Published at https://ijate.net/   https://dergipark.org.tr/en/pub/ijate   *Research Article*

Assessment tools and strategies used by Jamaican secondary school teachers

Clavia T. Williams-McBean [1,*]

[1]University of the West Indies, Mona Campus, Faculty of Humanities & Education, School of Education

Abstract: There is an increasing understanding that assessment is an integral part of teaching and learning and that teachers are largely not adequately prepared for their assessment responsibilities. Consequently, there is a need for research on what teachers need to improve their assessment practices. To determine what Jamaican secondary school teachers need, this mixed methods study was conducted to describe the assessment tools and strategies used by secondary school teachers of various subjects and in different types of schools as the basis for future interventions. Data was collected from a survey of 1088 secondary school teachers of varying subjects and school types and further explored through interviews and observations of 32 teachers of English. Analysis of the data using descriptive statistics and ANOVA in the quantitative phase revealed that secondary school teachers primarily used traditional assessment tools and strategies, particularly tests, despite school type. Pattern coding and pattern matching in the qualitative phase confirmed these results. The findings also revealed statistically significant differences in the frequency of use of traditional and alternative assessment tools and strategies based on the subject the teachers taught. Qualitative explorations revealed that school policies that require a quota of grades and state or express positive attitudes towards tests influenced teachers despite school type to use traditional methods. The findings imply that school administrators need to implement supportive school-level policies and display positive attitudes toward alternative assessments to maximize the use of assessment to improve learning.

1. INTRODUCTION

Assessment has been given international attention in recent times as the need for educational accountability increased (Kubiszyn & Borich, 2013), and as the use of assessment to improve student learning (i.e., formative assessment) has been promoted, investigated, and reported (Monteiro et al., 2021). Because of the promise of assessment, particularly formative assessment, in improving student learning, there have been worldwide efforts to improve teachers' assessment knowledge and skills. However, repeated reports have confirmed that teachers' knowledge and skill in engaging in effective assessment practices that can make the promise of assessment a reality need improvement (Acar-Erdol & Yıldızlı, 2018; Sewagegn, 2019; Organisation for Economic Cooperation and Development [OECD], 2019). Research has also indicated the need for empirical studies on what teachers need to improve their assessment

practice (Jiang, 2020). To determine what teachers need, it is crucial to determine, understand and describe current assessment practices. Understanding where the teachers are makes efforts to determine and take them where they are supposed to be more effective.

In the Jamaican context, The National Education Inspectorate (NEI) has repeatedly reported that teachers' use of assessment needs improvement (National Education Inspectorate [NEI], 2013; 2014; 2015; 2016; 2017). Additionally, there is a dearth of empirical research on assessment in Jamaica, in general, and the formative use of assessment in the Jamaican context (Williams-McBean, 2021). I found only one paper on the assessment strategies Jamaican teachers used, and it focused on only one of the seven education regions in Jamaica (Onyefulu, 2018). Additionally, Onyefulu (2018) reported that she could find no published article on Jamaican teachers' classroom assessment practices. Therefore, I embarked on the Spotlight on Assessment in Jamaica Project (SAAJP). The project aims to study existing assessment principles and practices in Jamaican schools in all seven regions, use the information to determine what teachers need to improve their assessment practices, then design and implement interventions to improve assessment policies and practices in Jamaica. The premise is that understanding what exists will increase the effectiveness of later interventions to improve teaching and learning through assessment. This paper is the first in a series that shares the results of the first phase of the project that describes the existing nature of assessment in secondary schools across Jamaica. In describing the existing nature of assessment, I focused on what assessment tools and strategies secondary school teachers used, how they used them, and, the factors that influenced their choice of assessment. However, because of the extensiveness of the data, this paper only focused on the assessment tools and strategies used by secondary teachers. Subsequent papers will report the findings on the other two areas of focus. In seeking to describe the assessment tools and strategies used by Jamaican secondary school teachers, I sought answers to the following research questions:

1. What assessment tools and strategies do Jamaican secondary school teachers use most frequently?
2. What is the difference in teachers' reported frequency of use of the different types of assessment tools and strategies based on subject?
3. What is the difference in teachers' reported frequency of use of the different types of assessment tools and strategies based on school type?

1.1. Assessment Tools and Strategies Used by Teachers

The 2014 Standards for Educational and Psychological Measurement distinguishes 'test' from 'assessment' by outlining that 'tests' refer to "scales, inventories, pen-and-paper tests, orals, free-format responses, and authentic assessments" (American Educational Research Association [AERA] et al., p. 2) and defining 'assessment' as "a process that integrates test information with information from other sources (e.g., information from other tests, inventories, and interviews; or the individual's social, educational, employment, health, or psychological history)" (AERA et al., p. 2). These definitions indicate that the 'test' is the instrument (traditional or alternative) that is used to measure learning. I agree with this definition despite the lack of consensus on the definition and specific differences between the two terms among researchers in the field (see, for example, Kubiszyn & Borich, 2013; Miller et al., 2013; Popham, 2018). However, since most people (including the respondents in this research) associate tests with the traditional pen-and-paper, one-shot examinations, I used it in that way for shared understanding. The term assessment tools and strategies refer to all the testing tools and techniques used to provide the measurement and qualitative data used in the assessment process. Assessment refers to the process by which the measurement and/or qualitative data on the nature and extent of students' learning are used by teachers, students, or administrators for formative, summative, and evaluative purposes.

Assessment tools and strategies are differentiated by their format (traditional or alternative, testing or performance), purpose (diagnostic, formative, summative, evaluative), location (internal or external, classroom or standardised), their relative weight or importance (high stakes or low stakes) and the interpretation of the results (norm-referenced and criterion-referenced) (Acar-Erdol & Yıldızlı, 2018; Kubiszyn & Borich, 2013). In this research, the focus is on the format of the assessment used in the classroom by teachers of varying subjects across Jamaica. In terms of format, assessment tools and strategies are classified as traditional or alternative. Traditional classroom assessment models the format and administration of standardised, public examinations and refers to pen-and-paper examinations that usually utilise items such as multiple-choice, true/false, matching, short answers, and essays (Dikli, 2003; Gronlund, 2006, Koh, 2017; Miller et al., 2013) although some writers classify essays as performance assessment (Wren & Gareis, 2019). In contrast, alternative assessment methods include authentic and open-ended performance assessment that requires students to use or apply their knowledge and skills while performing a task in a realistic setting. It also requires direct observation of the performance by the assessor, who uses a rubric to evaluate the quality of the performance (Brookhart, 2009). Examples of alternative assessment strategies include performances, concept maps, open-ended questions, interviews, exhibits, presentations, oral and practical demonstrations, hands-on execution of experiments, simulations (with or without the use of computers), observations, student journals, peer-assessment, self-assessment, projects, and portfolios (Adeyemi, 2015; Berry, 2008; Bland & Gareis, 2018; Dandis, 2013).

In education, over 30 years of research have reported that traditional assessment tools and strategies have dominated (Brookhart, 2013; Esomonu & Eleje, 2020; OECD, 2019; Stiggins & Conklin, 1992). However, there has been increasing advocacy for the increased use of alternative assessment methods. This advocacy is based on research results that alternative assessments impact more positively on students' intrinsic motivation and engagement than traditional assessments (Hess et al., 2020; Koh, 2017); promote and measure affective learning (Koh, 2017); more effectively allow for formative assessment (Black & Wiliam, 1998; Koh, 2017); have greater authenticity (Wren & Gareis, 2019;); and, that they are focused on deeper learning and higher-order thinking skills (Koh, 2017; Wren & Gareis, 2019;). This shift has accompanied the shift from behaviourism to constructivism and from a focus on summative assessment to formative assessment (Buhagiar, 2007; Dogan, 2011; Koh, 2017).

At the same time, some writers have taken a "middle of the road" stance. They argue that both are useful and should be used in conjunction to get the most accurate picture of student achievement (Popham, 2005; Wren & Gareis, 2019). In explaining his support for what he calls "balanced assessment," Burke (2009) posits that it should include three types of assessment: traditional (focusing on knowledge, curriculum, and skills), portfolio (process, product, and growth), and performance (standards, application, and transfer). In this way, a more comprehensive range of student skills is measured, and a more valid assessment of student achievement can be made.

