

Title:

High-Stakes Testing and its influence on Classroom Instructional Practices in Kenya

Submitted by: African Population and Health Research Committee (APHRC)

Principal Investigator: Moses Ngware - Senior Research Scientist, APHRC,
mngware@aphrc.org; P. O. Box 10787, 00100, Nairobi, Kenya

Co-Investigator: Jack Rossiter - Policy Fellow, Global Education, Center for Global
Development, jrossiter@cgdev.org

Co-Investigator: Amani Karisa - Postdoctoral Research Scientist, APHRC,
mkarisa@aphrc.org

Date: 24 April 2023

Table of Contents	
List of Tables	3
List of Figures	3
Abbreviations & Acronyms.....	4
Abstract	5
Lay summary	5
1. Introduction	6
1.2 Problem Statement.....	8
2. Review of literature.....	8
3. Study objectives	10
4. Research questions.....	10
5. Conceptual Framework	11
6. Methodology.....	12
A. Data collection methods	14
B. Ethical considerations.....	16
C. Data management and analysis	17
7. Study's limitations and mitigation, and assumptions	18
8. Findings dissemination strategy	18
9. Management and organization of the Study	18
10.2 Research tool	22
References	23

List of Tables

Table 1: Change in summative assessment at the primary school level in Kenya.....	7
Table 2: Work plan.....	20
Table 3: The budget.....	21

List of Figures

Figure 1: Potential effects of high-stakes testing.....	11
Figure 2: Test-washback on instructional practice.....	12

Abbreviations & Acronyms

CBA:	Competency Based Assessment
CBC:	Competence Based Curriculum
JSS:	Junior Secondary School
KCPE:	Kenya Certificate of Primary Education
KCSE:	Kenya Certificate of Secondary Education
KNEC:	Kenya National Examination Council
KPSEA:	Kenya Primary School Education Assessment
SSA:	Sub-Saharan Africa
TPD:	Teacher Professional Development

Abstract

Background: Tests are known to influence instructional practices in the classroom. However, most literature on this phenomenon is from the Global North, with gaps in knowledge from the Global South.

Objectives: The broad objective of this study is to generate evidence that improves our understanding of the influence of assessments on teachers' teaching behavior inside the classroom with a focus on Kenya. The following research questions will be addressed:

1. How do assessments influence classroom instructional practices in primary school grades?
2. Why do assessments influence (if at all they do) classroom instructional practices in primary school grades?
3. How are teachers changing their classroom practices as new tests are introduced in Grade 5 and Grade 6, and existing tests phased out in Grade 8 in Kenya?

Methods: This is an exploratory study that will collect data from primary school teachers in Kenya through surveys to examine their perceptions of the influence of assessment and assessment data on their teaching decisions. The study design will be a descriptive survey. The sample will consist of 200 primary school teachers. Data collection methods will include a structured interview using closed-ended questions to collect quantitative data. These data will be analyzed using statistical techniques that employ small sample sizes to explore relationships/correlations. Nonparametric tests will also be used to examine the possible non-normal distributions likely to be encountered with the small sample size.

Results: The findings of this exploratory study are likely to enhance an understanding of what is happening in regard to tests and teaching behaviours. While the limited generalizability of the findings is acknowledged due to the small sample size, our study will shed light on areas that teachers have identified as more or less important with regards to tests and classroom teaching behaviour. By considering their perspectives, the study can offer a meaningful input to the ongoing and active policy debate surrounding the implementation of competence-based curriculum and exam reforms, allowing decision-makers to make more informed choices

Lay summary

Tests can affect how teachers teach in the classroom. However, most of what we know about this comes from research done in wealthy countries, so we do not have enough information about how it works in poorer countries in Africa. This study wants to fill that gap by talking to teachers in Kenya and finding out how they think tests and test results affect their teaching decisions. We will use surveys to ask 200 teachers questions and look for patterns in their answers. We will use numbers to describe what we find. This information will help us better understand how tests and teaching practices are connected in schools.

1. Introduction

In the last two decades, the world has witnessed unprecedented education reforms at various levels – global, regional, national and even at local levels. Some of these reforms have focused on changing the teaching behaviour given the critical role played by teachers in student learning (see for example Dalton et al., 2012, Herman et al, 2022; Nseibo et al, 2022). In almost all education reforms, learning outcomes are used as the measure of success (see for example United Nations Sustainable Development Goals, 2015, Goal 4 Target 1 and the emphasis on learners achieving relevant and effective learning outcomes). In fact, decision-makers, education managers, program staff, monitoring and evaluation experts, and researchers among others place a lot of emphasis on learning outcomes (Halász, 2017; Koh, et al., 2015). It is used as the ‘barometer for the health status’ of the education system, education reforms, the school system, and/or education interventions. This is not to say that it is the only way good outcomes of the education systems should be measured. Teachers therefore, have an overwhelming task of ensuring students learn better so that learning outcomes improve. Teachers use assessments (both formative and summative) to establish the levels of learning and/or to evaluate how well an education reform or an intervention is doing. Furthermore, the education stakeholders expect teachers to ‘produce’ good results after exposing them (teachers) to interventions informed by Teacher Professional Development (TPD) and/or other forms of education reforms.

The foregoing synopsis presents an unintended tension between learning outcomes as a measure of accountability and teacher professional behaviour inside the classroom. This tension, mostly exhibited on the floor of the classroom, influences the way teachers use learning assessment and assessment data. On one hand, decision-makers hold teachers accountable for better learning outcomes. Available evidence shows that teachers teach the test with a view to boosting test scores (a measure of learning outcomes) (Berliner, 2011; Slomp, 2020). This way they are perceived to be accountable. However, gains in test scores driven by the behaviour of ‘teaching to the test’ are short-lived, contrary to expectations of the education reforms and/or TPD. There also exists evidence to the effect that frequent testing/assessment of students boosts learning outcomes. For instance, because of this positive relationship between testing and learning outcomes, it can be postulated that teachers like frequent testing of students.

