and publicity rates are utilized in politics to determine voter representation. Lastly, profit records may serve as proxy measures in the business market. However, Campbell's Law suggests that proxy measures can be distorted due to society's emphasis on scores. The COVID-19 pandemic provides a recent applicable scenario of Campbell's Law. Best (2021) argues that manipulating statistical measures by government organizations to avoid criticism or gain recognition is a prime example of Campbell's Law in action.

Recent studies by Emler et al. (2019) and Koretz (2017) have uncovered the compromised nature of standardized tests due to rampant corruption, including cheating, providing answers, and exclusion of low-achieving students from tests. Even data reporting under Campbell's Law is distorted by selective evidence gathering (Best, 2021). Furthermore, Hess and McShane (2018) observed that educators prioritize tested content, move strong teachers to teach tested courses, and focus on "bubble students" sitting below the cut-score level. Thus, emphasizing Campbell's (1979) contention that testing becomes the focus of teaching and learning, losing its value as a progress indicator and distorting education. The use of Campbell's Law on standardized tests perpetuates a testing culture and raises concerns about score validity and reliability. Decisions regarding student placement, graduation, resources, and teacher accountability based on test scores can result in score interpretation uncertainty (Byrd & Varga, 2018; Cunningham, 2019). Despite this, standardized testing remains ingrained in American culture. However, test preparation and coaching practices can contribute to score inflation, particularly in schools with disadvantaged populations (Koretz, 2019). Ultimately, the emphasis on standardized test scores can distort the educational process and raise questions about the intention and use of these measures.

Validity is the appropriateness and credibility of the interpretations and uses of test results (Kane & Wools, 2020). Kane (2000) indicated a clear justification of scores' intended use must be provided. A three-step systematic approach should be taken to ensure the validity of test scores: a defined use of the test results; assessments designed to serve the intended uses of the test results; and a checks-and-balances system to ensure the designed assessment supports the intended uses (Kane & Wools, 2020). Furthermore, test score interpretations require five inferences: (a) evaluation, (b) generalization, (c) extrapolation, (d) explanation, and (e) decision (Kane, 2000). Each of these inferences requires different types of supporting evidence, and it is beneficial to combine test results with other performance tasks and classwork to gain a more complete understanding of a student's abilities.

Assessments serve different purposes, such as providing feedback on students' performance, evaluating their ability to perform specific tasks, and identifying their strengths and weaknesses (Kane & Wools, 2020). However, when applied to larger groups of students, like standardized tests, they evaluate overall achievement in specific areas. Hence, drawing conclusions from a single test is not ideal for assessing student performance (Kane & Wools, 2020). Due to the impacts of COVID-19, students and teachers may have experienced post-traumatic stress (Huber & Helm, 2020; Miller & Hui, 2022). With a disrupted routine for 57 million students (Donohue & Miller, 2020), some struggled to meet their basic needs (Ansorger, 2021). The effectiveness of teaching virtual classes with a condensed curriculum varied based on teachers' expertise. Given the variances in educational conditions and Kane's validation description, it is questionable if state standardized tests' intended measures and scores obtained during the pandemic are accurate.

Based on Heisenberg's Uncertainty Principle, there are many uncertainties regarding test score pollution sources and the impact on education (Amrein & Berliner, 2003). To prevent corruption and distortion of test scores, Campbell (1979) called for multiple measures of performance. Additionally, Campbell's Law suggests that society may have accepted the constant cycle of state standardized tests. The degree to which interpretations are supported by evidence is defined by Kane's test validity (Kane, 2000). Under COVID-19 conditions, Kane's Validation Model challenges the intended uses of current test scores. Amrein-Beardsley and Barnett (2012) urged education policymakers to move towards a holistic measurement system to address the limitations and utility of standardized tests.

Methodology

According to Nassaji (2020), qualitative studies are conducted in natural settings and use inductive processing to analyze data sets and identify emerging themes. Unlike quantitative studies, qualitative studies require a smaller number of participants, often chosen through purposeful sampling. Data sets are collected through interviews, observation, and review of archival data to establish data triangulation. Stake (1995) emphasized that qualitative research focuses on a holistic study of a phenomenon through interpretations of interviews, observations, and archival records and formulating themes and patterns. Therefore, qualitative studies are subjective and not intended to generalize the results obtained (Stake, 1995; Yin, 2003).

Merriam and Grenier (2019) assert that qualitative research is essential for providing a comprehensive understanding of social sciences by examining personal experiences and perspectives. Qualitative studies serve various functions, including contextual, explanatory, evaluative, and generative. This qualitative exploratory case study serves a contextual purpose by exploring perceptions and answering "what" questions. It describes the nature of a naturally occurring phenomenon and examines the reasons and connections within it. Furthermore, it generates new theories and concepts within the phenomenon under investigation (Yin, 2016).

Qualitative studies focus on understanding the dynamics within a given setting, as the researcher is the primary data collection instrument. Merriam and Tisdell (2016) noted that the inductive data analysis of qualitative studies provides an ornately descriptive end product and are characterized by rich and descriptive data, researchers as the primary instrument, and the search for meaning. Yin (2003) posed social science research is guided by the problem of the research rather than the methodology. Qualitative research enables the researcher to seek in-depth knowledge of the phenomenon. The problem is COVID-19-induced changes in teaching modality, delivery, and curriculum content may result in test score pollution and affect state standardized test scores (Middleton, 2020). Therefore, the purpose of this qualitative exploratory case study was to explore teachers’ and administrators’ perceptions of COVID-19-related test score pollution in state standardized tests in grades six through 12 in an Atlantic Coast state. As such, conducting a qualitative study allows research designs aligned to answer the research questions and permits smaller sample sizes to obtain in-depth knowledge from the participants.

A stratified sampling method was used to ensure proper representation of teachers and administrators from middle and high school levels. Data collection was completed through semi-structured interviews. Ten interviews were conducted for this research. The continual data analysis and collection of new data provided different lenses for interpreting the data, assigning codes, and forming categories and themes. Data collection and analysis was a simultaneous process requiring adaptability as needed. The interview questions were formulated based on the concepts and theories outlined in the conceptual framework, which served as a guide throughout the study. During analysis, the emergent themes were thoroughly compared to these concepts.

Table 1

Comparative Analysis: High-Frequency Codes

| Middle School Teachers | High School Teachers | Middle School Administrators | High School Administrators |
|------------------------|----------------------|------------------------------|----------------------------|
| Test Preparation | Test Preparation | Data drive | Test Preparation |
| Cultural shift | Academic gap | Academic gap | Academic gap |
| Lowered expectations | Lowered expectations | Cultural shift | Support |
| Academic gap | Emotional well-being | Support | Uses of test scores |
| Same tests | Virtual learning | Emotional well-being | Same tests |
| Emotional well-being | Cultural shift | Lowered expectations | Magnified disparities |

At the outset, Microsoft Excel was used for manual coding to ensure precision. The NVivo software was then utilized to analyze file references, visualization, and coding matrix for the subgroups. This process guaranteed accurate manual coding. The coding cycle was broken into two cycles. During the first cycle, the In Vivo Coding method was used to capture key phrases provided by the participants. However, this method alone limited perspectives on data analysis. The In Vivo codes were reviewed and rephrased using the Descriptive method. Each code was assigned a noun to summarize the participants' comments (Saldaña, 2016). This method laid the foundation for creating categories and themes. Using a “splitter” approach, both In Vivo and descriptive coding

methods were analyzed line by line. As a result, each transcription yielded between 70 and 90 codes.

Focused and axial coding techniques were used in the second coding cycle to create categories. Focused coding involved collecting the most frequently used codes from the descriptive coding phase to compare data frequency among the participants. Axial coding consisted of grouping similar codes from focused coding to create fewer codes and categories (Saldaña, 2016). During this phase, the analytic memo was heavily used to identify dominant codes and categories, which was the first step toward theme formation. The codes obtained from axial coding were changed to actionable words in the process coding method. Table 1 represents a comparative analysis of high-frequency codes by the participants.

Demographics

Six participants were teachers; four were from middle school, and two were from high school. The remaining participants were school and district-level administrators. Each potential participant was screened to meet the criteria of teachers and administrators, grades six to 12, at least 25 years old, and a minimum of five years of experience in classroom teaching or administration in this particular Atlantic Coast state (see Table 2). The teacher group consisted of three males and three females. The administrator group had one male. Six of the participants were in the age range of 40-49, and four were in the 50-59 range. Three participants identified as Black or African American for race or ethnicity. One participant identified as Multiracial or Biracial, and six identified as White or Caucasian. The most common years of experience for administrators were 9-15. This experience was in addition to the years of teaching experience. The most significant amount of experience was 16-20. Pseudonyms were assigned to each participant to ensure confidentiality.