The more positive impact of alternative assessment indicates that improved educational outcomes can result from its use. However, most of the studies reviewed found that despite the pedagogical shifts and the curricula rewrites, teachers' assessment practices at the secondary level have remained predominantly traditional, with tests being the most frequently used type of assessment (Acar-Erdol & Yıldızlı, 2018; Bramwell-Lalor, 2019; Brookhart, 2013; Dandis, 2013; Esomonu1 & Eleje, 2020; Guskey & Link, 2019; Saefurrohman, 2017; Vlachou, 2018). Berry (2010) reported that even when teachers used strategies or tools labelled as alternative assessments, for example, projects, their objective was to measure lower-order thinking skills and to assess knowledge acquisition and retention. Other researchers also found that elementary teachers used more varied assessment methods, including informal evidence and observations,

while secondary teachers used paper-pencil objective tests, whether commercially prepared, teacher-made, or derived from textbooks (Brookhart, 2009, 2013; Guskey & Link, 2019; Ong, n.d.; Vlachou, 2018; Zhang & Burry-Stock, 2003).

I found only one study investigating teachers' assessment practices in the Jamaican context: Onyefulu (2018). Onyefulu (2018) surveyed 157 primary and secondary school teachers in Region 1 in Jamaica and confirmed testing dominance with 51% of the primary school teachers and 85% of the secondary school teachers surveyed reporting that they most frequently used closed-book tests to assess their students. However, there were only eight assessment methods included on the research instrument. Five of the eight were a type of test (closed book test, open-book test, collaborative or negotiated test, cooperative testing, and take-home test). The other three methods were portfolio assessment, peer-assessment, and self-assessment. This research includes the reported frequency of use of 22 assessment tools and strategies from secondary school teachers from all seven educational regions in Jamaica. Additionally, since all except one of the studies reviewed were conducted outside of Jamaica and none included teachers from across the country, it was prudent to investigate if the same obtained in Jamaica. Nevertheless, these studies helped identify various assessment tools and strategies and classify them as traditional or alternative.

1.2. The Difference in Assessment Tools and Strategies Used Based on Subject

Most of the studies reviewed focused on the assessment tools and strategies used in a single subject. Therefore, they did not allow for comparisons across subjects. This inclusion is another way in which this study contributes to the existing body of literature. Additionally, among the studies reviewed that included different subjects, the findings are contradictory Some researchers reported that teachers of Mathematics indicated that they used alternative assessment methods with greater frequency than all other subject areas (Bol et al., 1998) or more than teachers of language arts, science, and social studies (Zhang & Burry-Stock, 2003). Bol et al. (1998) explained that the greater use of alternative assessment tools and strategies resulted from the Mathematics teachers' greater focus on process than product. On the other hand, researchers have found that teachers of Mathematics use predominantly traditional assessment tools and strategies (Dandis 2013; Senk et al., 1997; Watt, 2005).

The contradiction is evident for other subjects as well. For example, Zhang and Burry-Stock (2003) corroborated Marso and Pigge's (1988) study and reported that "language-arts teachers used paper-pencil tests more often than did teachers in nonacademic subjects" (p. 333). McMillan (2001) also reported that English teachers reported more frequent use of constructed-response assessment strategies than both mathematics and science teachers. Constructed response items include essays, which may be classified as traditional assessment. The term, however, also includes alternative assessment tools and strategies. Therefore, it is unclear what type of assessment (traditional or alternative the teachers in this study were using. Furthermore, Brookhart (2009) reported that teachers of Social Studies used traditional assessment (constructed-response items) more frequently than all other subjects. Berry (2010) corroborated the difference in assessment tools and strategies based on subject. However, she did not assess which subject area had a greater propensity toward what type of assessment tool or strategy. She did, however, establish that subject content played a role in the assessment tool and strategy selection of the participants in her study. According to Berry (2010), teachers reported using alternative assessment strategies if the content was "activity-based" (p. 104). To add to the contradiction, Duncan and Noonan (2007) and Ong (n.d.) reported no difference based on the subjects taught. Therefore, the results on the difference in assessment tools and strategies used by secondary school teachers based on subject are conflicting and worthy of further investigation, especially in the Jamaican context where this area is mainly unaddressed. Consequently, it was an area of focus in this research.

1.3. Difference in Assessment Tools and Strategies Used based on School Type

It is essential to consider school-type differences in Jamaica because there are grave disparities in student academic achievement among the different types of secondary schools: traditional high schools for boys, traditional high schools for girls, coeducational traditional high school, upgraded high schools, and technical high schools (Clarke, 2011; Williams-McBean, 2021). Top performers in the primary-level exit examinations are usually placed in traditional high schools. As students' academic achievement (as measured by the exit examinations) decreases, they are placed in upgraded high schools and technical high schools (Clarke, 2011; Williams-McBean, 2021). However, individual upgraded and technical high schools outperform some traditional high schools, and technical high schools outperform some upgraded high schools. In addition, research purport that the use of alternative assessment can increase student achievement (e.g., Guha et al., 2018). Since school type and academic achievement are so interconnected in the Jamaican context, investigating the types of assessment used in the different types of schools would be useful. Therefore, this research disaggregated schools based on the five types of secondary schools in Jamaica.

2. METHOD

Data was collected using a multiphase mixed methods design, which began with a quantitative phase, followed by a qualitative phase, followed by an intervention phase. However, the data presented in this paper are from the first two phases.

2.1. The Quantitative Phase

In this phase, the researcher surveyed 1,088 secondary school teachers on the types of assessment tools and strategies used, the frequency of use, the factors that influenced their choice of assessment tools and strategies, and the types of feedback they give to students.

2.1.1. *The sample*

The quantitative sample consisted of 1,088 secondary school teachers from 45 secondary schools across Jamaica. The schools were ranked (above average, average and below average) based on a three-year average of students' performance in Caribbean Secondary Examinations Certificate English A examinations – the exit examination for English at the secondary level. Therefore, the schools were stratified according to school type and rank and a sample was selected using proportionate, stratified random sampling. Of the 1,088 teachers, 587 or 54% teachers were from upgraded high schools, 213 or 19.6% from coeducational traditional high schools, 60 or 5.5% from traditional high schools for boys, 100 or 9.2%, from traditional high schools for girls, and 128 or 15.5% from technical high schools. The quantitative sample consisted of male and female teachers with varying years of experience who reported teaching various subjects categorized into nine different groups: English (English Language, English Literature, Communication Studies), Mathematics, Social Sciences (e.g., Social Studies, Religious Education, History), Sciences (e.g., Biology, Chemistry, Physics, Integrated Sciences), Business (e.g., Principles of Business, Principles of Accounts, Office Administration, Information Technology), Practical Arts (e.g., Physical Education, Woodwork, Electrical and Electronic Technology, Food and Nutrition), Performing Arts (e.g., Dance, Drama, Art), Modern Languages (Spanish and French) and Mixed (a combination of any of the categories) (see Table 1). The disproportionality within each sample variable represents the disparity that exists in the teacher population of Jamaica. There was a 95% overall response rate.

Table 1. *The quantitative sample.*

Sample Characteristic		N	%
Gender	Male	325	31
	Female	726	69
Age	Young adult	149	18
	Middle-aged	913	82
Years of Experience	0 – 5 years	275	216
	6 – 10 years	328	32
	11 – 15 years	163	16
	16 – 20 years	112	11
	≥ 20 years	154	15
School Type & Rank	**Traditional High school (Coed)**	**213**	**20**
	Above Average	47	4
	Average	85	9
	Below Average	81	7
	Traditional High school (Boys)	**60**	**6**
	Above Average	20	2
	Average	20	2
	Below Average	20	2
	Traditional High school (Girls)	**100**	9
	Above Average	33	3
	Average	33	3
	Below Average	34	3
	Upgraded High School	**587**	**54**
	Above Average	195	18
	Average	196	18
	Below Average	196	18
	Technical High School	**128**	**11**
	Above Average	37	3
	Average	52	5
	Below Average	39	3
Subject	English	191	18
	Mathematics	132	13
	Social Sciences	177	17
	Sciences	115	11
	Business	119	11
	Practical Arts	175	17
	Performing Arts	34	3
	Modern Languages	43	4
	Mixed	60	6