A historical perspective of Kenya’s Education system

Kenya's education system has a strong historical connection to the British system of education. During the colonial period, the British introduced their education system in Kenya, which was heavily based on examinations. This meant that only those who performed well in exams were allowed to progress to higher levels of education. This system was designed to produce a small elite class of educated individuals who would serve as administrators and civil servants in the colonial government. After Kenya gained independence in 1963, the education system remained largely unchanged, with a continued emphasis on examinations (Imana, 2020).

However, in recent years, there has been a shift towards a more holistic approach to education, with a focus on developing critical thinking skills and practical knowledge (Rono & Bwamoni, 2020; Amukune, 2021; Kretzer & Oluoch-Suleh, 2022). Despite this shift, the legacy of the

British education system can still be seen in Kenya's current education system, with a continued emphasis on examinations and academic achievement (Imana, 2020; Kavua, 2020). However, there are ongoing efforts to reform the education system and move towards a more balanced approach that values both academic achievement and practical skills (Rono & Bwamoni, 2020; Wanjala, 2021).

Kenya's Competence Based Curriculum

Kenya rolled out our education reforms dubbed Competence Based Curriculum (CBC) in 2017. The overall aim of the CBC is to bring into being a citizen who is not only engaged, empowered, and ethical but who can also succeed in a competitive world that requires innovations, creativity, critical thinking, positive vision, adaptability as well as transferable skills (Ministry of Education, 2022). As of 2023, the first cohort of CBC is in its 7th year/grade, which is the first grade of Junior Secondary School (JSS).

Assessment under CBC

CBC came with a new approach to assessment that de-emphasized examinations – dubbed as Competency Based Assessment (CBA). CBA is characterised by continuous assessment at various stages with a view to providing constant feedback on students' learning needs. CBA are administered by teachers (classroom-based assessments) as well as the national examining body (school-based assessments) – Kenya National Examination Council (KNEC), and assess what the learners know and can do including knowledge, skills, values, and attitude. Formative assessments at both classroom and school levels are conducted in grades 4, 5, and 6; while a summative assessment is done at the end of grade 6. School-based assessments are standardised and downloaded from a KNEC website for administration by teachers. At JSS (grades 7 to 9), teachers administer classroom-based continuous assessments (formative), while the KNEC will administer a summative assessment at the end of grade 9. End of grade 9 assessment informs post-JSS choice of academic career pathways.

The main change in the CBC in terms of summative assessment at the primary school level is summarised in Table 1.

Table 1: Change in summative assessment at the primary school level in Kenya

Old curriculum	CBC
The Kenya Certificate of Primary Education examination (KCPE) was done at the end of class 8. It had national performance rankings and determined the students' secondary school choice. Students who scored highly were typically admitted to prestigious schools (with presumably better resources) while the rest were admitted to less prestigious schools.	The Kenya Primary School Education Assessment (KPSEA) is done at the end of grade 6 (or class 6 in the terminology of the old curriculum). KPSEA has no national rankings and does not influence the students' future school placements (after grade 6, students move to JSS in the same school. The current study is delimited to assessments done before grade 6 of the CBC, which is the period before

	JSS. Assessment during/after JSS will not be under focus).
--	--

It is envisaged that the policy changes brought by adopting CBC, including the types of assessments and their use, would have implications on teachers' classroom instructional practices. The CBC policy directs teachers to use the assessment results to inform how they teach. Whether this is actually happening or not remains unclear. As a result, the current exploratory study seeks to understand whether and/or how the prescribed CBC assessment policy changes influence teachers' classroom instructional practices.

1.2 Problem Statement

From the introduction, what we do not know, especially in sub-Saharan Africa, is how such testing/assessment influences teaching behaviour inside the classroom. Furthermore, it would be useful for decision-makers and teachers to understand the extent to which assessments influence teaching behaviour - and how this is beginning to change under the new CBC and CBA regimes. Such evidence is useful in managing the said tension, expectations, and designing TPDs that use learning outcomes to inform instructional strategies.

2. Review of literature

Accountability policies and practices increasingly focus on measurable performance outcomes, which, consequently, provides the conditions for teachers to be subjected to ways for quantifying their performance as well as giving them incentives to increase the test scores of students (Smith, 2016; Verger & Curran, 2014). Exams have been argued to motivate teachers (Glewwe et al., 2010) and give an objective indicator to evaluate school performance (Cilliers et al., 2021). For teachers, exams can set expectations and targets in various academic domains (Jones & Egley, 2004). However, a large (predominantly US) literature documents that high-stakes assessment systems may be inducing teachers to teach in ways that are not best for their students. In 2000, the American Educational Research Association issued a position statement defining high-stakes testing as assessment programs mandated by states and school districts to gather data on student achievement and hold schools and students accountable, where certain uses of test results have serious consequences, such as public praise or financial rewards for high school-wide scores, or students being held back in grade or denied a diploma for low scores (Marchant, 2004). Low-stakes tests, on the other hand, are assessments that have little to no impact on the test-taker's life opportunities or outcomes. Low-stakes tests are typically used for formative assessment, to provide feedback and support for learning, rather than for making high-stakes decisions (Penk, et al., 2014).

High-stake tests can produce curriculum narrowing by teachers (Berliner, 2011; Slomp et al., 2020), fragmentation, and teacher-centred pedagogies (Taylor, 2023). A report on Uganda suggested low coherence between teacher instruction and the curriculum, and a focus on rote learning. The implication was that teachers concentrate teaching on the narrow set of topics on

the exam rather than on the mastery of the broader content in the curriculum (Atuhurra & Kaffenberger, 2019). This reflects the observation by Slomp et al. (2020) that teachers often narrow their instruction and assessment practices to mirror the characteristics of the exam (including the scope of outcomes measured in the exam and type of items used). In addition, exams may lead teachers to focus their instruction on children whose learning improvements will likely improve the teachers' evaluation as educators (Gilligan et al., 2019). Exams can also lead to coaching (Neal, 2013), in which case the teachers focus class time on activities that would improve performance on a target exam without necessarily improving the mastery of a subject. There is also competition for instructional time with tested subjects, with non-testable subjects (like health education in the case of Michigan elementary schools) being given limited attention because they are viewed as secondary (Boguslawski et al., 2021).