Table 2

Participant Demographics

| Name | Age Group | Race | Gender | Current Teaching/ Admin. Grade Level | Years of Teaching Experience | Years Administrative Experience | of |
|----------------|-----------|--------|--------|--|------------------------------|---------------------------------|----|
| Cardinals | 40-49 | White | Male | 9-12 Admin | 5-8 | 9-15 | |
| Bass Pro | 40-49 | White | Male | 6-8 Teacher | 16-20 | 0 | |
| Football Fan | 50-59 | White | Female | 9-12 Teacher | 30+ | 0 | |
| Boxer | 40-49 | White | Male | 9-12 Teacher | 9-15 | 0 | |
| Dolphins | 40-49 | White | Female | 6-8 Teacher | 16-20 | 0 | |
| Flogging Molly | 50-59 | White | Male | 6-8 Teacher | 30+ | 0 | |
| Angels | 50-59 | Black | Female | 9-12 Admin | 9-15 | 21-29 | |
| Bestie | 40-49 | Black | Female | 6-8 Admin | 16-20 | 5-8 | |
| Amiga | 50-59 | Multi- | Female | 6-8 Teacher | 9-15 | 0 | |
| Chy | 40-49 | Black | Female | 6-8 Admin | 16-20 | 9-15 | |

Findings

All participants were asked about their familiarity with “test score pollution.” Regardless of the participants’ positions and years of experience, no participant was familiar with the term. Upon sharing the definition of the term, all participants indicated the relevance of the term to their position and the impact of the factors described on test scores. Participants could not provide in-depth information on the different factors of test score pollution.

Six themes emerged with the coding process. The themes were inadequately preparing students for state standardized tests, devaluing of education, understanding students’ emotional well-being, providing data-driven support, questioning test validity, and recognizing magnified disparities among students. Losing academic progress was an overarching theme across all subgroups. Although losing academic progress is a factor that may impact student test scores on state standardized tests, the theme is a broad term to be considered as an external factor. Furthermore, external factors such as home life, emotional well-being, cultural shift, absenteeism, lack of support, economic disparities between students, and lack of teachers may have contributed to the perceived academic and learning loss. The analysis of this common theme was outside the scope and purpose of this study. The themes relevant to test score pollution factors were presented to that end. Indicative of the literature review, the emergent themes were reflective examples of Heisenberg’s Uncertainty Principle, Campbell’s Law, and Kane’s Validation model, further supported by a comprehensive literature review of the results.

Theme 1: Inadequately Preparing Students for State Standardized Tests

Teacher participants noted they embed continuous review into the pacing guides to ensure the repetition of content throughout the year. Test blueprints provided by the state education department are used to align instruction and pacing. Bass Pro referred to test blueprints as the “bedrock and Bible” of their teaching as teachers spend more time on content areas covered on the tests. All teachers commented that continuous reviews include examination of previously released state standardized test questions. This process allowed students to get familiar with test structures, decoding test questions, the cognitive level of questions, and learning test-taking strategies.

Due to school closures and changes to the learning environment since the pandemic, teachers agreed that teaching test-taking strategies and preparing for state standardized tests have become more challenging. Dolphins said, “Amount of material covered on tests is unrealistic; there are differences in exposure to materials and gaps in learning.” Teachers noted that “bridging the gap” and teaching the current content took time away from adequately preparing students for the tests. Flogging Molly stated, “We are trying to backtrack leading up to the test plus our material.”

Theme 2: Devaluing of Education

The theme of devaluing of education became apparent through the initial codes, such as cultural shift, absenteeism, mind shift, opt-out, lowered expectations, student indifference, and home life. Participants described students as apathetic toward themselves, their peers, and teachers. Regarding state tests, Bass Pro noted that students do not take them seriously, while Boxer mentioned a prevalent “I don’t care” attitude among the students. Furthermore, Amiga stated that there has been a cultural shift regarding the value of education, and students no longer view teachers as authority figures. According to the participants, this shift was further influenced by parents’ attitudes toward education and their lack of support.

Theme 3: Understanding Students' Emotional Well-being

The importance of paying attention to students' emotional well-being and recognizing the “whole child” was another theme. Teachers are unaware of the personal trauma students might have incurred during the pandemic. One of the emotional challenges mentioned was students' self-confidence. Football Fan stated, “We have to build the students' confidence up” and “Students have a defeated attitude before tests.” Cardinals discussed the lack of support at home for some students. Furthermore, administrators noted that some students suffer from anxiety and depression, and some students' transition from home to school has not been successful. Lastly, the school staff might not be aware of the challenges students face outside of the school environment.

Theme 4: Providing Data-Driven Support

Terms such as “remediation” and “differentiate instruction” were used to provide students with instruction geared toward individual needs. For example, Bestie stated, “Differentiate instruction to improve scores.” The remediation and differentiation of instruction were centered around data-driven teaching. Administrators discussed the importance of teachers having the correct data to review. Diagnostic data to assess weaknesses and gaps in learning was mentioned as a requirement for teaching. Chy referred to the process as monitoring progress and said, “Assess and reassess as we move forward in instruction.” Furthermore, Angels discussed administrators' critical role in assisting teachers in “preparing instruction to meet the needs of students.”

Theme 5: Questioning Test Validity

The validity and reliability of the tests were an area of concern, especially for administrators. The administrators indicated that no content or structural changes have been made to state standardized tests since the pandemic. Chy noted tests are an instruction guide and a guide to determine the necessary resources, but not the “sole tool” in decision-making. Angels added, “We cannot keep giving the same kind of tests and expect something different; that is the definition of insanity.” The administrators agreed that scores were inaccurate after the pandemic, and the state was not assessing the students, in the same manner teachers were expected to teach. Furthermore, the current tests were not setting the students up for success. Cardinals stated, “Education is in a state of emergency; we are in survival mode.”

Theme 6: Recognizing Magnified Disparities Among Students

The inequities and disadvantages across economic levels and races and the impact on student academics were discussed by teachers and administrators. The academic achievement gap between minority and low-socioeconomic students has been an issue before the pandemic. Participants noted that the current tests were biased. For example, Cardinals stated, “Tests are geared toward certain races and nationalities. Students' personal experiences are different based on their culture and socioeconomic background.” Bestie added, “The pandemic showcased the disparities across many things. Some parents didn't have the resources to support students' education. Some parents had the resources for tutors and private schools.”

Discussions

According to the American Psychological Association (2020), educators should consider the administration and scoring of tests, and individual student context and exercise caution when making clinical decisions. Teacher participants expressed concern about the lack of time available to adequately prepare students for state standardized tests. Despite having a curriculum pacing guide aligned with state standards and a test blueprint, teachers found it challenging to address the learning gap in test preparation. The learning gap was a central concern in the study, with all participants acknowledging that learning came to a halt when schools closed. Relaxed grading policies during school closures contributed to students not participating in the coursework, failing to complete assigned work, and irregular attendance in virtual sessions. The relaxed accountability policy further widened the learning gap among students. Despite the significant learning gap, state standardized tests have remained unchanged. Participants agreed that the usual four to six weeks dedicated to test preparation was no longer available. Instead, teachers had to re-teach and remediate missed content during the pandemic to bridge the learning gap and teach the current content.

The intent of accountability measures is to monitor student progress. Given the circumstances the pandemic has created, states' testing requirements should be assessed. In February 2021, the U.S. Department of Education issued a notice to states concerning accountability testing for elementary and secondary education (U.S. Department of Education Office of Elementary and Secondary Education, 2021). The notice suggested a shortened version of statewide assessments, remote testing, and an extension of the testing window. Individual states were invited to request waivers for testing requirements. The review of course materials and assessments based on the circumstances is imperative. In addition, the different teaching modalities, social justice, and equity issues require examination in testing decisions. Tests' short-term and long-term potential impacts on students' academic plans, teachers' evaluations, and school policies necessitate further evaluations and modifications.

Participants noted that the inability to close the learning gap and sufficiently prepare students for state standardized tests has contributed to student anxiety. Ewell et al. (2022) discovered six factors contributing to test anxiety related to COVID-19. These factors include difficulty maintaining attention, inability to use academic supports, difficulty constructing meaning, shifting priorities, difficulty organizing academic tasks, and limited time for test preparation. Anxiety is commonly associated with decreases in motivation and engagement. Situational test score pollution factors, such as test anxiety, stress, fatigue, motivation, and test administration practices, have been documented by Abrams et al. (2003), Chalak and Tavakoli (2010), and Haladyna (1992).