2.1.2. *Quantitative data collection method*

A self-developed Teacher Assessment Practices Questionnaire was used to collect data in this phase. The questionnaire was developed by relying heavily on the literature (e.g., Alkharusi, 2011; Berry, 2010; Dandis, 2013). The questionnaire contained 41 questions that were divided into four sections. The first section presented four items to capture demographic details that researchers identified as influencing teachers' choice and frequency of use of different assessment tools and strategies: gender (Alsarimi, 2000); age; years of service (Alkharusi, 2011), subject(s) taught (Alkharusi, 2011; Berry, 2010; Dandis, 2013). The second section

consists of assessment strategies and techniques scale: 22 items on a 6-point Likert scale ranging from *Don't know* (to be selected if the respondent does not know the assessment strategy) to *Frequently use*d. Each assessment tool or strategy was identified in previous studies. The tools and strategies were also classified as traditional or alternative based on Gronlund's (2006) specification. To increase the clarity of the items, the method 'Test' was used to refer to the traditional pen-and-paper test because that is how it is understood by the respondents, and (multiple-choice, true/false, matching, short answers) were included in brackets to clarify further. These item formats were identified in previous literature as items commonly used on traditional tests (see, for example, Koh, 2017; Miller et al., 2013). Additionally, putting clarifying terms in brackets after a concept is recommended by Cobern and Adams (2020) as part of the basic steps to instrument validation. Though essays are also frequently used on pen-and-paper tests, it was separated because some writers classify essays as performance assessment (e.g., Wren & Gareis, 2019). The separation allowed for more specific identification and examination of teachers' frequency of use of traditional as differentiated from alternative assessment tools and strategies. Section three consisted of 19 items on a 4-point Likert scale ranging from *Least influential* to *Extremely influential*) that listed factors, also identified from the literature, that influenced teachers' choice of assessment tools and strategies. Section four consisted of one item with five types of feedback: Grades (e.g., 70%, 9/10, B+), Ticks and Xs, Oral feedback, Written feedback on students' strengths and weaknesses and Grades accompanied by written feedback. These types of feedback were identified from the literature, and the respondents were required to select the type of feedback they most frequently gave their students.

The validity and reliability of the instrument were assured using data from two pilot studies, member checking, expert checking and a literature-validated theoretical model. According to Cobern and Adams (2020), theoretical models where the researcher uses the literature to create a model on which the survey instrument is developed "provide the first line of validity evidence for the survey" (p. 408). The selected demographic details, the assessment tools and strategies and their classification as traditional or alternative, the factors and types of feedback were all derived from the related literature. In addition, experts in quantitative research and educational assessment checked the questionnaire for content validity since expert checks are the best way to ensure content validity (Zohrabi, 2013). The educational assessment experts, who had at least a master's degree in educational measurement and taught in the area, confirmed the grouping of the tools and strategies as traditional and alternative and suggested no change to the instrument. The experts also affirmed the logical groupings of the individual factors into three categories: Student Factors (students' grade level, students' academic abilities, students' behaviour, students' motivational levels, number of students in the class, number of students in the school and expectations of the students' parents), Teacher Factors (formal teacher training, teachers' experiences as teachers, teacher's experiences as learners, teachers' knowledge of current research, teacher content knowledge, and Assessment Factors (the format of standardized tests (e.g. CSEC), availability of past papers, workload of the assessment strategy, national assessment practices, the school's assessment policy, time constraints and the demands of the national curriculum). No additions were suggested by the 10 secondary school teachers, including five heads of department, who were interviewed about the clarity and completeness of the tools and strategies, factors and types of feedback They suggested adding 'please turn over' on the first page of the instrument and increasing the spacing. Both suggestions were implemented before the questionnaire was administered to the main sample. Expert and respondent feedback was also used to ensure face validity (Oluwatayo, 2012).

The reliability of the instrument was assessed using Cronbach's alpha in SPSS. The two sub-scales in Section 2 had acceptable reliability of .65 for frequency of use of traditional assessment tools and strategies and .83 for frequency of use of alternative assessment tools and

strategies. The subscales in Section three also had acceptable reliability of .73, .60 and .71 for Student Factors, Teacher Factors, Assessment Factors, respectively. Though alpha of or greater than .70 is usually considered acceptable, researchers have also purported that alphas of .60 are acceptable (Churchill Jr. & Peter, 1984; Taber, 2018) especially for newly developed measures (Nunnally, 1978, 1988).

2.1.3. *Quantitative data analysis and presentation*

To answer the first research question: (*What assessment tools and strategies do Jamaican secondary school teachers use most frequently?*), I calculated the mean score of the non-missing values for each assessment strategy (Bryman & Cramer, 2011). Then, using Gronlund's (2006) specifications, I categorised the 22 individual assessment tools and strategies as "Traditional Assessment Strategies" and "Alternative Assessment Strategies" in SPSS. The traditional assessment tools and strategies were tests, questioning, oral quizzes, teacher observation, and essays. Concept maps, checklists, flow charts, peer evaluations, portfolios, speech/debate/drama, case studies, research reports, rubrics, self-evaluations, practical tests, role plays, student journals, contracts, conferences, anecdotal records, and interviews were categorized as alternative assessment tools and strategies. Descriptive statistics were then used to answer the question in the quantitative phase. To respond to questions 2: (*What is the difference in teachers' reported frequency of use of the different types of assessment tools and strategies based on subject?*), a one-way between-groups ANOVA with a post-hoc test was done as all the subjects were collapsed into nine categories: English, Mathematics, Social Sciences, Science, Business, Practical Arts, Performing Arts, Modern Languages and Mixed. Finally, a two-way between-group analysis of variance was used to assess the difference in teachers' reported frequency of use of the different types of assessment tools and strategies based on school type. This technique was suitable because, in this study, school type referred to the type of school (traditional, technical, upgraded) as well as the rank of the school (above average, average, below average). The results of the quantitative phase were used to select the sample for the subsequent qualitative phase.

2.2. The Qualitative Phase

2.2.1. *Research design*

Qualitative data was also collected to answer the research questions and to add depth to the research. I observed teachers in their natural settings (to determine if they used the same assessment tools and strategies, and with the same frequency, as they had reported — to add credibility to the quantitative findings and the overall conclusions from this study. A multiple-case instrumental case study design (Creswell, 2014; Yin, 2014) was used in this research phase. The cases (teachers) were embedded within the context of the schools, and they were deliberately selected to unearth different perspectives about the issue of teachers' assessment practices. Hence, they were "instrumental cases" (Creswell, 2014, p. 493).

2.2.2. *The participants*

I selected the participants in the qualitative phase through stratified purposive sampling (Patton, 1990). The six schools I selected were stratified from the quantitative phase by school type and rank. The quantitative findings showed no difference in the frequency with which teachers used traditional or alternative assessment tools and strategies based on school type. Therefore, I reduced the number of school types represented from five to three. However, I maintained the three major types – traditional, upgraded, and technical – to explore possible school type differences qualitatively. I randomly selected two schools from the three different school types retained. Of the two schools, one was from the above average rank and the other from the below average. Five or six language teachers from each selected school were observed and interviewed to explore further the methods of assessment teachers used and explain the quantitative findings.

I selected all the teachers from each school who had participated in the initial survey and were willing to continue into the qualitative phase. I interviewed all the teachers of English in each school even after saturation was achieved. After these schools had been selected, one teacher from an average, traditional high school for boys requested to continue participating in the study. Therefore, 32 teachers of English, two males, and 30 females from four types of schools, with varying years of experience, were interviewed and observed. I selected the English department for further investigation because it is the area in which I am most knowledgeable and skilled, having been a teacher and researcher of issues in English Language and Literature education at the secondary level for approximately nine years. It was also the area in which the formative assessment intervention was to be subsequently implemented. The English group also represented the largest subject group from the quantitative sample: 191 or 18%. Consequently, while the qualitative findings provide useful insights into why teachers from different school types predominantly used traditional assessment tools and strategies, the specific findings are reflective of the teachers of English within these schools.

2.2.3. Qualitative data collection methods and procedures

I collected data through interviews, observations, and document analysis. I interviewed the participants using in-depth, semi-structured interviews guided by an interview schedule (DiCicco-Bloom & Crabtree, 2006). The interview questions were informed by the findings of the quantitative phase, the literature reviewed, and the research questions. Participants were asked about the types of assessment used, the factors that influenced their choices in general and specific factors in their schools that would have influenced them to select any assessment tool or strategy most frequently. The interviews lasted 20–90 minutes, with a mode of 45 minutes. At the end of the day or week of each interview, I transcribed and emailed the transcripts to each participant for their verification. I intertwined data collection and analysis to allow the analysis results to guide subsequent interviews and observations. After I interviewed the participants, I observed each of them three times while they taught three different classes, with class periods lasting 45 minutes (single session) to 90 minutes (double session). However, the first observation for each teacher was not recorded to reduce reactivity (Fraenkel & Wallen, 2003; Johnson & Turner, 2003). In the other two observations, I observed classroom practices without participating in the activities. I tape-recorded each classroom observation and supplemented the recordings with my field notes. I also observed other school functions, such as prize-giving ceremonies, devotions, student activity during recess, and school paraphernalia (notice boards and paintings on the walls) — to understand the context better. I extended my field notes immediately after the observations or at the end of the day when the information was fresh in my mind. No more than three observations were conducted for a day and the observations were transcribed at the end of the day. It took five months to complete all the observations. These observations provided a direct picture of the teachers' assessment practices. It also allowed me to corroborate, refute or extend the assessment practices the teachers reported on the questionnaires and in the interviews (Charmaz, 2006). The teachers' lesson plans were also analysed to increase the accuracy of the findings through triangulation.