Teachers have also reported that state testing programs have led them to teach in ways that contradict their own views of effective classroom instructional practice (Abrams et al., 2003). For example, in Sweden, Jonsson and Leden (2019) report some science teachers are reluctant to include argumentation skills in their teaching because this subject has not been part of their past pedagogical practice. Nevertheless, because they do not want students to be disadvantaged when doing the tests (as argumentation is a part of the national test), the teachers find ways to include only enough for the students to pass the test. Additionally, how teachers perceive their students may become inseparable from the test-scores of those students, making teachers normalise the use of deficit language such as "low" and "middle" achievers, which potentially affects how they relate with such students in the classroom (Datnow et al., 2018; Fisher-Ari et al., 2017; Schmidt et al., 2017).

Tests are noted to increase the teachers levels of stress and weaken their teaching morale (Collins, 2014; Wronowski & Urick, 2021), predicaments that affect classroom instructional practice. Nevertheless, Salloum and BouJaoude (2017, p. 855) note that teachers may see teaching to the test "as a moral duty as stakes are high for students who do not pass such exams; hence, it may be better for teachers to be supported in developing methods for balancing goals for conceptual learning and for test preparation". Thus, teachers may negotiate teaching to the test with teaching for understanding, through dialectic and implicit shifting among factual and conceptual knowledge on one hand and through making overt shifts toward specific instructions to aid students to crack the exam, on the other hand (Salloum & BouJaoude, 2017). This echoes the observation by Taylor (2023) of a teacher who, although critical of high-stakes testing, nevertheless appreciates 'the reality' of the test mandates and her limited ability to resist and disrupt the system. Thus, this teacher is "compelled to seek 'balance' between these dichotomies, positioning herself as responsible for managing the competing demands of neoliberal accountability (as evidenced by high-stakes testing demands) and the emancipatory pedagogies (alternatives to high-stakes tests) she sought to enact" (Taylor, 2023, p. 8). This teacher's pedagogical sense-making is significantly shaped by the neoliberal logics of high-stakes testing, and this is true for many teachers who operate within and resist the existing neoliberal order (Taylor, 2023).

Consequently, high stakes testing affects the teachers' perceptions of what to teach (curricular content) and teacher's assessment practices, which are often structured to reflect the breadth and depth of the high stakes test. Accordingly, high-stakes tests influence the content taught by teachers in the classroom to the extent of even teachers using the tests as a substitute for the state-wide curriculum (Jonsson & Leden, 2019). It should also be emphasised that teachers degree of teaching and assessing to the test compromises the validity of test scores, undermining the argument that test results are reflective of student ability with respect to the set of curriculum outcomes being tested by the exams (Slomp et al., 2020).

Though the studies cited in this literature review are mostly from the Global North, particularly the US, they offer invaluable insights into the influence of assessments on teachers' classroom instructional practices. Yet, in SSA, there is little known regarding how changes in assessments, for instance, those brought by the implementation of CBC in Kenya, are influencing the classroom instructional practices of teachers. To address this gap, we are embarking on an exploratory study that aims to enhance our understanding of the relationship between assessments and teaching behaviours. The insights gained from the exploratory study would inform the design of a larger study that represents a broader population of teachers in Kenya and other SSA jurisdictions adopting new educational approaches. The ultimate goal is to contribute to the existing body of knowledge and provide valuable insights for educational practices in SSA.

3. Study objectives

The broad objective of this study is to generate evidence that improves our understanding of the influence of assessments on teachers' teaching behaviour inside the classroom with a focus on SSA. Specifically, the study will:

1. Explore how assessments influence classroom instructional practices.
2. Assess why assessments influence (if at all they do) classroom instructional practices.
3. Examine the changes in teachers' classroom practices in response to the introduction of new tests in Grade 5 and Grade 6, as well as the phasing out of existing tests in Grade 8.

4. Research questions

To address the two research objectives, we shall seek answers to the following research questions:

1. How do assessments influence classroom instructional practices in primary school grades?
2. Why do assessments influence (if at all they do) classroom instructional practices in primary school grades?
3. How are teachers changing their classroom practices as new tests are introduced in Grade 5 and Grade 6, and existing tests phased out in Grade 8?

Conceptual Framework

Shih's (2009) model of test-washback on instructional practice and the classroom environment can help to think about the links between testing and classroom instructional practice. It is shown here to guide the research design.