Castro et al. (2020) and Townsley (2020) pointed out that not all communities have an equal opportunity to access high-quality distance learning. Hence, it is imperative that grading policies prioritize the principle of "do no harm" to prevent students' academic standings from being negatively impacted by the pandemic. Districts have implemented this principle in the Atlantic Coast state in their grading practices. When in-person instruction was suspended in March 2020, cumulative grading ceased, and grades were only given for assignments that contributed to enhancing the existing baseline grades. This policy remained in effect for the remainder of the school year.

The "do no harm" grading policy was intended to help students, but the participants had differing opinions. According to teachers, some students were promoted to the next grade level simply because of their age, without considering their academic performance or learning gaps. Fisher et al. (2022) observed that although teachers were more flexible with grading policies during

the pandemic, students' academic achievement declined. School administrators were accused of not holding students accountable and maintaining stability in terms of work completion. The participants generally believed that lowering expectations had altered students' and parents' attitudes toward education.

Participants agreed that chronic absenteeism poses a significant challenge to students' academic success and that the problem of student attendance started with the shift to online learning and the more lenient grading policies. One participant emphasized that they have no control over external factors and the student's home environment, making it challenging to address the issue. It is well-known that chronic absenteeism negatively impacts test scores (Santibañez & Guarino, 2021), and in the 2021-2022 academic year, the state reported 20.1% of students as chronically absent (Virginia Department of Education [VDOE], 2023). According to the U.S. Department of Education, chronic absenteeism refers to a student's absence from school for 15 or more days in a school calendar year (U.S. Department of Education, 2023). McDonald et al. (2022) suggested that the rise in student absenteeism may be due to COVID-related anxiety, new rules and procedures, and the awareness of the academic gap. Some special education students found coping with sensory stimulation, social situations, and academic demands overwhelming. Additionally, one participant mentioned that some students flourished in home environments and faced difficulties transitioning back to school. While some students prefer online learning, it may not be a viable option for everyone due to their home circumstances.

Numerous studies conducted since 2020 have emphasized the significance of tending to the psychological needs of COVID-affected students. Jones et al. (2023) have identified various stressors that can potentially lead to post-traumatic stress disorders, anxiety, depression, and even suicide in adolescents, including losing family members, disruption in the grieving process, fear of infecting loved ones, and separation from friends and family. Additionally, Rogers et al. (2021), Naff et al. (2022), Jones et al. (2023), and Minkos and Gelbar (2020) have discovered that traumatic life events can result in disruptive behavior, substance abuse, and lower academic achievement. As a result of social distancing and quarantine protocols, students have been isolated and unable to participate in extracurricular activities, connect with others emotionally, and achieve significant life milestones. Furthermore, the traumatic events that students have witnessed during the pandemic have adversely affected their emotional well-being.

The pandemic's transition to online learning has significantly impacted students' sense of belonging and connection to the school. Emotional struggles have been a common occurrence for many students due to unexpected and prolonged changes in their education and social lives. As per Fisher et al. (2022), 20% of students have reported experiencing depression and anxiety, resulting in a 30% decline in academic performance. Students' home environment also plays a crucial role in their emotional well-being, as reported by Rogers et al. (2021), who found that a lack of privacy and personal space can cause additional anxiety and depression. Additionally, Jones et al. (2023) and Pincus et al. (2020) reported increased conflicts among family members and domestic violence, which contribute to these issues. A longitudinal study by Theodorou et al. (2022) revealed a rise in child abuse cases from 8.3% pre-COVID to 13.5%. The numbers highlight the need for educators to be aware of the potentially traumatic events that students may have encountered during the pandemic. In addition, disruption to routines contributed to students' emotional well-being during the pandemic (Jones et al., 2023; Naff et al., 2022; Rogers et al., 2021; Scott et al., 2021). Rogers et al. (2021) posited that the lack of routine led to feelings of lethargy and sadness among adolescents. Ewell et al. (2022) and Naff et al. (2022) found students struggled with maintaining attention, attending school, and academic performance. The lack of routine might contribute to students' challenges transitioning to in-person learning.

Teachers in the study raised concerns regarding the over-reliance on technology by their students, which may contribute to learning gaps and communication difficulties. Studies by Ravens-Sieberer et al. (2014) and Pincus et al. (2020) have found a significant link between increased screen time and adverse effects on physical well-being, intellectual impairment, emotional challenges, and speech difficulties. Adibelli and Sümen (2020) have also noted that prolonged screen use and a sedentary lifestyle can lead to long-term physiological and psychological health issues. For instance, children's sleep patterns are related to their self-esteem scores. Due to the pandemic, school closures have necessitated using computers for academic work and social interaction. While technology has become the primary means of communication during the pandemic, excessive use may have hindered students' ability to communicate face-to-face.

The pandemic has led to a decline in communication between students and teachers at the middle and high school levels, according to a study by Lessard and Puhl (2021). This makes it crucial to create trauma-informed learning environments to ensure long-term academic and social success for students (Avery et al., 2021). Teachers play a critical role in identifying students' emotional and academic needs early on. Lack of communication between teachers and students can hinder the early detection of necessary student support. Trauma-informed teachers can improve the learning environment by recognizing individual needs, promoting a safe in-person and online environment, fostering positive connections, and building a support system (Scott et al., 2021).

Learning Policy Institute declared that the Whole Child framework dictates child development is malleable and the brain changes in response to experiences and relationships. Human relationships are essential to healthy development and learning, each child's development is unique based on experiences, and adversity affects learning. Pandemic-related school closures, social distancing, the death toll, and the educational impact manifest as traumatic experiences. The traumatic events caused by the pandemic had physiological and mental health connections. For example, diet, weight gain, lack of exercise, disturbed sleep patterns, and headaches are associated with increased emotional difficulties and apathy (Adibelli & Sümen, 2020; Moulin et al., 2022). Moulin et al. (2022) added that emotional difficulties and sleep disturbance are perpetual cycles that increase anxiety. Amalgamated, the factors could also be contributors to test score pollution and inaccurate test scores.

Gadermann et al. (2022) noted that some students faced family unemployment, financial pressure, and responsibilities of caregiving and homeschooling siblings during the pandemic. The dynamics led to decreased levels of an optimistic outlook on life and overall life satisfaction in adolescents. Pisano et al. (2021) found that adolescents showed clinically significant anxiety levels, depression, and psychopathology during the pandemic. Females were more symptomatic than males. Differences in care between racial and ethnic groups were also evident. Low socioeconomic students showed more emotional difficulty, hyperactivity, inattention, and anxiety (Moulin et al., 2022).

Support was crucial to student success during the school closures and the transition back to in-person learning. Schnieders (2023) found that only six percent of students (n=1881) received check-ins about mental health support. 10% of the students noted receiving mental health resources from educators, and 26% had access to school counselors for mental health support. Approximately 50% of students reported support for academics during the first year of the pandemic. Considering the health, social, emotional, and academic challenges students faced, participants in this study agreed on a well-structured support program to use social workers, school counselors, and school psychologists to be established, and the sole focus could not be academics.

The use of unchanged state standardized tests in light of the challenges and learning setbacks brought about by COVID-19 is a notable concern. Yang and Xin (2022) noted that the cognitive ability structure, knowledge structure, and interpretation of current tests are prerequisites

for their validity and reliability. Criterion-related validity, content validity, and construct validity are three measures used to gauge the extent to which the tests measure what they are intended to. Kane (2000, 2013a) stated that criterion and construct validity are not enough to ensure the validity of tests and that an argument-based approach to interpreting scores is necessary. Furthermore, the validity of the interpretation relies on the population taking the test (Kane, 1994). Administering the same tests to students with different needs, challenges, social-emotional and academic losses poses a challenge to the validity and reliability of the current state standardized tests.

COVID-19 drastically changed how students received educational services, attendance rates, social-emotional well-being, family dynamics, availability of equal and equitable resources along various subgroups, and instructional materials. Sireci and Suarez-Alvarez (2022) provided criteria for evaluating the quality of educational tests proctored after the COVID-19-related school closures and identified six consistency criteria for the validity of standardized tests administered to all students; (a) consistency in student participation, (b) consistency in instruction, (c) consistency in the opportunity to learn, (d) consistency in testing format, (e) consistency in students' psychological characteristics, and (f) validity argument. The student participation rate was inconsistent among subgroups. Student participation, learning opportunity, social-emotional well-being, and academic expectations were inconsistent. Administration of the same statewide standardized tests to all students who had different experiences and access to education challenges the validity of the tests. Using disrupted data with varied academic expectations to support the pre-COVID accountability measures illustrates Campbell's Law.