2.2.4. Qualitative data analysis and presentation

Marshall and Rossman (2016) purported that typical procedures for analyzing qualitative data involve "immersion in the data, generating categories and themes, coding the data, offering interpretations through analytical memos, searching for alternative understanding and writing the report or other format for presenting the study" (p. 209). These were the methods I employed in this study. I read through the transcripts for each case to get an overall impression of the teachers' assessment practices. Then, beginning with my research questions, I listed possible theory-generated codes and categories (Marshall & Rossman, 2016). For example, for the research question, *(What assessment tools and strategies do Jamaican secondary school*

teachers use most frequently?), the individual assessment tools and strategies: *questioning, tests, oral quizzes, teacher observation, and extended writing (essays, written speeches, and stories)* were listed as theory-generated codes under the category of *traditional assessment tools and strategies*. I also classified the other tools and strategies from the questionnaire used in the quantitative phase as alternative assessment tools and strategies. Since I was using the qualitative phase to corroborate the findings of the quantitative phase, I used all the assessment strategies on the questionnaire as codes. However, I was keen to identify tools and strategies that were not on the questionnaire. Therefore, I coded the data deductively and inductively (Saldaña, 2016). My literature review, the quantitative results, and my initial exploration of the qualitative data generated many of the deductive codes.

Using QDAMiner, I first coded sentences and chunks and employed independent coding (Thomas, 2006) by a lecturer and veteran qualitative researcher to validate my codes and coding. Then, I categorised the codes using pattern coding before seeking answers to the research questions through pattern matching (Yin, 2014). Pattern matching is where the researcher "compare[s] an empirically based pattern — that is, one based on the findings from your case study — with a predicted one made before you collected your data (or with several alternative predictions)" (Yin, 2014, p. 143). I made predictions for each research question. For example, for the first research question (*What assessment tools and strategies do Jamaican secondary school teachers use most frequently?*), I predicted that teachers used predominantly traditional assessment methods, with pen-and-paper tests being the most frequently used assessment method. This prediction was based on the review of the extant literature, the findings of the quantitative phase of this research, and the findings of the qualitative pilot study. I ran a code frequency on the category, "*Assessment tools and strategies used," to match the empirical data from the qualitative data.* This output combined the assessment tools and strategies reported by all the teachers and those I observed them using. I then separated the frequency of use across the different types of data (interview and observations) to ascertain the difference between teachers' reported and observed frequency of use. After that, I classified the tools and strategies in this list as traditional or alternative, based on Gronlund's (2006) specifications, as was done in the quantitative phase. With information on the types of assessment and the frequency of use, I assessed whether and to what extent the empirical data matched my initial prediction. I also used pattern matching to identify possible answers to the other research questions.

After the individual case analyses, I conducted cross-case analyses within the context (type of school) and across cases and contexts. These analyses were done by using the same set of categories and profiles of each case, arranging them in a matrix, and then checking for replications (similarities) and contrasts (differences) across cases (Yin, 2014). In doing the cross-case analyses, I utilised explanation building (Yin, 2014) because I wanted to explain the findings from the quantitative phase, particularly why there was no difference in teachers' reported frequency of use of assessment tools and strategies based on school type. Explanation building allowed me to provide these explanations and explore rival explanations while strengthening the credibility of the findings by showing how "these rival explanations cannot be supported given the actual set of case study findings" (Yin, 2014, p. 150). It also allowed me to "build a general explanation that fits each case, even though the cases will vary in their details" (Yin, 2014, p. 148).

To interpret the data, I looked at patterns in the data (causes and effects, sequence, hierarchy, frequencies) and extrapolated possible explanations for these relationships. I also used the "most useful data segments to support the emerging story, to illuminate the questions being explored" (Marshall & Rossman, 2016, p. 219). I looked for alternative explanations throughout, as supported by the data collected. According to Yin (2014), the findings of

multiple case studies may be reported as an overall cross-case analysis with separate sections devoted to different topics. I used this reporting format in this study. I also interspersed exemplars from the individual cases throughout the different sections.

3. FINDINGS

3.1. Types and Frequency of Use of Assessment Tools and Strategies

Based on the data analysis in the quantitative and qualitative phases, the prediction that teachers predominantly used traditional assessment tools and strategies, especially pen-and-paper tests, was corroborated. In the quantitative phase, the teachers reported using traditional forms of assessment most frequently, with tests (98.9%, n = 1072), questioning (98.4%, n = 1077), teacher observations (95.2%, n = 1063), practical tests (92.8%, n = 1053), and oral quizzes (94.4%, n = 1081) being the most frequently used tools and strategies (see Table 2). Though a higher overall percentage of the sample reported that they used oral quizzes over practical tests, practical tests were ranked higher because more teachers reported that they *always* used them. This level of frequency resulted in a higher mean score for practical tests (M = 3.68, SD 1.23) than for oral quizzes (M = 3.63, SD = 1.14). The percentage of teachers who indicated that they always used tests and questioning, 51.8 (n = 1072) and 51.6 (n = 1077), respectively, further underscored the high frequency of reported use of traditional assessment tools and strategies.

Table 2. *Assessment tools and strategies used by classroom teachers (Quantitative phase).*

Tools & Strategies	n		M	SD	DK	NU	SU	U	FU	AU	%
	Valid	Missing									
Tests	1072	16	4.32	0.856	0.4	0.7	1.9	11.8	33.4	51.8	98.9
Questioning	1077	11	4.27	0.946	0.7	0.7	3.4	12.7	30.7	51.6	98.4
Teacher observations	1063	25	3.85	1.194	1.5	3.3	6.4	25.9	23.4	39.5	95.2
Practical tests	1053	35	3.68	1.234	0.1	6.2	8.5	24.3	28	32	92.8
Oral quizzes	1081	7	3.63	1.137	0.7	4.9	8	29.5	31.1	25.8	94.4
Self-evaluations	1063	25	3.49	1.199	0.9	4.7	13.5	32.3	22.6	26	94.4
Essays	1064	24	3.23	1.404	1.3	14.8	14.1	22.7	23.8	23.4	84
Peer evaluations	1053	35	3.1	1.156	1.5	7.8	18.1	35.8	25.2	11.6	90.7
Roleplays	1071	17	3.1	1.262	0.7	12.4	17.7	31.5	21.4	16.3	86.9
Rubrics	1053	35	3.01	1.421	5.4	10.7	16.8	30.3	17.9	18.8	83.8
Checklist	1060	28	2.8	1.181	2.5	12.3	22.1	36.7	18.7	7.8	85.3
Speech/Debate/Drama	1053	35	2.76	1.245	1.1	18.5	20.4	32.5	17.9	9.5	80.3
Research reports	1043	45	2.74	1.222	2.2	14.9	23.9	34.4	15.2	9.4	82.9
Portfolios	1050	38	2.67	1.168	1.1	16.8	25.5	34.6	15.2	9.4	84.7
Concept maps	1058	30	2.62	1.205	2.7	16.8	26.2	30.1	18.4	5.8	80.5
Flow charts	1052	36	2.5	1.205	2.9	19.8	27.3	30.1	14.1	5.9	77.4
Student journals	1054	34	2.41	1.201	1.6	25.1	26.9	29.6	10.1	6.6	73.2
Interviews	1070	18	2.34	1.245	2.2	29.1	24.3	27.7	9.9	6.8	68.7
Case studies	1057	31	2.21	1.24	3	32.3	26.6	22	10.6	5.5	64.7
Anecdotal records	1031	57	1.93	1.402	15.5	29.7	20.6	20.7	7.6	6	54.9
Conferences	1046	42	1.85	1.207	8.2	40.6	21.9	19.5	6.6	3.2	51.2
Contracts	1040	48	1.53	1.144	13.7	47.5	19.7	13	3.6	2.6	38.9

Note. DK = Don't Know, NU = Never Used, SU = Sometimes Used, U = Used, FU = Frequently Used and AU = Always Used

On the other hand, the five least reportedly used assessment strategies were interviews (68.7%, *n* = 1070), case studies (64.7%, *n* = 1057), anecdotal records (54.9%, *n* = 1031), conferences (51.2%, *n* = 1046) and contracts (38.9%, *n* = 1040). It is also noteworthy that anecdotal records and contracts are the two strategies that were most frequently left unanswered — with 57

(0.05%) and 49 (0.05%) missing responses, respectively. This omission could indicate that more teachers did not know about these strategies but were unwilling to indicate their lack of knowledge.