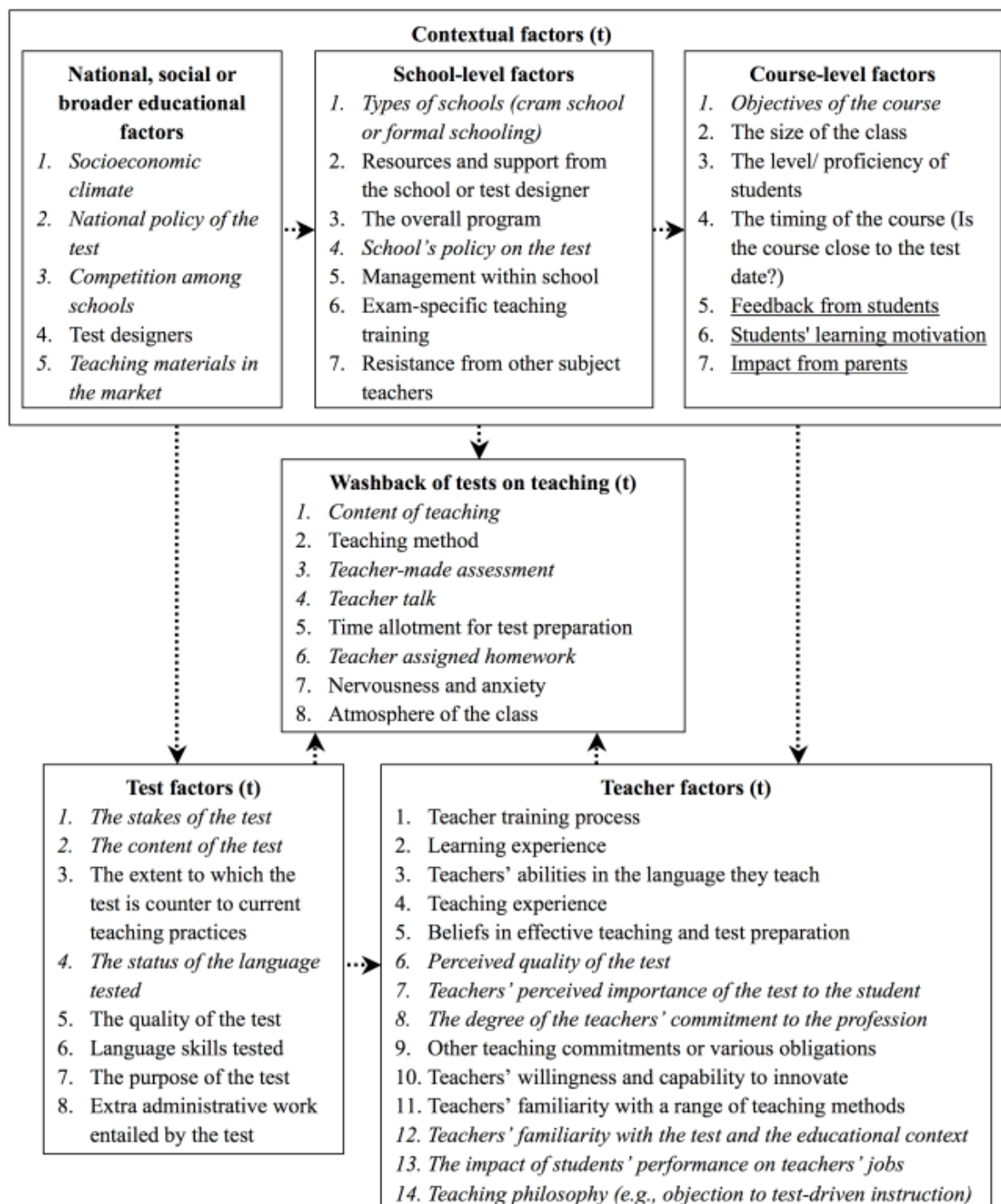


Figure 2: Test-washback on instructional practice

5. Methodology

6.1 Study design and sampling strategy

This study will employ the descriptive survey design, which aims to provide an accurate description of the existing situation as it is (Cohen, Manion & Morrison, 2007). Using this design will enable the researchers to conduct an exploratory study and collect primary data in order to investigate teachers' perspectives on the nexus between assessments and teaching behaviour.

Quantitative data will be collected to answer our research questions. As this is an exploratory study, survey questionnaires will be administered to 200 teachers spread across Grades 5, 6, and 8. This sample size is chosen conveniently based on budgetary and time constraints. While the sample size is relatively small, the intention of this study is not to generalize, but to understand what is happening in regard to the phenomenon, in order to inform the design of a larger study that will be representative of the larger population of teachers in Kenya. In other words, we wish to determine if there are any recurring patterns or trends of perceptions in a specific subset of teachers. By studying this subset, we aim to gather insights that can guide hypothesis generation. While we understand that these teachers are not a representative sample of the entire country, their experiences are valuable for initial analysis. If our findings indicate that there are no significant changes among these teachers, it would not allow us to make broad claims about the entire country. However, it would still serve as a starting point for generating hypotheses. On the other hand, if we observe a consistent and widespread response among these teachers, it would strongly suggest that further investigation and discussion are warranted.

Grade 5, 6, and 8 are purposely selected to allow for comparisons on how teachers' classroom instructional practices are influenced by assessments (grade 5: no high stakes test, grade 6: there is the new KPSEA based on CBC, and grade 9: there is the old high stakes KCPE).

The targeted teachers will be those teaching English and Mathematics in schools located in counties where APHRC conducts education research. Specifically, three counties will be selected: Nairobi, Kajiado and Embu.

Nairobi is the capital city of Kenya, located in the South-central part of the country. It is the largest city in Kenya and is home to various government offices, international organizations, and education institutions. Nairobi is an urban county with a diverse population, and its residents are a mix of different ethnic groups and cultures. Kajiado county borders Nairobi and to its South it borders the Tanzanian regions of Arusha and Kilimanjaro. It is largely semi-urban and has a mix of agricultural and pastoral communities. Its economy is driven by livestock keeping, tourism, and small-scale trade. Embu county is located in the South-eastern slopes of Mount Kenya. It is predominantly rural, its economy based on subsistence agriculture.

The selection of these three counties for the study will provide a level of variation in respondent background, which will add to the richness of the data collected in the exploratory study. In addition, the counties were chosen because APHRC already conducts education research in

these areas. Thus, the study will benefit from APHRC's established research infrastructure and networks in the region, facilitating efficient data collection and analysis.

English and Mathematics subject/learning areas are purposely chosen in response to the call to focus on foundational skills (such as literacy and numeracy skills), necessitated by the backlogs in learning, resulting from the disruptions that happened in the COVID-19 period in SSA (Herbert, et al. 2021). At least four teachers will be selected from a school taking into consideration gender and subject. Approximately 50 public primary schools will be targeted for this study (inclusive of possible non-response): 20 from Nairobi, 15 from Kajiado and 15 from Embu. Nairobi has been allocated more schools because of its big population of students (being a capital city) as compared to the other counties. In this county, schools will mainly be urban but serving various populations (upmarket as well as in the informal settlements). In Kajiado and Embu, we shall proportionally allocate urban/rural schools. A list of schools will be accessed from the MOE county officials for purposes of identifying actual location of schools and the proportion of rural/urban after we get the research permit from the National Commission for Science, Technology and Innovation. To be included in the study, a school must be a public school and must offer grades 5, 6, and 8. The list of schools from MOE county officials will also include information on grade 6 and 8 performance and the administrative locations of the schools. Using the performance league tables within a county, schools will be categorised as low performing (bottom 33%), medium performing, and high performing (top 33%). We shall ensure representation from each category of performance in our selection of the sample. In addition to distribution of sample by examination performance categories, administrative location (sub-counties and divisions) will be taken into consideration in the selection of the schools. Thus, the study population will be all primary school teachers who are teaching English and Mathematics in Nairobi, Kajiado, and Embu counties in Kenya. The following inclusion and exclusion criteria will be used to select the sample of teachers.