According to Sireci (2020), accurately interpreting test scores requires considering the differences among student populations, such as their access to home resources, family dynamics, cultural and language backgrounds, communication norms, and educational experiences. Unfortunately, no federal laws are in place to address accommodations for these differences among student subgroups. Kane's Validation Model supports the argument that interpreting test scores is crucial for ensuring their validity, especially in the aftermath of the pandemic.

The pandemic exasperated economic, social, and racial disparities in education (Dimitrijevska-Markoski, 2022; Lund et al., 2020; Moulin et al., 2022; Rogers et al., 2021; Santibañez & Guarino, 2021; Schnieders, 2023; Scott et al., 2021; Sosu et al., 2020). Frohn (2021) identified areas in which student disparities could have been heightened during the pandemic: (a) physical space for students to complete school work, (b) access to technical equipment, (c) access to non-technical resources, (d) skills and competencies of parents/caregivers, (e) attitude of parents/caregivers towards school, (f) subject-specific skills of students and caregivers, (g) motivation of students and caregivers, (h) self-sufficiency skills of students, (i) digital and media skills of students and caregivers, (j) language skills of students and caregivers, and (k) social-emotional skills of students. For example, Schnieders (2023) noted that students whose parents were not college graduates were adversely impacted in academics and resilience.

Mast (2018) found that students living in low-income housing areas perform academically below their peers. Supovitz et al. (2023) identified seven inequalities impacting students' academic success. The inequities are related to race, differences in economic status, cultural capital, social capital, digital, health, and policy. For example, school funding which relies on state and local taxes further exasperates the inequities. The fiscal budget of school districts determines the availability of resources, hiring highly qualified teachers, and ensuring continued academic and mental support for students and teachers. Hence, student support is determined by socioeconomic placement. Jensen (2009) stated that socioeconomic status is strongly associated with IQ, academic achievement, literacy, and grade retention.

E. KALEMDAROGLU-WHEELER

Moreover, impoverished children are more likely to act out, be more impulsive, and have less empathy for others. Ewell et al. (2022) stated that the shift in student concern during the pandemic was to survival issues of housing and food insecurity and added caregiving roles for siblings and elderly family members. An equally significant aspect was parents' financial difficulties and association with children's emotional challenges, including hyperactivity and inattention (Moulin et al., 2022).

Although students across various races, ethnic, and socioeconomic backgrounds reported challenges with academic progress, the most prevalent challenges differed by socioeconomic subgroups. Scott et al. (2021) concluded that Black/African American and Asian students reported higher academic challenges, and Hispanic/Latino students indicated higher challenges in physical health. Pre-existing disparities might have caused the challenges. A participant noted, "Inequities and disproportionate things there had been, was brought to light and showcased through the pandemic." COVID-19 disproportionately impacted Black and Hispanic communities (Center for Disease Control [CDC], 2023). Black and Hispanic students were also affected by disproportionate school spending from racial inequalities (Haderlein et al., 2021). Oster et al. (2021) concluded that non-Hispanic students had the most access to full-time in-person instruction.

During the 2021-2022 school year, a significant disparity was observed in student absenteeism. Out of the 245,000 chronically absent students in the Atlantic Coast state, about 158,000 were economically disadvantaged. Additionally, 44,000 students had disabilities, and 6,000 were homeless. A comparison of race and ethnicity revealed that approximately 149,000 students identified as Asian, Black, Hispanic, Native Hawaiian, or of multiple races (VDOE, 2023). Sosu et al. (2020) noted that COVID-19 caused school absenteeism, further perpetuating socioeconomic disparities and exacerbating the educational achievement gap.

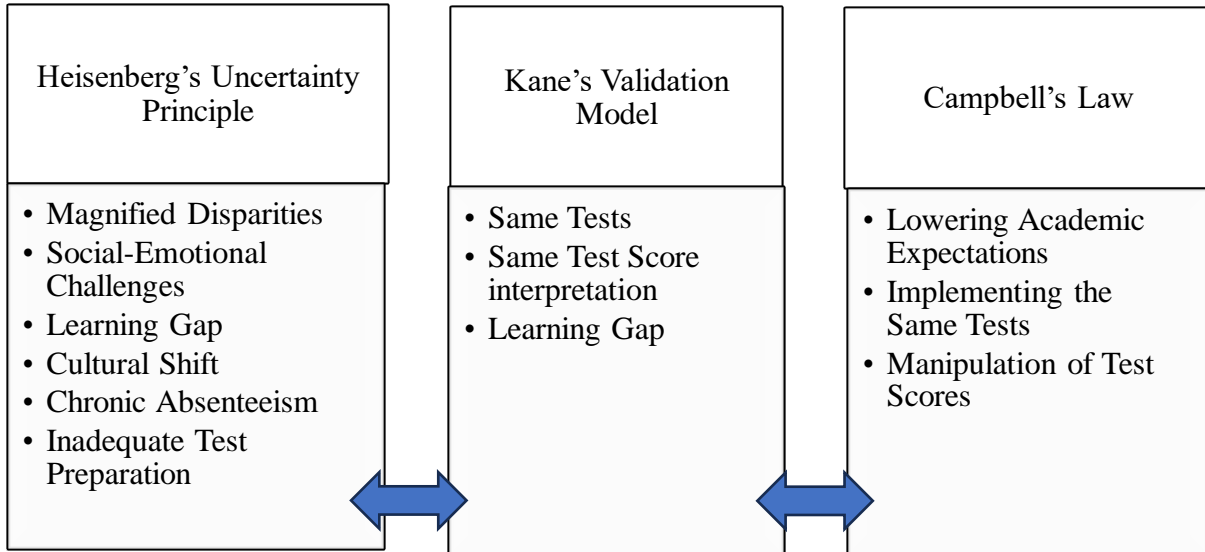
School absences directly impact students' test scores and social-emotional skills (Santibañez & Guarino, 2021). Perhaps the impact is an incessant cycle. The more social-emotional and academic challenges students encounter lead to increased absences. Rogers et al. (2021) indicated adolescents from lower-income households experienced more significant family conflicts and higher levels of anxiety and depression. Lower-income families had additional stress on obtaining resources and quarantining in crowded homes. Lund et al. (2020) added systematic health and social inequities increased the risk of getting COVID and other health-related issues. Low-income students, students with disabilities, foster youth, and homeless students were more subject to learning loss and chronic absenteeism, given the challenges of the pandemic.

The 2022 test scores in this state highlighted a concerning achievement gap among students. As per the National Center for Education Statistics (2022), Black students obtained an average score of 31 points lower than White students on the state-mandated 8th-grade math tests. In the same year, Hispanic students scored 23 points lower than White students, and those who qualified for free-or-reduced lunch earned an average of 31 points lower than their non-eligible peers. Likewise, these subgroups also demonstrated lower performance on the reading assessments.

The disparities in educational services at the height of the pandemic were captured shortly after the school closures. In a survey completed by Educators for Excellence (2020), educators indicated that only nine percent of students participated in virtual instruction 91 to 100% of the time. In addition, 67% of the students showed a decline in work completion. In response to equity and vulnerable populations, the same survey showed schools could meet the needs of minority students 52% of the time, students from low-income at 46%, and students with disabilities around 35% of the time. Schools could meet the needs of homeless students 21% of the time and English as second language learner students 34% of the time. Social inequality during the pandemic was persistent through the digital, economic, structural, and cultural divides (Goudeau et al., 2021).

Given the data, equal and equitable education expectations for all students had waived during the pandemic. Figure 1 outlines the emergent themes aligned with the conceptual framework.

Figure 1
Alignment of Emergent Themes to Conceptual Framework



Recommendations

Although state standardized tests are designed according to specific curriculum standards, pacing, and students' cognitive development, unexpected changes or disruptions can affect the testing process. Glickman et al. (2018) noted that despite having plans and procedures in place, disturbances within or outside the established system could impact school improvement efforts and potentially cause them to fail. The COVID-19 pandemic presented unforeseen challenges that required immediate changes to learning modalities and environments. These changes are examples of Heisenberg's Uncertainty Principle. According to Sireci and Suarez-Alvarez (2022), there are various plausible explanations for the changes in student performance during the pandemic, and these disturbances can be considered pollution factors in test scores. Despite the learning gap caused by interruptions in teaching and learning, state standardized tests are still expected to be implemented.