The qualitative results confirmed that traditional assessment tools and strategies were reported and observed being used more frequently by the participants. In the qualitative phase, all the traditional assessment tools and strategies, except Oral Quiz, were in the top five, with 'Test' (selected response and short answer items only), 'Teacher Observation,' 'Questioning' and 'Extended Writing (essays, written speeches, and stories) ranked 1–4, respectively, as the most frequently used assessment tools and strategies (see Table 3). When I asked the participants which assessment tools or strategies they used most frequently (participants were allowed to select more than one assessment tool or strategy), 20 of the 32 participants responded 'Test'.

However, 'Teacher Observation' was the most frequently observed strategy (88 times by 21 participants), although only three teachers reported using it most frequently. Therefore, although overall traditional assessment tools and strategies were reported and observed to be the more frequently used, the specific traditional assessment tool and strategy used differed. Tests were the most frequently reported (20 counts in 20 cases), and Teacher observation (88 counts in 23 cases) was the most frequently observed. Conversely, the only alternative assessment strategy in the top five was peer assessment, listed at number five among the most frequently observed assessment tools and strategies but reported by none of the participants as the most frequently used strategy.

Table 3. *Comparison of teachers' reported and observed frequency of use of assessment tools and strategies (Qualitative phase).*

Top 10 Assessment Tools and Strategies Reported and Observed	Reported Use		Observed Use		Total	
	Counts	Cases	Counts	Cases	Counts	Cases
Test (MCQS, T/F, Short answer)	20	20	62	32	82	32
Teacher Observation	3	3	88	23	91	23
Questioning	4	4	50	27	54	27
Extended Writing	8	8	20	17	28	19
Peer-assessment	-	-	25	15	25	15
Presentation	4	4	23	13	23	13
Oral Quiz	-	-	7	4	7	4
Research Report	-	-	6	6	6	6
Dramatization	1	1	4	4	5	4
Game/Puzzle	1	1	1	1	1	1

Note. - = none was reported

3.2. Different Subject, Different Assessment Tools and Strategies

The second research question assessed differences in teachers' reported frequency of use of the different types of assessment tools and strategies based on the subject the teachers taught. This was analysed quantitatively using a one-way between-groups ANOVA with a post-hoc test as all the subjects were collapsed into nine categories: English, Mathematics, Sciences, Social Sciences, Business, Performing Arts, Practical Arts, Modern Languages, and Mixed and traditional assessment and alternative assessment as dependent variables in separate analyses.

3.2.1. *Subject differences for traditional assessment tools and strategies*

The assumptions of normality and homogeneity of variance for ANOVA had been violated for the frequency of use of traditional assessment tools and strategies. However, both the Welsh and Brown-Forsythe tests revealed a significant difference between teachers' reported frequency of use of traditional assessment tools and strategies based on subject ($p < 0.001$ for

both tests). The non-parametric, Kruskal-Wallis test which both Field (2013) and Pallant (2013) recommend instead of a one-way between-group ANOVA when the distribution is not normally distributed, also showed a significant difference (see Table 4). Therefore, it was concluded that there was a significant difference in teachers' reported frequency of use of traditional assessment tools and strategies based on subject.

Table 4. *Kruskal-Wallis test for subject*traditional assessment tools & strategies.*

Null Hypothesis	Test	Sig.	Decision
The distribution of traditional assessment tools and strategies is the same across categories of subject.	Independent-Sample Kruskal-Wallis Test	< 0.001	Reject the null hypothesis

Note. Asymptotic significances are displayed. The significance level is .05.

The post hoc test results revealed that the differences between teachers of English and teachers of Mathematics and Practical Arts were significant at a confidence interval of .05. The teachers of English reported using traditional assessment tools and strategies more frequently than teachers of Mathematics and Practical Arts ($M = 4.04$, *SD* .64 for teachers of English and $M = 3.54$ and 3.75, $SD = .56$ and $.69$ for teachers of Mathematics and Practical Arts, respectively). There were significant differences among other subject areas as well. The teachers of Mathematics reportedly used traditional assessment tools and strategies significantly less frequently than the teachers of Social Studies, Science, and Business. There was also a significant difference between the Social Sciences teachers ($M = 4.07$, *SD* .64) and the teachers of the Practical Arts ($M = 3.79$, $SD = .69$). The effect size was moderate at .06. Consequently, I concluded that there were practical differences.

Overall, the teachers of Mathematics reported using traditional assessment tools and strategies with the least frequency ($M = 3.54$, $SD = .56$). On the other hand, the Social Sciences and English teachers reported the highest use of traditional assessment tools and strategies. This result is arguably because 'essay', which is frequently used as an assessment tool by English teachers, was classified as a traditional assessment. This probability was supported by another ANOVA with subjects as the independent variable and essays as the continuous, dependent variable. It revealed that the mean score for English was the highest, ($M = 4.07$, $SD = .9$) followed by Social Sciences ($M = 3.94$, $SD = 1.06$). Additionally, the teachers of English had significant differences in the reported frequency of use from all the other subject areas except Social Sciences and Modern Languages. Predictably, the teachers of Mathematics reported using essays with the least frequency, which was significantly different from all the other subject groups. Since only the teachers of English participated in the qualitative phase of the research, subject differences were not explored in this phase.

3.2.2. *Subject differences for alternative assessment tools and strategies*

The difference in teachers' reported frequency of use of the alternative assessment tools and strategies based on subject was also analysed quantitatively using a one-way between-groups ANOVA with a post-hoc test. The assumption of normality had been violated ($p = .002$ on the K-S test). However, as Elliott and Woodward (2007) and Pallant (2013) stipulated, when the sample size is greater than 30 or 40, parametric tests can be used even if there is a violation of the assumptions of normality. Therefore, I proceeded with the ANOVA since the sample was 1088.

The results of the ANOVA revealed that there was a significant difference ($p < 0.001$). A subsequent examination of the post hoc test results revealed significant differences between teachers of English and teachers of Mathematics and Science at a confidence interval of 0.05. The teachers of English reported using alternative assessment tools and strategies more frequently than teachers of Mathematics and Science ($M = 2.75$, *SD* .56) for teachers of English

and M = 2.26 and 2.49, SD = .63 and .60 for teachers of Mathematics and Science, respectively. Significant differences were also found between teachers of Mathematics and all the other subject groups except Science and Modern Languages. The teachers of Mathematics reportedly used alternative assessment tools and strategies less frequently than all the other subject groups, including Science and Modern Languages. This meant that the teachers of Mathematics reported that they used alternative assessment tools and strategies with the least frequency. A mean score of 2.27 out of 5 meant that, on average, the teachers of Mathematics reported that they sometimes used the alternative assessment tools and strategies on the instrument. Apart from the significant difference between teachers of Social Science and Mathematics discussed earlier, there were also significant differences between the teachers of Social Sciences (M = 2.84, SD .63) and the teachers of the Science (M = 2.49, SD .60), Practical Arts (M = 2.58, SD .57) and Modern Languages (M = 2.48, SD = .58). The teachers of Social Sciences reported using alternative assessment tools and strategies more frequently than these other subject areas as well. The teachers of the sciences and the teachers of subjects categorized as the performing arts differed significantly as well, with the performing arts teachers reporting a higher frequency of use. This significant difference is in addition to the significant differences found between English and Social Sciences teachers reported earlier. Business differed significantly from Mathematics and Performing Arts. While the reported frequency of use by business teachers was higher than that of teachers of Mathematics, it was lower than that reported by the performing arts teachers. The performing arts teachers reported the highest frequency of use of alternative assessment tools and strategies (M = 3.03, SD = .69). However, while this is the highest, it is much lower than the highest mean score for the reported frequency of use of traditional assessment (M = 4.07, SD = .64 for Social Sciences). It is also lower than the lowest mean score for reported frequency of use of traditional assessment tools and strategies (M = 3.54, SD = .56) for Mathematics.

In continuing, the results also showed that performing arts teachers reported using alternative assessment tools and strategies significantly more frequently than practical arts (M = 2.58, SD = .58), modern languages (M = 2.48, SD = .58) and mixed teacher (M = 2.61, SD = .65). This is in addition to all the other subjects discussed earlier (Mathematics, Business, Science and Practical Arts). Finally, the teachers who taught more than one category of subjects (Mixed) differed from the teachers of Mathematics and the performing arts, as was discussed earlier. They reported using alternative assessment tools and strategies more frequently than teachers of Mathematics but less frequently than the performing arts teachers. The effect size was moderate at .09, which indicated that the differences were not by chance.