Inclusion criteria

- Primary school teachers teaching in public schools in Nairobi, Kajiado, and Embu counties in Kenya.
- Primary school teachers teaching English and Mathematics.
- Primary school teachers teaching grades 5, 6, and 8.
- Male and female primary school teachers (to include different gender perspectives).
- Primary school teachers with at least six months teaching experience.

Exclusion criteria

- Primary school teachers who have a known history of cognitive or mental impairment that could affect their ability to provide accurate responses.
- Primary school teachers who are currently on leave or are expected to be absent during the data collection period.

- Primary school teachers who are not currently employed by TSC as full-time primary school teachers (they may not have the same level of experience and exposure to the current teaching practices and assessments as full-time teachers).

If a school has a significantly higher number of teachers meeting the inclusion criteria than the available time to conduct interviews, random sampling will be conducted. This will involve numbering each of the eligible teachers and picking out of the hat until the desired number of respondents is reached. Once the potential respondents are identified, they will be contacted in-person and informed about the study and its objectives. The study team will thereafter seek the respondents' consent by handing them the consent forms to append their signatures as a confirmation of consent (see attached consent form). For the consenting, the researchers will suggest several time options and the participants will choose when they want the interview(s). A respondent(s)' preferred time will be used for the interview.

A. Data collection methods

We shall conduct an exploratory survey using a structured interview to solicit teachers' responses based on the research questions. A questionnaire with closed-ended questions will be used in this survey because these questions provide standardized responses that can be easily quantified and analysed (Dalati & Gómez, 2018). They are particularly useful when the researcher wants to gather specific information about a particular topic or variable (such as the influence of assessments on teachers' teaching behaviour) and when the researcher wants to compare responses across different participants or groups (Baburajan, et al., 2020). The specific questions in the research tool (see subsection 8.2) are informed by a preliminary review of the literature on the influence of assessments on classroom instructional practices that was conducted to guide this study. The closed-ended survey questions will solicit information on the teachers' social and professional backgrounds, their pre-CBC and CBC teaching and assessment practices, as well as how teaching style decisions are arrived at. The data collection instrument will be administered in person to the study respondents. The SurveyCTO, a mobile data collection software platform that allows users to create customized digital forms, collect and manage data in real-time using mobile devices, and export data to various formats for analysis, will be used for the data collection. The survey items will be sequenced in a flowing order, with each question worded in an easy-to-understand structure. As such, we will endeavour to frame the questions in a 'role play' way to enable the informants to respond in a first-person perspective. The language of communication during the data collection exercise will be English, which is the medium of instruction in Kenyan schools.

To ensure the effectiveness and efficiency of the data collection process, a pilot survey will be conducted before the actual data collection begins. The pilot study will be conducted with 30 teachers from Kiambu, a county neighbouring Nairobi and Kajiado counties. Conducting the pilot study in a different location from the study locations can reduce potential bias in the data collection process. This is because participants may be familiar with the survey items, which could introduce bias and influence their responses based on prior experiences or knowledge, leading to inaccurate or inconsistent data. By conducting the pilot study in a different location, the researchers aim to minimize this familiarity and increase the validity of the study. The data

collection tool will be revised based on the pilot survey feedback, and the final version will be used during data collection in the field.

The data collection process will involve recruiting 15 field interviewers, with five assigned to each of the three selected counties (Nairobi, Kajiado, and Embu). To be eligible for the position, field interviewers must have completed form four and attained a minimum grade of C in their Kenya Certificate of Secondary Education. Moreover, they must be residing within the target counties to ensure familiarity with the local environment and culture.

Once the field interviewers are selected, they will undergo a three-day training program on research ethics and data collection procedures. This will be followed by the actual data collection exercise after the necessary clearances and permissions have been obtained (see Ethical Considerations section below). The research team will approach the head teachers of the targeted schools to inform them about the study and the importance of their teachers' participation. The research team will seek an audience with the targeted teachers during a scheduled meeting to introduce the study, provide details about the study's procedures, and answer any questions that the teachers may have. This will be followed by inviting the targeted teachers to participate in the study and issuing consent forms to those who show interest. To account for the possibility of non-response from study participants, the researchers have decided to aim for 300 interviews, even though the targeted sample size is 200. That is to say, although we target 300 interviews, we will be okay with getting 200 interviews. The decision to increase the sample size by 50% of the initial 200 interviews to account for non-response is made in consideration of practical constraints (budgetary and time-related) and the aim to obtain meaningful results. Thus a total of 300 interviews will be conducted with teachers, distributed among 20 schools in Nairobi, 15 schools in Kajiado, and 15 schools in Embu. With 15 field interviewers conducting approximately 4 interviews per day, the researchers plan to conduct the 300 interviews over the course of five days. This approach will ensure that the researchers have enough data to answer their research questions, even if some study participants decline to participate or are unavailable.

Acknowledging the biases of self-reported data (Yoong, et al., 2013), this study is meant to reflect the issues that teachers tell us are most important regarding the changes that will be happening related to assessment and their classroom instructional practices. In that respect, although there might be bias, effort will be made to minimise the incentive to report with bias. We will seek to build trust and rapport with participants prior to the data collection exercise in order to increase the likelihood that they will provide honest and accurate information. We will assure the anonymity and confidentiality of the participants, as participants may be more likely to give honest responses if they feel that their answers will not be linked back to them personally (Brenner, et al. 1995; Durmaz, et al, 2020). It will be stated explicitly in the consent forms (discussed in the ethical considerations section below) and explained verbally that the participants responses will be kept confidentially and will not be linked to their identity. In addition, the questions asked during the data collection exercise will focus on short-term recall, which can mitigate inaccurate recall and enhance the reliability of self-reported data (Schrimshaw, et al., 2006).

B. Ethical considerations

The APHRC's internal Ethical Review Committee (ERC) will first review the study protocol and the research tool (see attachment), before submission to AMREF ESRC. After receiving ethical clearance from the AMREF ESRC, the research team will apply to the National Commission for Science, Technology, and Innovation (NACOSTI) for a research permit. In addition to seeking clearance from NACOSTI, the research team will obtain a letter of authorisation from the Principal Secretary of the Ministry of Education in Kenya to access the schools. We will take the letter to the County Directors of Education in Nairobi, Kajiado and Embu for us to enter the schools.