Per Campbell's Law, lowering passing scores on state tests and implementing verified credit options in lieu of passing the standardized test aided in modifying scores in the interest of accountability (Best, 2020). In addition, Kane (2000) posited criterion-based assessments do not provide a solid basis for validation because test developers, not the teachers, make decisions regarding test content. Furthermore, based on Kane's Validation Model, the interpretation of test results is pivotal in establishing the validity of the tests. The education field is intertwined with public policy, curriculum and instruction, and the guidelines which oversee the daily operations of schools. The recommendations to leaders and practitioners in educational policy, school leadership, and curriculum and instruction are derived from the study's findings and literature review.

Teachers and school administrators have identified inadequate test preparation as a critical factor contributing to low test scores on state standardized tests. With the challenges posed by the pandemic and the resulting learning gap, administering these tests has become even more difficult. While Pan and Sass (2020) suggested extending school days to address this issue, the shortage of teachers and the impact on working parents make this solution impractical. Instead, extending the

E. KALEMDAROGLU-WHEELER

school year to include summer months could help bridge the learning gap and better prepare students for standardized tests. Studies by Campbell et al. (2019) and Fitzpatrick and Burns (2019) have shown that summer learning programs can help students make academic gains in reading and mathematics, especially for students from low socioeconomic backgrounds. By providing additional instruction time, students can catch up on what they missed due to the pandemic, and teachers can teach test-taking strategies to improve results.

Prioritizing students' mental health, particularly during the pandemic, is paramount for educators. This task involves considering each student's emotional well-being and acknowledging their unique experiences and developmental stages. To ensure that students receive the necessary support, schools should increase the number of counselors, psychologists, and social workers. Solely monitoring academic progress is inadequate. There is a critical need for educational systems to establish an intervention system to track the social and emotional progress of the students. Teaching self-regulation skills and prioritizing social-emotional well-being should be the main objectives, with trauma-informed pedagogy employed in classrooms through creating a physically and emotionally safe environment, transparent decision-making, and supportive peer groups. Professional development and changes in educational practices may be necessary. Consistent implementation of social-emotional lessons, with a tier-system approach to provide support, ranging from daily check-ins to small group sessions to whole group lessons, is the key to success.

Community outreach programs are crucial for families of different socioeconomic backgrounds to access necessary resources. Effective communication between educators, school counselors, school psychologists, social workers, students, and parents is essential to ensure students' academic and social-emotional growth, in addition to providing them with stable technology and connectivity. A recommendation is for schools to employ a full-time school and community liaison who can establish communication with families and build stronger community relationships to support educational goals. By prioritizing educational values, stakeholders may be more likely to support schools in their expectations of student attendance, work completion, and accountability. Education policymakers must recognize the disparities among students and focus on allocating resources and financial support to enhance mental health resources, close the achievement gap, and provide families with the necessary training to support students. Furthermore, professional development and training in cultural competency to address structural and social inequities should be adopted.

It is essential for all schools to implement coordinators of assessment and remediation to monitor students' academic and social-emotional progress using data-driven approaches. The coordinators can assist in tracking progress and creating personalized goals for every student. Additionally, providing each student with an academic plan based on diagnostic assessments and progress monitoring, akin to individualized education plans for special education students, is highly recommended. Lastly, restructuring courses to meet students' academic needs by modifying the usual sequence of courses and adding classes designed to bridge content and skill gaps is necessary.

One issue with preparing students for state standardized tests is the inadequate administration of said tests. Compounded by the loss of learning and content gaps, successful student performance becomes challenging. A possible solution is to assess the uses and implementation of pre-COVID-prepared standardized tests. This assessment should involve all stakeholders, including educators, parents, and students, to build agreement and redirect the academic accountability system. Educators in this study also raised concerns about test validity and reliability, which refer to the accuracy and consistency of test scores. Given the unpredictable circumstances created by the pandemic, it is crucial for states and test developers to recreate tests that align with state standards and address the academic gap created by the loss of instruction. Additionally, test developers should be more aware of the disparate effects of the pandemic on

diverse populations and ensure that test items address cultural competency and multicultural perspectives in education.

Limitations and Conclusion

It is imperative to acknowledge that the study's timing significantly limited its scope, as it solely concentrated on the perceptions of educators and administrators concerning COVID-19-related issues. It is plausible that the study would have yielded different themes if it had been conducted during the in-person learning transition phase, thus enabling participants to provide more comprehensive feedback. Additionally, the language utilized in recruitment materials may have influenced potential participants. Furthermore, none of the participants were familiar with test score pollution, which could have potentially intimidated some educators during the recruitment phase.

Standardized tests could be used as diagnostic tools for teachers and districts to determine educational support and program needs. Factors such as anxiety, administration of remote assessments, and testing supervision may have impacted the reliability of the scores (Blad, 2021). Although there are fairness issues concerning standardized tests, assessments are needed to identify learning gaps, monitor student progress, and adjust support accordingly (Goldhaber & Özek, 2019; Hegazy et al., 2021). While recent test questions are not available for review, the tests could identify content area gaps among the students.

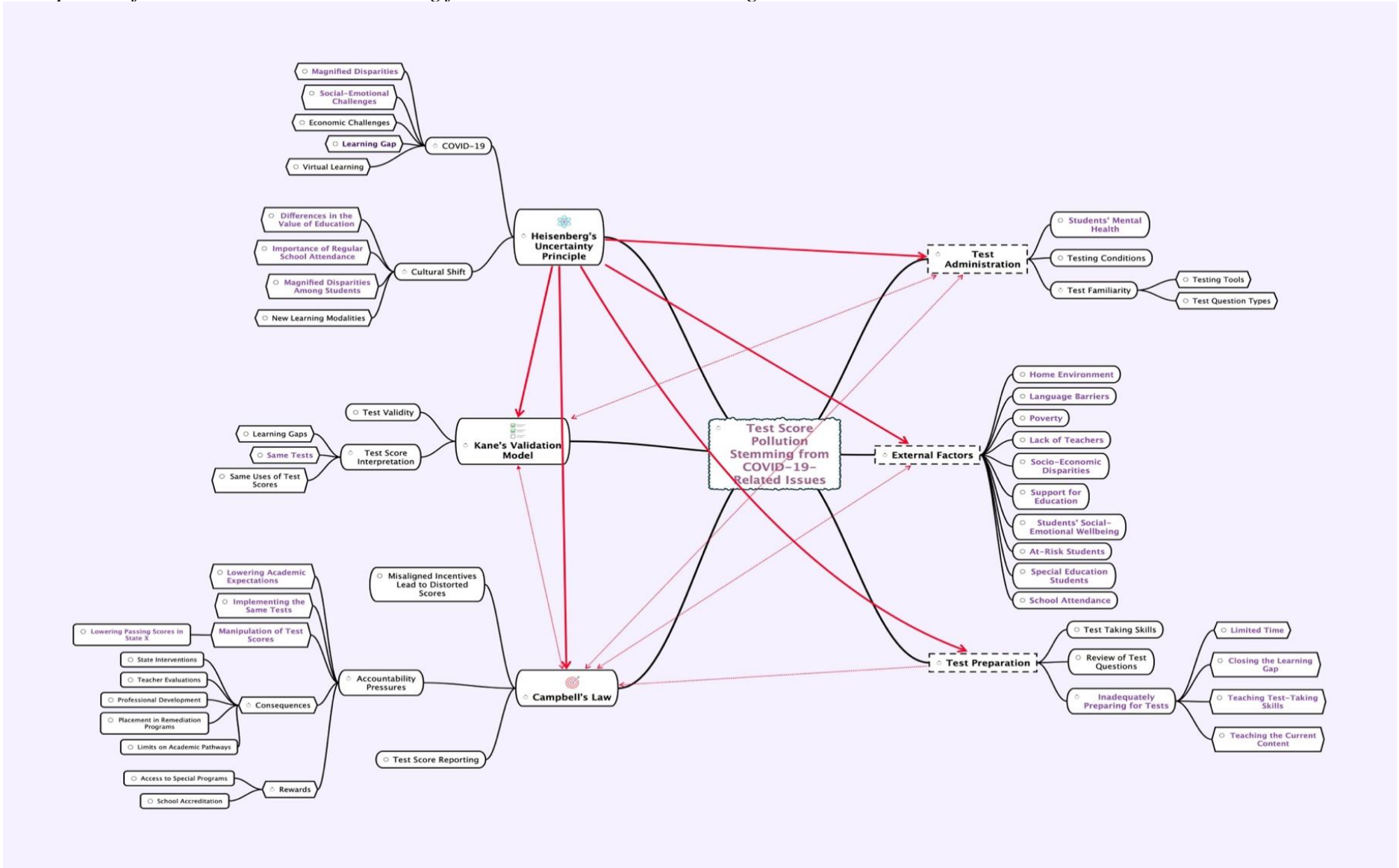
The COVID-19 pandemic has undoubtedly caused disruptions in learning and testing, leading to gaps in academic content and unequal outcomes among students. Kane's Validation Model (Kane, 1994) and Campbell and Stanley's (1963) research have identified various factors that can impact test scores' validity and interpretation, such as student selection, history, maturation, instrumentation, and mortality. The pandemic has resulted in the loss of educational time due to limited resources and learning methods, which could compromise the validity of tests. Furthermore, the pandemic has adversely affected students' well-being, educational abilities, and test administration, all of which could threaten test scores' validity. Lastly, comparing students' test scores before and after the pandemic could introduce bias issues that require careful attention.