3.3. Different School Type, Same Assessment Tools and Strategies, Same Assessment Policy

A two-way between-group analysis of variance was used to determine if there were differences in teachers' reported frequency of use of the different types of assessment tools and strategies based on school type. This technique was suitable because, in this study, school type referred to the type of school (traditional, technical, and upgraded) as well as the rank of the school (above average, average, below average). All the assumptions except normality and homogeneity of variance for Frequency of Use of Traditional Assessment Tools and Strategies (FUTATS) were met. However, since Elliott and Woodward (2007) and Pallant (2013) purport that with a larger sample, the assumption of normality is frequently violated, and ANOVA is robust to violations of the assumption of normality and "reasonably robust" to violations of the assumption of homogeneity of variance (Pallant, 2013, p. 204), I continued with the ANOVA.

The results of the ANOVA showed that the interaction effect was not significant (p = .74). There was also no significant difference in FUTATS based on school type or school rank (p = .20 and .27, respectively). There was also no significant difference in FUAATS based on school

type or rank (*p* = .64 for SchoolType*SchoolRank, .72 for School Rank, and .29 for School Type. Therefore, the quantitative analyses revealed no significant difference in teachers' frequency of use of either traditional or alternative assessment methods based on school type.

3.3.1. *Qualitative explanations of the absence of significant difference based on school type*

Based on the quantitative findings, I used the subsequent qualitative phase to explain why there was no difference in the frequency of use of traditional and alternative assessment tools and strategies based on school type. This quantitative finding was surprising given the grave disparities in student academic ability, infrastructural development and support, parental and alumni support, and teacher qualification among the different types of schools: traditional, upgraded, and technical high schools. When I analysed contextual data in the qualitative phase, I observed that traditional high schools benefited from better infrastructural development, alumni, and parental support and had teachers with higher qualifications. They also had more well-behaved students with higher overall academic achievement and achievement in English. For example, each classroom in the top-performing traditional high school was outfitted with projectors and HDMI connections for technology integration, while there were insufficient classrooms and, desks and chairs in the low-performing upgraded and technical high schools. Additionally, while all the teachers in the traditional high schools had a degree in English Language Education and some a master's degree, some of the teachers in the technical and upgraded high schools only had teaching diplomas. Some of the teachers in the below-average upgraded high school were trained to teach at the primary level and not to teach English. (For an extended discussion, see Williams-McBean 2021). Therefore, I wanted to find out why there was no difference in teachers' frequency of use of traditional and alternative methods despite the contextual differences. The data revealed similarities in the schools' assessment policies that led teachers to select traditional assessment tools and strategies more frequently than alternative assessment tools and strategies. These similarities include mandatory, standardised testing and a quota of grades.

3.3.1.1. **Mandatory, Standardised Testing Led to Greater Use of Traditional Assessment Tools and Strategies.** In all the participating schools, the schools' assessment policies propelled teachers into using traditional assessment tools and strategies by stipulating mandatory tests and essays. All the teachers reported that their schools' assessment policy required that teachers administer monthly or six weekly tests in addition to end-of-term and end-of-year examinations, which are usually standardised pen-and-paper tests. Even when not specified, the administrators' negative attitude to other assessment tools and strategies propelled teachers to use written assessments (tests and essays). This negative attitude is typified in the explanation provided by Mrs. Moody, from the below-average traditional high school, as to why she used written tests most frequently. She explained,

> I used to like doing a lot of drama first time …. But it's difficult now because the push is about the homework, the classwork, the test. It is more now of an academic institution right throughout, instead of making the students whole. I think the culture of the school is dying and where we can be creative that is basically taken away because when a drama presentation with students was suggested as the graded test for grade nine, it was shunned by the Head of Department and administrators. (Interview with Mrs. Moody)

Another example was seen when Mrs. Black from the above-average traditional high school shared that her school's assessment policy stipulated that teacher's term assessment classwork or homework "must include at least one essay and one comprehension task." The school administrators stated or expressed a preference for written tasks influenced teachers to select traditional assessment tools and strategies more frequently. Since this preference was evident in all the schools in the study, it partially explained why there was no statistically significant

difference in teachers' frequency of use of traditional or alternative assessment tools and strategies based on school type.

3.3.1.2. Higher Quota of Grades Led to Greater Use of Traditional Assessment Tools and Strategies. Schools that require a quota of grades from teachers also influenced teachers to use predominantly traditional assessment tools and strategies, despite school type. For accountability purposes, in each school, each teacher was required to input a set number of grades into the school's grading system per month, six-week period, or term (see Table 5).

Table 5. *The quota of grades required for each school.*

School Name (pseudonyms)	School Type & Rank	Number of Grades Required per Term		Type of Grades
		Language	Literature	
Sunnydale High School	Traditional Above Average	14		Three tests, two classwork, and two homework per subject
		7	7	
James Stewart High School	Traditional Average	12		One homework or classwork and one test per subject every six weeks.
		6	6	
Harrison High School	Traditional Below Average	12		One homework or classwork and one test per subject every six weeks.
		6	6	
Roaring River High School	Upgraded Above Average	36		Two homework, two classwork, one test and one affective every six weeks.
		18	18	
Willow High School	Upgraded Below Average	6		One homework, one classwork, one test.
		3	3	
Hill Top High School	Technical Above Average	24		One homework, one classwork, one test per month
		12	12	
Northside High School	Technical Below Average	4		Midterm and end of term exams.
		2	2	

These grades usually come from classwork, homework, and tests and were sent to parents on report cards. While the number of grades varied in each school, ranging from four to 36 per term for three terms (Christmas, Easter, and Summer terms), the impact of the quota requirement was similar in most of the schools. The more grades required, the higher the likelihood of teachers assessing students using traditional assessment tools and strategies.

When I asked the participants how the school's assessment policy impacted their choice of assessment, most of them explained that the required number of grades led them to use traditional assessment tools and strategies. These traditional assessments were primarily selected-response items with one correct answer because they were easier to mark. In that way, they could meet their grade quota more easily. This impact was most evident in Roaring River High School, which had the highest required number of grades per term (36). The explanation was typified in the response from Mrs. Turner. She explained:

> It has a lot of influence on it [her classroom assessment practices] because I teach so many classes, and I have so many grades to give in for the month. What I do is I plan some assessments that are not time-consuming to mark, especially for literature. So, what happens is that it is not as meaningful as I would like it to be. Because when I would give them like an essay or something, or have them do some extended writing, with the number of grades … If I have to give in five pieces of grades for literature, I have to give the students some questions based on the chapter and

> I give them like one to ten and so on, and they use a couple minutes and answer those questions, short answer questions. Or I give them something that is multiple choice if I have like a paper that I set before – a past paper that is multiple choice. I give them like from one to a certain number and have them answer the questions and in quick time I finish marking it and I give them a grade.

This teacher gave the students easy-to-mark assignments just to get a grade, and because traditional assessments are easier to mark and less time-consuming, they would be used more frequently.

Another teacher, Ms. Hall, also from Roaring River, explained how she changed from using activities that focused on the students' ability to speak in English to written pieces that were easy to mark to meet the quota of grades. I watched Miss Hall give her students a test comprised of 10 short answer questions on a chapter from the novel the class was reading in Literature class. I asked her why she decided to use a written test. She responded:

> Let me just say something. What I'm accustomed to in the classroom … my focus was mainly on learning. Well, that's what I believe, teaching and learning are the focus, right? So, before I came here, we spent more time teaching the concepts and evaluating the students on actually understanding the concepts. And evaluation didn't mean like four pieces or five pieces for the month. It would probably be like three pieces over a six-week period or something like that, so it was not that frequent. So, the whole speaking aspect of it came into play because then I had them speak more. They had the chance to take part more and not be afraid that I was going to mark every piece of work they did. That's what I'm used to. That's the kind of environment that I'm used to. So, this (*She points to the test paper*.) is a shock to me, and so I'm gradually getting accustomed to it. That is all I can tell you.

From the excerpt, it is evident that the other school in which Miss Hall taught (that was not included in this qualitative phase of this study) also required a quota of grades from the teachers. It is also evident that using traditional assessment tools and strategies becomes more likely when the quota is higher. The higher the quota of grades required, the higher the marking load and the less time the teachers have to focus on alternative assessments that take more time to administer and score.