At the schools, permission and/or consent will be sought from all the targeted teachers before any data collection. Verbal consent will be sought from all the respondents from the first contact, which will be followed by written consent before the actual data collection and/or signed on-site/during data collection (the consent form is attached). During the consent process, the purpose of the study will be explained to each research participant, including the fact that participation will be voluntary. It will be explained that no compensation will be given to the participants. This decision is made to avoid any potential biases or undue influence that compensation may have on the participants' responses. It will also be explained that, although there are no direct benefits that the targeted respondents will get from participating in the study, the information gathered from this research can help to support teachers through informing professional development programs that use students' learning outcomes to inform classroom instructional strategies. It will also be made explicit that withdrawal is allowable, and that the respondents will not be penalised in any way upon declining participation.

The withdrawal procedures will be clearly outlined to all participants during the informed consent process. Participants will be informed that they have the right to withdraw from the study at any time and for any reason, without any consequences or penalties. They will also be informed that they can request the withdrawal of their data at any point during or after the study. If a participant chooses to withdraw from the study, their data will be excluded from any further analysis. However, any data that has already been collected and anonymized may be retained for analysis, as long as the participant has not explicitly requested for their data to be deleted. In case a study participant withdraws part way through the study, they will be given the option to withdraw their data up until the point of withdrawal. The researchers will also respect their decision to continue with the study but exclude any further data collection from that point on. The researchers will ensure that the participants' decision to withdraw does not affect their access to any services or benefits they are entitled to.

The data collection process will be conducted in private spaces to ensure the confidentiality and privacy of the participating teachers. The interviews will take place at the school, either in a secluded room or in an open space away from other teachers, students, or any other potential distractions.

Despite the above precautions, this study is cognizant of the fact that there might be emerging gaps, challenges, and opportunities, pertaining to the context of assessment and teaching across schools. These issues will be explored during the analysis of the data, to allow for constructive delivery of findings and recommendations for improvement to be made. In doing so, the study will highlight a chain of events that contributes to the identified gaps, challenges, and opportunities for assessment and teaching inside the classroom to enable duty-bearers to address the challenges.

Moreover, the study will embrace the ‘Code of Ethics and Professional Conduct’ of inclusivity, and be respectful, considerate, and devoid of a harassing tone. The report, while remaining factual, will also adopt a balanced tone that is not accusatory, or reflect the respondents in bad faith due to age, health [mental and physical ability], social and economic class, and training levels among others. In exploring the nexus between assessment and instructional delivery, the study will also adopt a fair reporting approach to teachers’ in/capabilities in using assessment data to make teaching decisions inside the classroom. During the data collection exercise, consent will be sort from the participants to use their phone numbers. If provided, the phone numbers, will only be used by APHRC staff and/or field staff contracted by APHRC to reach the participants for purposes of future follow up in regard to this research.

C. Data management and analysis

The quantitative survey data will be cleaned for possible missing values and/or inconsistencies. Thereafter labelling will be done and preliminary tables generated to inform further statistical analyses. Analytical techniques that employ small sample sizes to explore relationships/correlations will be used – e.g. using a limited number of variables (the key ones and/or theoretically important variables). We will also explore nonparametric tests (for instance Wilcoxon Rank sum test, Spearman correlation, Kruskal Wallis test, and Friedman’s ANOVA test). Although they are less robust, they are more suited to the possible non-normal distributions we are likely to encounter with our small sample size. Thus, inferential statistics will not be used. This is because our sample is not representative of the population (a requirement for inferential statistics to be meaningful). Our sample is geared to an exploratory study that will help us design a much bigger study after understanding what is going on, and what are the best lines of inquiry. However, this will not stop us from using regression techniques to explore associations/correlations (not causality). The [Statistical Product and Service Solutions](#) (SPSS) software will be employed to manage the data analysis. We will work closely with the Ministry of Education’s county officials to collect data and also validate our results. After the study, names and other personally identifiable information (PII) will be stored in secure locations (password-protected computers and secure cloud storage) accessible only to the authorised researchers. This storage will be maintained for a period of five years, after which the PII will be permanently discarded to maintain privacy. However, anonymized data, where all identifiable information is removed, will be retained beyond the five-year period. This decision is based on the potential usefulness and value of anonymized data for future analysis and research purposes, even after the specified timeframe.

7. Study's limitations and mitigation, and assumptions

Possible limitations of the study include:

- Biased information/data: the presence of a 'stranger' researcher/interviewer might bias participants' responses during interviews – social desirability. To forestall this limitation, we will inform the study respondents of the time and date of our interviews to enable them adequately prepare including selection of interruption-free or private space that will accord them the latitude to speak freely and honestly.
- Generalisation: Due to the small sample size and the use of convenient sampling methods, the generalizability of the findings will be limited. Nevertheless, the intention is not to claim that this survey represents a nationally representative sample. There is an ongoing and active policy debate surrounding the implementation of CBC and exam reforms. In light of this policy context, the findings from our study can still provide valuable insights and contribute to the ongoing discussions. While we acknowledge the inherent limitations and caveats, our study can shed light on areas that teachers have identified as more or less important. By considering their perspectives, we can offer a meaningful input to the policy discourse, allowing decision-makers to make more informed choices.

The assumptions for this study include that all the study participants will be successfully reached through onsite visits and that the interview conversations will go on successfully and uninterrupted to the end. To ensure that we achieve this, the respondents will be assured of anonymity and confidentiality of their responses and their institutional affiliations.

8. Findings dissemination strategy

After data analysis, a succinct report/working paper, a policy brief, and a blog will be developed to highlight the findings of the study. These knowledge products will be made available in soft copy in online platforms/websites for access by the education stakeholders and general public. The study stakeholders' attention will be drawn to the online documents via emails and social media (Twitter, LinkedIn & WhatsApp). APHRC, as well as the funder, will also disseminate the findings of the study in either international, regional and/or other national forums/conferences to sensitise scholars, education stakeholders (governmental and non-governmental), and teacher training practitioners on the impact of assessment on classroom-based instructional decisions. Additionally, the research team will also use the study findings to inform the development of related journal articles, opinion pieces, and fact sheet(s) that will be shared online and/or through print media for access by education stakeholders, and scholars.