Regarding reliability, testing errors can arise from test administration or test instruments. Loeb and Byun (2019) noted that students' test scores might not reflect the learning or learning contribution of the school. Hence, consideration should be given when using students' test scores in regulatory, political, and educational decision-making. Because test scores can influence public education policy, identify school districts and regions by performance level, and assign schools a performance grade, interpreting scores becomes significant in education policymaking. As test scores are of high consequence, distortion of scores becomes a possibility per Campbell's Law (Fahle et al., 2019). The more incentives and consequences are attached to quantitative performance, the more amenable to corruption (Stone, 2020).

This study revealed a lack of awareness among educators regarding test score pollution and contributing factors, as the participants were unfamiliar with the term. The current state standardized tests cover materials students were not exposed to due to school closures. As such, the content of the tests and external factors students may compromise the validity of the scores. Furthermore, student remediation based on diagnostic data is essential to compensate for the learning gap. Lastly, recognizing the disparities among students, addressing social-emotional needs through trauma-informed teaching, and involving school counselors and social workers is vital. Figure 2 depicts a model of the conceptual framework, themes, and recommendations for leaders.

Figure 2

Perceptions of Test Score Pollution Stemming from COVID-19 & State Testing



References

- Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice, 42*(1), 18–29. https://doi.10.1207/s15430421tip4201_4
- Adıbelli, D., & Sümen, A. (2020). The effect of the coronavirus (COVID-19) pandemic on health-related quality of life in children. *Children and Youth Services Review, 119*. <https://doi.org/10.1016/j.chilyouth.2020.105595>
- Alimorad, Z. (2014). Ethical conflicts experienced by Iranian EFL teachers in the classroom context. *Teflin Journal, 25*(1), 1–15. <http://teflin.org/journal/index.php/journal/article/view/167/157>
- American Psychological Association. (2020, June). *Teaching and supervising assessment beyond COVID-19*. <https://www.apa.org/topics/covid-19/teaching-supervising-assessment-beyond.pdf>
- Amrein, A. L., & Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership, 60*(5), 32–38.
- Amrein-Beardsley, A., & Barnett, J. H. (2012). Working with error and uncertainty to increase measurement validity. *Educational Assessment, Evaluation and Accountability, 24*(4), 369–379. <https://doi.org/10.1007/s11092-012-9146-6>
- Ansorger, J. (2021). An analysis of education reforms and assessment in the core subjects using an adapted Maslow's Hierarchy: Pre and post COVID-19. *Education Sciences, 11*(8), Article 376. <https://doi.org/10.3390/educsci11080376>
- Avery, J., Morris, H., Galvin, E., Misso, M., Savaglio, M., & Skouteris, H. (2021). Systematic review of school-wide trauma informed approaches. *Journal of Child & Adolescent Trauma, 14*(1), 381–397. <https://doi.org/10.1007/s40653-020-00321-1>
- Bai, H., Bowering, S., Miyakawa, M., Cohen, A., & Scott, C. (2020). Unleashing the elephant out of the closet and into the wildness of inner work. In O. Ergas & J. K. Ritter (Eds.) *Exploring self toward expanding teaching, teacher education and practitioner research: Advances in research on teaching* (Vol. 34, pp. 37–54). Emerald Publishing Limited. <https://doi.org/10.1108/S1479-368720200000034002>
- Best, J. (2021). How to lie with Coronavirus statistics: Campbell's Law and measuring the effects of COVID-19. *Numeracy: Advancing Education in Quantitative Literacy, 14*(1), 1–17. <https://doi.org/10.5038/1936-4660.14.1.1378>
- Blad, E. (2021, March 5). *How two years of pandemic disruption could shake up the debate over standardized testing*. Education Week. <https://www.edweek.org/teaching-learning/how-two-years-of-pandemic-disruption-could-shake-up-the-debate-over-standardized-testing/2021/03>
- Braganza, O. (2022). Proxyeconomics, a theory and model of proxy-based competition and cultural evolution. *Royal Society Open Science, 9*(2), Article 211030. <https://doi.org/10.1098/rsos.211030>
- Bridges, K., & Woolcock, M. (2022, June). *Measuring what matters*. World Bank Group. <https://doi.org/10.1596/1813-9450-10051>
- Byrd, M., & Varga, B. A. (2018). The manifestation of Campbell's Law: Consequences of eliminating of social studies from the curriculum. *Social Studies, 109*(1), 27–33. <https://doi.org/10.1080/00377996.2017.1406322>
- Campbell, D. (1979). Assessing the impact of planned social change. *Evolution and Program Planning, 2*(1), 67–90. [https://doi.org/10.1016/0149-7189\(79\)90048-X](https://doi.org/10.1016/0149-7189(79)90048-X)

- Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. McNally & Company.
<https://archive.org/details/experimentalquas0000camp/page/n1/mode/2up>
- Campbell, L., Sutter, C., & Lambie, G. W. (2019). An investigation of the summer learning effect on fourth grade students' reading scores. *Reading Psychology*, 40(5), 465–490.
<https://doi.org/10.1080/02702711.2019.1629516>
- Castro, M., Choi, L., Knudson, J., & O'Day, J. (2020). *Grading policy in the time of COVID-19: Considerations and implications for equity*. Policy and Practice Brief. California Collaborative on District Reform. <https://cacollaborative.org/publication/grading>
- Center for Disease Control. (2023). *COVID-19 weekly cases and deaths per 100,000 population by age, race/ethnicity*. COVID Data Tracker. <https://covid.cdc.gov/covid-data-tracker/#demographicsovertime>
- Chalak, A., & Tavakoli, M. (2010). Sources of test score pollution: State of the art. *Curriculum Planning and Research in Educational Sciences*, 7(26), 13–34.
<https://www.sid.ir/en/journal/ViewPaper.aspx?id=246104>
- Cuban, L. (1991). *The misuse of tests in education*. Office of Technology Assessment, U.S. Congress. <https://eric.ed.gov/?id=ED340780>
- Cunningham, J. (2019). Missing the mark: Standardized testing as epistemological erasure in U.S. schooling. *Power and Education*, 11(1), 111–120.
<https://doi.org/10.1177/1757743818812093>
- Dimitrijevska-Markoski, T. (2022). Reducing COVID-19 racial disparities: Why some counties make data-driven decisions and other do not? *Public Performance & Management Review*, 45(5), 1235–1257. <https://doi.org/10.1080/15309576.2022.2101131>
- Donohue, J. M., & Miller, E. (2020). COVID-19 and school closures. *JAMA*, 324(9), 845–847.
<https://jamanetwork.com/journals/jama/fullarticle/2769033>
- Educators for Excellence. (2020). *Voices from the virtual classroom: A survey of America's teachers on COVID-19-related education issues*. <https://e4e.org/news/survey-americas-educators/voices-virtual-classroom>
- Emler, T. E., Zhao, Y., Deng, J., Yin, D., & Wang, Y. (2019). Side effects of large-scale assessments in education. *ECNU Review of Education*, 2(3), 279–296.
<https://doi.org/10.1177/2096531119878964>
- Ewell, S., Josefson, C., & Ballen, C. (2022). Why did students report lower test anxiety during the COVID-19 pandemic? *Journal of Microbiology & Biology Education*, 23(1), 1–11.
<https://doi.org/10.1128/jmbe.00282-21>
- Fahle, E., Shear, B., & Shores, K. (2019). Assessment for monitoring of education systems: The U.S. example. *The ANNALS of the American Academy of Political & Social Science*, 683(1), 58–74. <https://doi.org/10.1177/0002716219841014>
- Fisher, H., Hawkins, G., Herts, M., Sliwa, S., & Beresovsky, V. (2022). Student and school characteristics associated with COVID-19-related learning decline among middle and high school students in K-12 schools. *Journal of School Health*, 92(11), 1027–1039.
<https://doi.org/10.1111/josh.13243>
- Fitzpatrick, D., & Burns, J. (2019). Single-track year-round education for improving academic achievement in U.S. K-12 schools: Results of a meta-analysis. *Campbell Systematic Reviews*, 2019(15), 1–28. <https://doi.org/10.1002/cl2.1053>
- Frohn, J. (2021). Troubled schools in troubled times: How COVID-19 affects educational inequalities and what measures can be taken. *European Educational Research Journal*, 20(5), 667–683. <https://doi.org/10.1177/14749041211020974>