While the impact of the quota of grades was most evident in Roaring River High School, it was evident in the other schools. In most schools, the teachers found the school's assessment policy "challenging" because of the amount of marking required or because of the frequency of the assessment coupled with the large class sizes. As Mrs. Peart from Sunnydale High School explained:

> Sometimes it is challenging to ensure that you have the number of pieces because you must have two classwork pieces as well, and I think two homework pieces for both language and literature, so it takes a lot. It's a lot of marking. (Interview with Mrs. Peart)

Ms. Khan from Hill Top High School gave a similar explanation: "I think some [classes] probably have like forty-six or so. I think the lowest number is forty-five. Yeah, so you can just imagine having all those books to mark, and all those assignments".

Ms. Hunter from Harrison High School also explained:

> So alright, the term starts in September. Six weeks take us to mid-October, and I teach, and I test at that time. It's going to take me to — and I have to mark all of those pieces. While marking, I must still be teaching, and still, I have to be setting another set of six-week work again. The testing time is too much! (Interview with Ms. Hunter)

In essence, the stipulated grades caused the teachers to view the policies as challenging because to ensure they met their grade quota, the teacher had to be marking students' work much more frequently while teaching and engaging in other school activities. The challenge was also associated with large class sizes, as seen in the excerpt taken from Ms. Khan's interview. To overcome this challenge, many of the teachers used selected-response items.

The challenge of the grade quota system was evident in all the schools except Willow High School (the below-average upgraded high school), where there was no formal assessment policy and three grades were required per subject per term. At Willow High School, most teachers of English were required to submit three grades per term for English Language because English Literature was only taught to the top-streamed class in each grade. In this school, the teachers' preference for written assessment was primarily influenced by the format of the internal exams (End of Term and End of Year) and national assessments (CSEC, City & Guilds). The teachers taught to prepare students for these assessments. Therefore, they tested using similar formats (primarily tests and essays) but introduced projects after a project-based school-based assessment was added to the CSEC English examinations. This explanation is exemplified in the excerpt taken from the interview with Mrs. Downs. She explained:

> Sometimes you give them homework and projects. We try to give them at least one project per term so that they can get used to it, especially for the SBAs. Because we're having a problem with them at grades ten and eleven when they are to do the SBAs, we are trying from grade seven to say, okay you must do projects, and we're going to teach you skills for doing projects, so we trying to do that.

In sum, the assessment policies in the schools that participated in the qualitative phase of the research were largely similar in requiring or expressing a preference for traditional assessment tools and strategies and specifying a quota of grades that the teachers had to supply per month, six weeks, or term. These policy requirements influenced teachers of English to use traditional assessment tools and strategies more frequently because they were more manageable and less time-consuming to administer and score. The higher the grade quota, the more likely teachers would use selected-response and short-answer questions to assess students. The focus of assessment became to provide grades rather than to assess students' learning meaningfully. Tests (consisting of selected-response items only) were easier to mark, save teachers time, and ensured they met their grade quota. This largely accounted for the absence of differences across school types. The absence of difference in teachers' frequency of use of traditional assessment tools and strategies was also due to the format of internal and external summative examinations. Since those were primarily traditional, the teachers used traditional assessment formats as well. However, efforts were made to introduce projects since it was introduced as a part of the secondary exit English examinations offered by the Caribbean Examinations Council.

4. DISCUSSION of THE FINDINGS, IMPLICATIONS, and RECOMMENDATIONS

The findings of both the quantitative and qualitative phases of this mixed methods study confirmed the findings of previous studies that secondary school teachers primarily used traditional assessment tools and strategies. Among the traditional assessment methods, pen-and-paper tests which primarily included selected-response items, were most frequently used. Previous international researchers also reported the dominance of testing (see, for example, Acar-Erdol & Yıldızlı, 2018; Berry, 2010; Brookhart, 2013; Dandis, 2013; Esomonu1 & Eleje, 2020; Guskey & Link, 2019; OECD, 2019; Saefurrohman, 2017; Vlachou, 2018). The same was reported in the lone local study conducted by Onyefulu (2018). This dominance has persisted despite pedagogical shifts, curricular rewrites, and increased advocacy for the greater use of alternative assessment tools and strategies. Since classroom assessment is primarily supposed to be used to improve teaching and learning (Acar-Erdol & Yıldızlı, 2018) and that improvement can be increased by using alternative assessment tools and strategies (Berry 2010; Black & Wiliam, 1998; Koh, 2017; McMillan, 2014), there is need for research on why teachers continue to use traditional assessment tools and strategies with far greater frequency.

The explanations provided by the teachers of English who participated in the qualitative phase of this research provided some useful insights. The teachers primarily used tests to assess their students despite variation in students' academic ability, infrastructure which allowed for

innovations in assessment, teacher qualification, and parental support because the school's assessment policies required or expressed more positive attitudes towards traditional tests. They also used selected-response and short answer tests to meet the school administration's quota of grades per month, six weeks, or term. The higher the grade quota, the more frequently these tests were used, even if the administrators allowed teachers to choose the assessment format. Finally, the teachers used traditional tools and strategies more frequently because they modelled internal and external, standardised, summative examinations, which primarily used written examinations. However, as the format of these examinations changed, for example, to include school-based assessments, teachers included alternative assessments (i.e., projects). Other studies have also reported that the format of external, standardised assessment has influenced teachers to select and create and use traditional tests in the classroom (Berry, 2010; McMillan, 2003; Ong, n.d.).

Consequently, changes in assessment must be accompanied by policy changes at the school level to allow teachers time to administer and score alternative assessment tools and strategies. School administrators must also demonstrate more positive attitudes towards alternative assessment tools and strategies in practice and reduce the required number of grades. The focus on grades should be replaced with a focus on learning, from the summative use of assessment to the formative use of assessment. However, in the absence of supportive school-level attitudes and practices, researchers must focus on how traditional assessment tools and strategies can be created and used to improve learning (i.e., formatively) and not just for grading (i.e., summatively). Empirical studies on best practices related to the formative use of traditional and summative tests and their impact on students' learning are also needed to improve educational outcomes. Additional research should also be done to find out if the explanations provided by the teachers of English hold true for teachers of other subjects.

This study also showed that teachers' frequency of use of traditional and alternative assessment tools and strategies differed significantly based on subject: English, Mathematics, Sciences, Social Sciences, Business, Practical Arts, Performing Arts, Modern Languages, and Mixed. There were many differences among the groups that were discussed in this paper. Most notable were that teachers of Mathematics reported using both traditional and alternative assessment tools less frequently than teachers of all other subject groups, and the teachers of Social Sciences and English reported the highest use of traditional assessment tools and strategies. The result for the teachers of Social Sciences and English is arguably because 'essays' which are frequently used as an assessment tool by English teachers, were classified as traditional assessments. Some writers classify essays as traditional assessments (Dikli, 2003; Gronlund, 2006, Koh, 2017) while others do not (Frey & Schmitt, 2010; Wren & Gareis, 2019). In this study, essays were classified as traditional assessment primarily because it is popular on the external, standardised examinations offered by the CXC for secondary schools in the Caribbean. In this study, teachers of English and the Social Sciences reported that they used essays with a significantly higher frequency than all other subjects, which largely accounted for their significantly greater use of traditional assessment tools and strategies. In contrast, the teachers of the Performing Arts used alternative assessment tools and strategies with the highest frequency, and the teachers of English also reported using alternative assessment tools and strategies more frequently than teachers of Mathematics and Science.

There have been conflicting reports from previous studies on differences based on subject, with some researchers reporting significant differences (Alkharusi, 2011; Berry, 2010; Bol et al., 1998; Dandis, 2013; Duncan & Noonan, 2007; Zhang & Burry-Stock, 2003) and others reporting finding no significant difference (Duncan & Noonan, 2007; Ong, n.d.). There are contradictions among those who previously reported significant differences as well. Some researchers reported that teachers of Mathematics indicated that they used alternative

assessment methods with greater frequency than all other subject areas (Bol et al., 1998) or more than teachers of language arts, science and social studies (Zhang & Burry-Stock, 2003), while others reported that teachers of Mathematics use predominantly traditional assessment tools and strategies (Dandis 2013; Watt, 2005). The findings of the latter group of researchers were confirmed in this study. The findings of previous studies that reported that teachers of English and Social Studies used paper-pencil tests and constructed-response items including essays (Berry, 2010, Brookhart, 2009; Zhang & Burry-Stock, 2003) was also confirmed in this study.