9. Management and organization of the Study

At APHRC, the project will be led by a Principal Investigator (PI), Moses Ngware, who will provide oversight and intellectual guidance of the project. Moses will work closely and with support from Co-PIs including Jack Rossiter of Center for Global Development and Amani Karisa of APHRC. This core team will receive institutional support from the Director of

Research at APHRC for the smooth running of the project. The core team will share responsibilities to ensure successful completion of the project – e.g. responsibility for quantitative aspects of the project, etc. This core team will be supported by a technical team - whose functions will include data collection and policy engagement, and an administrative team - whose functions will include accounting, auditing, and logistics. Table 2 shows the workplan.

Table 2: Work Plan

Year 2023			Ja	Fe	Ma	Ap	Ma	Ju	Jul	Au	Se	Oc	No	De
		Milestone & Output	n	b	r	r	y	n		g	p	t	v	c
Inception Phase														
	Research protocol completion (narrative)	i. Research protocol approved by IRB												
	Research protocol completion (data collection tools)	ii. Research tools												
	Research Ethics application & research permit	iii. All participants identified (schools, community groups, etc)												
Data collection														
	Pilot survey													
	Revise tools based on pilot survey													
	Advertisement for and shortlisting of field interviewers	i. Clean datasets												
	Training of field interviewers	ii. Research report												
	Main study data collection	iii. A synthesis report based on the literature												
	Data processing (cleaning and analysis)													
	Report writing													
	Literature review report													
Dissemination Products														
	Writing a policy brief	i. One policy brief published												
	Writing two blogs (1. on why the study is relevant/important in Kenya 2. on findings of the study)	ii. Two blogs published												
	Conference attendance	iii. Two conferences attended and findings shared (poststudy period)												
	Writing a journal article	iv. One journal article published (post study period)												

Table 3: The Budget

Item	Description	Budget (USD)	Budget (KES*)
Personnel	Senior research scientist, project manager, research assistant, programmer, statistician	72,978	7,662,690
Ethical Protocol fees	AMREF fees NACOSTI fees	1,200	126,000
Research Activities	Training of field interviewers, main study data collection, data processing (cleaning and analysis), report writing	8,147	855,435
Dissemination	Writing a policy brief Writing a Blog Conference attendance (international and regional conference) Writing a journal article (publish in free access-hence Article Processing Cost)	9,500	997,500
Total		91,825	9,641,625

****1 USD=105 KES***

10. Research tool

See separate attachment

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.
- Amukune, S. (2021). Preschool education in Finland and Kenya: A comparison within perspectives of educational quality. *International Journal of Early Childhood Learning*, 28(2), 51-67.
- Atuhurra, J. & Kaffenberger, M. (2019). System (in)coherence seen through a curriculum lens: Ugandan teachers face conflicting demands from curriculum and examination bodies. https://riseprogramme.org/blog/system_incoherence_curriculum
- Baburajan, V., e Silva, J. D. A., & Pereira, F. C. (2020). Open-ended versus closed-ended responses: a comparison study using topic modeling and factor analysis. *IEEE Transactions on Intelligent Transportation Systems*, 22(4), 2123-2132.
- Berliner, D. (2011). Rational responses to high stakes testing: The case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education*, 41(3), 287-302.
- Berryhill, J., Linney, J. A., & Fromewick, J. (2009). The effects of education accountability on teachers: are policies too-stress provoking for their own good? *International Journal of Education Policy and Leadership*, 4(5), 1-14.
- Boguslawski, M. K., Lohrmann, D. K., Sherwood-Laughlin, C., Eckes, S., Chomistek, A. K., & Applegate, T. (2021). Elementary school personnel and cultural factors affecting health education implementation in the high-stakes testing era. *The Journal of School Health*, 91(10), 846-856. <https://doi.org/10.1111/josh.13071>
- Brener, N. D., Collins, J. L., Kann, L., Warren, C. W., & Williams, B. I. (1995). Reliability of the youth risk behavior survey questionnaire. *American Journal of Epidemiology*, 141(6), 575-580.
- Cilliers, J., Mbiti, I. M., & Zeitlin, A. (2021). Can public rankings improve school performance? Evidence from a nationwide reform in Tanzania. *Journal of Human Resources*, 56(3), 655-685.
- Cohen, L., Manion, K., & Morrison, L. (2007). *Research methods in education* (6th ed.). Routledge Taylor & Francis Group.
- Collins, C. (2014). Houston, We Have a Problem: Teachers find no value in the sas education value-added assessment system (EVAAS®). *Education policy analysis archives*, 22(98), n98.
- Dalati, S., & Gómez, J. M. (2018). Surveys and questionnaires. In Gómez, J. M., & Mouselli, S. (Eds.) *Modernizing the Academic Teaching and Research Environment: Methodologies and Cases in Business Research*, 175-186. Springer.
- Dalton, E. M., Mckenzie, J. A., & Kahonde, C. (2012). The implementation of inclusive education in South Africa: Reflections arising from a workshop for teachers and therapists to introduce Universal Design for Learning. *African Journal of Disability*, 1(1), 1-7.