- Gadermann, A., Thomson, K., Gill, R., Schonert-Reichl, K., Petteni, M., Guhn, M., Warren, M., & Overle, E. (2022). Early adolescents' experiences during the COVID-19 pandemic and changes in their well-being. *Frontiers in Public Health*, *10*, Article 823303. <https://doi.org/10.3389/fpubh.2022.823303>
- Ghaicha, A. (2016). Theoretical framework for educational assessment: A synoptic review. *Journal of Education and Practice*, *7*(24), 212–231. <https://files.eric.ed.gov/fulltext/EJ1112912.pdf>
- Glickman, C. D., Gordon, S. P., & Ross-Gordon, J. M. (2018). *Supervision and instructional leadership: A development approach* (10th ed.). Pearson.
- Goldhaber, D., & Özek, U. (2019). How much should we rely on student test achievement as a measure of success? *Educational Researcher*, *48*(7), 479–483. <https://doi.org/10.3102/0013189X19874061>
- Goudeau, S., Sanrey, C., Stanczak, A., Manstead, A., & Darnon, C. (2021). Why lockdown and distance learning during the COVID-19 pandemic are likely to increase the social class achievement gap. *Nature Human Behavior*, *5*(1), 1273–1281. <https://doi.org/10.1038/s41562-021-01212-7>
- Haderlein, S., Saavedra, A., Polikoff, M., Silver, D., Rapaport, A., & Garland, M. (2021). Disparities in educational access in the time of COVID: Evidence from a nationally representative panel of American families. *AERA Open*, *7*(1), 1–21. <https://doi.org/10.1177/23328584211041350>
- Haladyna, T. (1992). Test score pollution: Implications for limited English proficient students. *Focus on Evaluation and Measurement*, *1*(2), 135–149. <https://files.eric.ed.gov/fulltext/ED349826.pdf>
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, *20*(5), 2–7. <https://doi.org/10.3102/0013189X020005002>
- He, S. Y., & Giuliano, G. (2018). School choice: understanding the trade-off between travel distance and school quality. *Transportation*, *45*(5), 1475–1498. <https://doi.org/10.1007/s11116-017-9773-3>
- Hegazy, N., Elrafie, N., Saleh, N., Youssry, I., Ahmed, S., & Yosef, M. (2021). Consensus meeting report "technology enhanced assessment" in Covid-19 time, MENA regional experiences and reflections. *Advances in Medical Education and Practices*, *12*, 1449–1456. <https://doi.org/10.2147/AMEP.S331829>
- Heisenberg, W. (1927). *The actual content of quantum theoretical kinematics and mechanics* (SCITRAN, Trans.). National Aeronautics and Space Administrations. (1983). <https://ntrs.nasa.gov/api/citations/19840008978/downloads/19840008978.pdf>
- Hess, F. M., & McShane, M. Q. (2018). The happy (and not so happy) accidents of Bush-Obama school reform. *Phi Delta Kappan*, *100*(4), 31–35. <https://doi.org/10.1177/0031721718815670>
- Huber, S., & Helm, C. (2020). COVID-19 and schooling: Evaluation, assessment and accountability in times of crises—reacting quickly to explore key issues for policy, practice and research with the school barometer. *Educational Assessment, Evaluation and Accountability*, *32*(2), 237–270. <https://doi.org/10.1007/s11092-020-09322-y>
- Jensen, E. (2009). *Teaching with poverty in mind*. ASCD.
- Jones, K., Mallon, S., & Schnitzler, K. (2023). A scoping review of the psychological and emotional impact of the COVID-19 pandemic on children and young people. *Illness, Crisis & Loss*, *31*(1), 175–199. <https://journals.sagepub.com/doi/pdf/10.1177/10541373211047191>

- Kane, M. T. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64(3), 425–461. <https://www.jstor.org/stable/1170678>
- Kane, M. T. (2000). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319–342. <https://doi.org/10.1111/j.1745-3984.2001.tb01130.x>
- Kane, M. T. (2013a). The argument-based approach to validation. *School Psychology Review*, 42(4), 448–457. <https://doi.org/10.1080/02796015.2013.12087465>
- Kane, M. T., & Wools, S. (2020). Perspectives on the validity of classroom assessment. In S. Brookhart, & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp.11–26). Routledge. <https://doi.org/10.4324/9780429507533-2>
- Koretz, D. (2017). *The testing charade: Pretending to make schools better*. The University of Chicago Press.
- Koretz, D. (2019). Measuring postsecondary achievement: Lessons from large-scale assessments in the K-12 sector. *Higher Education Policy*, 32(4), 513–536. <https://doi.org/10.1057/s41307-019-00142-4>
- Lessard, L. M., & Puhl, R. M. (2021). Adolescent academic worries amid COVID-19 and perspectives on pandemic-related changes in teacher and peer relations. *School Psychology*, 36(5), 285–292. <https://doi.org/10.1037/spq0000443>
- Lim, L. (2021). Tertiary education assessment amidst COVID-19: Adjustments to maintain score validity. *Journal of Higher Education Theory & Practice*, 21(15), 1–7. <https://doi.org/10.33423/jhetp.v21i15.4886>
- Loeb, S., & Byun, E. (2019). Testing, accountability, and school improvement. *The ANNALS of the American Academy of Political and Social Science*, 683(1), 94–109. <https://doi.org/10.1177/0002716219839929>
- Longmuir, F., Windsor, S., & Henning Loeb, I. (2021). Disrupted and challenged learning practices: Students' experiences of 2020 as their final year of secondary schooling. *International Journal of Educational Research*, 110. <https://doi.org/10.1016/j.ijer.2021.101879>
- Lund, E. M., Forber-Pratt, A. J., Wilson, C., & Mona, L. R. (2020). The COVID-19 pandemic, stress, and trauma in the disability community: A call to action. *Rehabilitation Psychology*, 65(4), Article 313322. <https://doi.org/10.1037/rep0000368>
- Madaus, G. (2009). The distortion of teaching and testing: High-stakes testing and instruction. *Peabody Journal of Education*, 65 (1988), 29-46. <https://doi.org/10.1080/01619568809538611>
- Mast, B. (2018). School performance of schools assigned to HUD-assisted households. *Cityscape: A Journal of Policy Development and Research*, 20(3), 189–221. <https://www.jstor.org/stable/26524879>
- McDonald, B., Lester, K., & Michelson, D. (2022). 'She didn't know how to go back': School attendance problems in the context of the COVID-19 pandemic—A multiple stakeholder qualitative study with parents and professionals. *British Journal of Educational Psychology*, 93, 386–401. <https://doi.org/10.1111/bjep.12562>
- McIntyre-Mills, J. J. (2021). The importance of relationality: A note on co-determinism, multispecies relationships and implications for COVID-19. *Systems Research and Behavioral Science*, 39, 339–353. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8661985/>
- Merriam, S. B., & Grenier, R. S. (Eds.). (2019). *Qualitative research in practice: Examples for discussion and analysis*. John Wiley & Sons, Incorporated
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research* (4th ed.). Wiley.