- Datnow, A., Choi, B., Park, V., & John, E. S. (2018). Teacher talk about student ability and achievement in the era of data-driven decision making. *Teachers College Record*, 120(4), 1-34.
- Durmaz, A., Dursun, İ., & Kabadayi, E. T. (2020). Mitigating the effects of social desirability bias in self-report surveys: Classical and new techniques. In M. L. Baran, & J. E. Jones. (Eds.). (2019). *Applied social science approaches to mixed methods research* (pp. 146-185). IGI Global.
- Fisher-Ari, T. R., Kavanagh, K. M., & Martin, A. (2017). Urban teachers struggling within and against neoliberal, accountability-era policies. *Penn GSE Perspectives on Urban Education*, 13(2).
- Gilligan, D. O., Karachiwalla, N., Kasirye, I., Lucas, A. M., & Neal, D. (2022). Educator incentives and educational triage in rural primary schools. *Journal of Human Resources*, 57(1), 79-111.
- Glewwe, P., Ilias, N., & Kremer, M. (2010). Teacher incentives in Kenya. *American Economic Journal: Applied Economics*, 2(3), 205-227.
- Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, 33(2), 14-18.
- Halász, G. (2017). The spread of the learning outcomes approaches across countries, sub-systems and levels: A special focus on teacher education. *European Journal of Education*, 52(1), 80-91.
- Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational research review*, 17, 50-62.
- Herbert, A., Saavedrale, A., Marr, A. & Jenkins, R. (2021). The urgent need to focus on foundational skills. [The urgent need to focus on foundational skills \(worldbank.org\)](https://www.worldbank.org/)
- Herman, K. C., Reinke, W. M., Dong, N., & Bradshaw, C. P. (2022). Can effective classroom behavior management increase student achievement in middle school? Findings from a group randomized trial. *Journal of Educational Psychology*, 114(1), 144.
- Imana, D. K. (2020). The politics of education reforms in Kenya: Critical assessment of the education system from 1963-2020. *Jurnalul Practicilor Comunitare Pozitive*, 20(2), 11-30.
- Jones, B. D., & Egley, R. J. (2004). Voices from the frontlines: Teachers' perceptions of high-stakes testing. *Education Policy Analysis Archives*, 12(39), n39.
- Jonsson, A., & Leden, L. (2019). The ambiguous influence of high-stakes testing on science teaching in Sweden. *International Journal of Science Education*, 41(14), 1926-1943. <https://doi.org/10.1080/09500693.2019.1647474>
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772-775.

- Kavua, M. M. (2020). Educational reforms in Kenya. In *Oxford Research Encyclopedia of Education*. <https://doi.org/10.1093/acrefore/9780190264093.013.435>
- Koh, K., Hadden, J., Parks, C., Sanden, L., Monaghan, L., Gallant, A., & LaFrance, M. (2015). Building Teachers' Capacity in Authentic Assessment and Assessment for Learning. University of Calgary. [content \(ucalgary.ca\)](https://content.ucalgary.ca)
- Kretzer, M. M., & Oluoch-Suleh, E. (2022). (Hidden) potentials for African languages in curriculum reforms: examples from Kenya and South Africa. *SN Social Sciences*, 2(154), 1-24. <https://doi.org/10.1007/s43545-022-00440-6>
- Marchant, G. J. (2004). What is at stake with high stakes testing? A discussion of issues and research. *Ohio Journal of Science*, 104(2), 2-7.
- Ministry of Education (2022). CBC Implementation Bulletin, Vol 1. Nairobi: MoE.
- Neal, D. (2013). The consequences of using one assessment system to pursue two objectives. *The Journal of Economic Education*, 44(4), 339-352.
- Nseibo K, Vergunst R, McKenzie J, Kelly J, Karisa A and Watermeyer B (2022) Learning together: The dynamics of training teachers for disability inclusion in South Africa. *Front. Educ.* 7(999814). doi: 10.3389/feduc.2022.999814
- Penk, C., Pöhlmann, C., & Roppelt, A. (2014). The role of test-taking motivation for students' performance in low-stakes assessments: An investigation of school-track-specific differences. *Large-scale Assessments in Education*, 2(5), 1-17.
- Rono, K. K., & Bwamoni, C. O. (2020). Implementation of Kenya's basic education curriculum framework: a pursuit of quality education for social economic and political growth. *The International Journal of Humanities & Social Studies*, 8(11). DOI: 10.24940/theijhss/2020/v8/i11/HS2011-036
- Salloum, S., & BouJaoude, S. (2017). The use of triadic dialogue in the science classroom: a teacher negotiating conceptual learning with teaching to the test. *Research in science education (Australasian Science Education Research Association)*, 49(3), 829-857. <https://doi.org/10.1007/s11165-017-9640-4>
- Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B., & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, 103, 101602.
- Schmidt, R., Jacobs, M. M., & Meyer, H. (2017). Sociopolitical testing discourses in elementary teachers' talk about reading assessment. *English Teaching: Practice & Critique*, 16(3), 391-406.
- Schrimshaw, E. W., Rosario, M., Meyer-Bahlburg, H. F., & Scharf-Matlick, A. A. (2006). Test-retest reliability of self-reported sexual behavior, sexual orientation, and psychosexual milestones among gay, lesbian, and bisexual youths. *Archives of sexual behavior*, 35, 220-229.

- Slomp, D., Marynowski, R., Holec, V., & Ratcliffe, B. (2020). Consequences and outcomes of policies governing medium-stakes large-scale exit exams. *Educational assessment, evaluation and accountability*, 32(4), 431-460.
- Smith, W. (2016). An introduction to the global testing culture. In W. Smith (Ed.), *The global testing culture: Shaping education policy, perceptions, and practice* (pp. 7-24). Symposium Books.
- Taylor, L. A. (2023). "If I was better at managing all this": The role of neoliberal logic in framing one teacher's narratives about accountability. *Teaching and teacher education*, 121. <https://doi.org/10.1016/j.tate.2022.103944>
- United Nations Sustainable Development Goals (2-15). Goal 4 | Department of Economic and Social Affairs (un.org)
- Verger, A., & Curran, M. (2014). New public management as a global education policy: Its adoption and re-contextualization in a Southern European setting. *Critical studies in education*, 55(3), 253-271.
- Wanjala, D. N. (2021). Assessment of funding mode as an integrated education management strategy on pupils' participation in public primary schools in kenya. *The International Journal of Humanities & Social Studies*, 9(4), 105-116. <https://doi.org/10.24940/theijhss/2021/v9/i4/HS2104-018>
- Wronowski, M., & Urick, A. (2021). Teacher and school predictors of teacher deprofessionalization and demoralization in the United States. *Educational Policy*, 35(5), 679-720.
- YanYan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., & Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, 28(3), 228-260.
- Yoong, S. L., Carey, M. L., D'Este, C., & Sanson-Fisher, R. W. (2013). Agreement between self-reported and measured weight and height collected in general practice patients: a prospective study. *BMC medical research methodology*, 13, 1-8
- Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: A literature review. *Education policy analysis archives*, 18, 19-19.