- Middleton, K. V. (2020). The longer-term impact of COVID-19 on K-12 student learning and assessment. *Educational Measurement: Issues and Practice*, 39(3), 41–44. <https://doi.org/10.1111/emip.12368>
- Miller, R. K., & Hui, I. (2022). Impact of short school closures (1–5 days) on overall academic performance of schools in California. *Scientific Reports*, 12(1), 1–13. <https://doi.org/10.1038/s41598-022-06050-9>
- Minkos, M., & Gelbar, N. (2020). Considerations for educators in supporting student learning in the midst of COVID-19. *Psychology in the Schools*, 58, 416–426. <https://doi.org/10.1002/pits.22454>
- Moulin, F., El-Aarbaoui, T., Bustamante, J., Héron, M., Mary-Krause, M., Rouquette, A., Galéra, C., & Melchior, M. (2022). Risk and protective factors related to children's symptoms of emotional difficulties and hyperactivity/inattention during the COVID-19-related lockdown in France: Results from a community sample. *European Child & Adolescent Psychiatry*, 31(1), 1103–1114. <https://doi.org/10.1007/s00787-021-01752-3>
- Naff, D., Williams, S., Furman-Darby, J., & Yeung, M. (2022). The mental health impacts of COVID-19 on PK-12 students: A systematic review of emerging literature. *AERA Open*, 8(1), 1–40. <https://doi.org/10.1177/23328584221084722>
- Nassaji, H. (2020). Good qualitative research. *Language Teaching Research*, 24(4), 427–431. <https://doi.org/10.1177/1362168820941288>
- National Center for Education Statistics. (2022). *Common core of data*. https://nces.ed.gov/ccd/districtsearch/district_detail.asp?Search=2&DistrictID=5100840&ID2=5100840&details=
- Neemati, N., Hooshangi, R., & Shurideh, A. (2014). An investigation into the learners' attitudes towards factors affecting their exam performance: A case from Razi University. *Procedia-Social and Behavioral Sciences*, 98, 1331–1339. <https://doi.org/10.1016/j.sbspro.2014.03.550>
- Oster, E., Jack, R., Halloran, C., Schoof, J., McLeod, D., Yang, H., Roche, J., & Roche, D. (2021). Disparities in learning mode access among K-12 students during the COVID-19 pandemic, by race/ethnicity, geography, and grade level - United States, September 2020-April 2021. *MMWR. Morbidity and Mortality Weekly Report*, 70(26), 953–958. <https://doi.org/10.15585/mmwr.mm7026e2>
- Pan, W., & Sass, T. (2020). *Potential remediation strategies in the wake of COVID-19 school closures: A review of the literature*. Georgia Policy Labs Reports, 3. <https://doi.org/10.57709/30728953>
- Pearson, P. D., Vyas, S., Sensale, L. M., & Kim, Y. (2001). Making our way through the assessment and accountability maze: Where do we go now? *The Clearing House*, 74(4), 175–182. <https://www.proquest.com/scholarly-journals/making-our-way-through-assessment-accountability/docview/196836948/se-2?accountid=35812>
- Pincus, R., Hannon-Walker, T., Wright, L., & Justice, J. (2020). COVID-19's effects on students: How school counselors rise to the rescue. *NASSP Bulletin*, 104(4), 241–256. <https://doi.org/10.1177/0192636520975866>
- Pisano, S., Catone, G., Gritti, A., Almerico, L., Pezzella, A., Santangelo, P., Bravaccio, C., Iuliano, R., & Senese, V. (2021). Emotional symptoms and their related factors in adolescents during the acute phase of Covid-19 outbreak in South Italy. *Italian Journal of Pediatrics*, 47(1), 1–8. <https://doi.org/10.1186/s13052-021-01036-1>
- Popham, J. (2020). *Classroom assessment: What teachers need to know*. Pearson.

- Ravens-Sieberer, U., Herdman, M., Devine, J., Otto, C., Bullinger, M., Rose, M., & Klasen, F. (2014). The European KIDSCREEN approach to measure quality of life and well-being in children: Development, current application, and future advances. *Quality of Life Research*, 23(3), 791–803. <https://www.jstor.org/stable/24726378>
- Rogers, A., Ha, T., & Ockey, S. (2021). Adolescents' perceived socio-emotional impact of COVID-19 and implications for mental health: Results from a U.S. based mixed-methods study. *Journal of Adolescent Health*, 68, 43–52. <https://doi.org/10.1016/j.jadohealth.2020.09.039>
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.) SAGE Publications.
- Santibañez, L., & Guarino, C. M. (2021). The effects of absenteeism on academic and social-emotional outcomes: Lessons for COVID-19. *Educational Researcher*, 50(6), 392–400. <https://doi.org/10.3102/0013189X21994488>
- Schnieders, J. (2023). *Disruptions and gains: Students' reflections on the effects of the pandemic*. ACT. <https://www.act.org/content/dam/act/unsecured/documents/R2295-Disruptions-Gains-Effects-of-Pandemic-02-2023.pdf>
- Scott, J., Jaber, L. S., & Rinaldi, C. M. (2021). Trauma-informed school strategies for SEL and ACE concerns during COVID-19. *Education Sciences*, 11(12), Article 796. <https://doi.org/10.3390/educsci11120796>
- Sidorkin, A. M. (2016). Campbell's Law and the ethics of immensurability. *Studies in Philosophy and Education*, 35(4), 321–332. <https://doi.org/10.1007/s11217-015-9482-3>
- Sireci, S. G. (2020). Standardization and understandardization in educational assessment. *Educational Measurement: Issues & Practice*, 39(3), 100–105. <https://doi.org/10.1111/emip.12377>
- Sireci, S. G., & Suarez-Alvarez, J. (2022). Deriving decisions from disrupted data. *Educational Measurement: Issues & Practice*, 41(1), 23–27. <https://doi.org/10.1111/emip.12499>
- Sosu, E., Dare, S., & Goodfellow, C. (2020). Socioeconomic status and school absenteeism: A systematic review and narrative synthesis. *Review of Education*, 2021(9), Article e3291. <https://doi.org/10.1002/rev3.3291>
- Stake, R. E. (1995). *The art of case study research*. SAGE Publications.
- Stone, C. N. (2020). Rhetoric, reality, and politics: The neoliberal cul-de-sac in education. *Urban Affairs Review*, 56(3), 943–972. <https://doi.org/10.1177/1078087419867165>
- Supovitz, J., Hemphill, A., Manghani, O., & Watson, C. (2023). Cogs of inequity: How structural inequities impeded school efforts to support students and families at the outbreak of the COVID-19 pandemic. *Equity in Education & Society*. Advance online publication. <https://doi.org/10.1177/27526461231153666>
- Tannenbaum, R. J., & Kane, M. T. (2019). Stakes in testing: Not a simple dichotomy but a profile of consequences that guides needed evidence of measurement quality. *ETS Research Reports Series*, 2019(1), 1–16. <https://doi.org/10.1002/ets2.12255>
- Theodorou, C., Brown, E., & Jackson, J. (2022). Child abuse and the COVID-19 pandemic. *Journal of Surgical Research*, 276, 18–23. <https://doi.org/10.1016/j.jss.2022.02.039>
- Townsley, M. (2020). Grading principles in pandemic-era learning: Recommendations and implications for secondary school leaders. *Journal of School Administration Research and Development*, 5(S1), 8–14. <https://files.eric.ed.gov/fulltext/EJ1301268.pdf>
- U.S. Department of Education. (2004). *Four pillars of NCLB*. <https://www2.ed.gov/nclb/overview/intro/4pillars.html>
- U.S. Department of Education. (2023). *Chronic absenteeism in the nation's schools*. <https://www2.ed.gov/datastory/chronicabsenteeism.html#intro>
- U.S. Department of Education Office of Elementary and Secondary Education. (2021). <https://www2.ed.gov/policy/elsec/guid/stateletters/dcl-assessments-and-acct-022221.pdf>

- Virginia Department of Education. (2023). *Virginia school quality profiles*.
<https://schoolquality.virginia.gov/virginia-divisions/division-at-a-glance>
- Walt, H., & Madaus, G. (1986). *Effects of standardized testing and the future of the national assessment of educational progress*. The National Assessment of Student Achievement.
<https://files.eric.ed.gov/fulltext/ED279680.pdf>
- Yang, L. P., & Xin, T. (2022). Changing educational assessments in the post-COVID-19 era: From assessment of learning (AoL) to assessment as learning (AaL). *Educational Measurement: Issues and Practice*, 41(1), 54–60. <https://doi.org/10.1111/emip.12492>
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). SAGE Publications.
- Yin, R. K. (2016). *Qualitative research from start to finish* (2nd ed.). Guilford Press.

Notes on Contributor

Dr. Elif Kalemdaroglu-Wheeler completed her B.A. and M.A. in Leadership in Education at the University of Notre Dame of Maryland, and a post-master's Educational Leadership and Administration certification from George Washington University. Her doctoral degree is in Educational Leadership, with a specialization in Curriculum and Instruction. She has 23 years of experience in K-12 teaching and school administration. She is the founder of Capella Educational Services, LLC. She is studying Education Law at St. Mary's University School of Law and pursuing further educational research as a Fellow in Residence with the University of Phoenix.

ORCID

Elif Kalemdaroglu-Wheeler, <https://orcid.org/0009-0000-4308-2614>

Manuscript received May 28, 2023
Final revision received July 24, 2023
Accepted August 15, 2023