The explanations provided by the teachers of English in the qualitative phase of this study, which were previously discussed, provide some insights as to why teachers of English used tests so frequently. However, since this qualitative exploration was not done with the teachers of other subjects in this study, future studies could provide said qualitative explanations. Even the qualitative explanations provided by the teachers of English in this study should be explored in other contexts as what obtains in one region, country, school, or classroom may differ from another. Diverse contextual issues not identified in these schools and Jamaica may become evident in future studies. It is through identifying and responding to these issues can we hope to improve teachers' assessment practices and improve teaching and learning through the formative use of assessment.

Acknowledgments

Declaration of Conflicting Interests and Ethics

The author declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author.

Orcid

Clavia T. Wlliams-McBean ⬤ https://orcid.org/0000-0003-3434-8913

REFERENCES

Acar-Erdol, T., & Yıldızlı, H. (2018). Classroom Assessment Practices of Teachers in Türkiye. *International Journal of Instruction*, *11*(3), 587-602. https://doi.org/10.12973/iji.2018.11340a

Adeyemi, B. (2015). The efficacy of authentic assessment and portfolio assessment in the learning of social studies in junior secondary schools in Osun state, Nigeria. *IFE Psychologia: An International Journal, 23*(2), 125-132.

Alkharusi, H. (2011). Teachers' classroom assessment skills: Influence of gender, subject area, grade level, teaching experience and in-service assessment training. *Journal of Turkish Science Education*, *8*(2), 39-48).

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Berry, R. (2008). *Assessment for learning*. Hong Kong University Press.

Berry, R. (2010). Teachers' orientations towards selecting assessment strategies. *New Horizons in Education*, *58*(1), 96-107.

Bland, L.M., & Gareis, C.R. (2018). Performance assessments: A review of definitions, quality characteristics, and outcomes associated with their use in K-12 schools. *Teacher Educators' Journal*, *11*, 52-69.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Bol, L., Stephenson, P.L., O'Connell, A.A., & Nunnery, J.A. (1998). Influence of experience, grade level, and subject area on teachers' assessment practices. *The Journal of Educational Research*, *91*, 323–330. https://doi.org/10.1080/00220679809597562

Bramwell-Lalor, S. (2019) Assessment for learning on sustainable development. In: Leal Filho W. (eds) *Encyclopedia of sustainability in higher education*. Springer, Cham. https://doi.org/10.1007/978-3-319-63951-2_1-1

Brookhart, S.M. (2009). Assessment and examinations. In L.J. Sasha & A.G. Dworkin (Eds.), *International handbook of research on teachers and teaching*. Springer Science & Business Media.

Brookhart, S.M. (2013). Comprehensive assessment systems in service of learning: Getting the balance right. In R. W. Lissitz (Ed.), *Informing the practice of teaching using formative and interim assessment: A systems approach* (pp. 165–184). Information Age Publishing.

Buhagiar, M. (2007). Classroom assessment within the alternative assessment paradigm: Revisiting the territory. *Curriculum Journal*, *18*(1), 39-56. https://doi.org/10.1080/09585170701292174

Burke, K. (2009). *How to assess authentic learning*. Corwin Press.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative data analysis*. Sage.

Cobern, W.W., & Adams, B.A. (2020). Establishing survey validity: A practical guide. *International Journal of Assessment Tools in Education*, 7(3), 404-419. https://doi.org/doi.org/10.21449/ijate.781366

Clarke, M.G. (2011). Rescue upgraded high schools – Gov't must address inequities in education sector. Retrieved from http://jamaicagleaner.com/gleaner/20110821/cleisure/cleisure2.html

Creswell, J. (2014). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4th ed.). Pearson Education Limited.

Dandis, M.A. (2013). The assessment methods that are used in a secondary mathematics class. *Journal for Educators, Teachers and Trainers*, *4*(2), 133–143.

DiCicco-Bloom, B., & Crabtree, B.F. (2006). The qualitative research interview. *Medical Education, 40(*4), 314–321. https://doi.org/10.1111/j.1365-2929.2006.02418.x

Dikli, S. (2003). Assessment at a distance: Traditional vs. alternative assessments. *The Turkish Online Journal of Educational Technology, 2*(3), 13–19.

Dogan, M. (2011). Student teachers' views about assessment and evaluation methods in mathematics. *Educational Research and Reviews*, *6*(5), 417–431.

Duncan, C.R., & Noonan, B. (2007). Factors Affecting Teachers' Grading and Assessment Practices. *The Alberta Journal of Educational Research*, *53*(1), 1–21.

Esomonu, N.P., & Eleje, L.I. (2020). Effect of diagnostic testing on students' achievement in secondary school quantitative economics. *World Journal of Education, 10*(3), 178-187. https://doi.org/10.5430/wje.v10n3p178

Fraenkel, J.R., & Wallen, N.E. (2003). *How to design and evaluate research in education* (5th ed.). McGraw-Hill.

Gronlund, N.E. (2006). *Assessment of student achievement* (8th ed.). Pearson.

Guha, R., Wagner, T., Darling-Hammond, L., Taylor, T., & Curtis, D. (2018). *The promise of performance assessments: Innovations in high school learning and college admission*. Learning Policy Institute.

Guskey, T.R., & Link, L.J. (2019). Exploring the factors teachers consider in determining students' grades. *Assessment in Education: Principles, Policy & Practice*, *26*(3), 303-320.

Hess, K., Colby, R., & Joseph, D. (2020). *Deeper competency-based learning: Making equitable, student-centered, sustainable shifts*. Corwin.

Jiang, Y. (2020). Teacher classroom questioning practice and assessment literacy: Case studies of four English Language teachers in Chinese universities. *Frontiers in Education*, *5*(23), 1-17. https://doi.org/10.3389/feduc.2020.00023

Johnson, B., & Turner, L.A. (2003). Data collection strategies in mixed methods research. In A. Tashakkori & C. Teddie (Eds.), *Handbook of mixed methods in social and behavioural research* (pp. 297–319). Sage.

Koh, K. (2017). Authentic assessment. *Oxford Research Encyclopedia of Education*. Retrieved from https://oxfordre.com/education/view/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-22.

Kubiszyn, T., & Borich, G. (2013). *Educational testing and measurement: Classroom application and practice* (10th ed.). John Wiley & Sons, Inc.

Marshall, C., & Rossman, G.B. (2016). *Designing qualitative research* (5th ed.). Sage.

McMillan, J.H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, *20*(1), 20-32. https://doi.org/10.1111/j.1745-3992.2001.tb00055.x

McMillan, J.H. (2003). Understanding and improving teachers' classroom assessment decision-making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, *22*(4), 34–43. https://doi.org/10.1111/j.1745-3992.2003.tb00142.x

McMillan, J.H. (2014). *Classroom assessment: Principles and practice for effective standards-based instruction* (6th ed.). Pearson.

Miller, D., Linn, R., & Gronlund, N. (2013). *Measurement and assessment in teaching* (11th ed.). Pearson Education: Upper

Monteiro, V., Mata, L., & Santos, N. (2021) Assessment conceptions and practices: Perspectives of primary school teachers and students. *Frontiers in Education*, *6*, 631185. https://doi.org/10.3389/feduc.2021.631185

National Education Inspectorate [NEI]. (2013, November). Chief inspector's baseline report. http://www.nei.org.jm/Portals/0/Content/Documents/Chief%20Inspector's%20Report%20November%202013.pdf?ver=2015-04-08-111059-667

National Education Inspectorate [NEI]. (2014, June). Chief inspector's baseline report. http://www.nei.org.jm/Portals/0/Content/Documents/Chief%20Inspector's%20Report-%20June%202014%20Final.pdf

National Education Inspectorate [NEI]. (2015, September). Chief inspector's baseline report. http://www.nei.org.jm/Portals/0/Chief%20Inspector's%20Baseline%20Report%202015.pdf?ver=2015-09-30-125548-787

National Education Inspectorate [NEI] (2016). Chief inspector's report. http://www.nei.org.jm/Portals/0/Content/Documents/C2R1%20Chief%20Inspector's%20Report%202016%20Final.pdf?ver=2018-04-19-115528-887

National Education Inspectorate [NEI] (2017). Chief inspector's report. https://www.nei.org.jm/Portals/0/Content/Documents/Chief%20Inspector's%20Report%202017.pdf?ver=2018-11-30-102446-537&ver=2018-11-30-102446-537

Oluwatayo, J.A. (2012). Validity and reliability issues in educational research. *Journal of educational and social research*, *2*(2), 391–400.

Ong, S.L. (n.d.). *Profiling Classroom Teachers Assessment Practice*. Retrieved from www.iaea.info/documents/paper_4d32f2cd.pdf

Onyefulu, C. (2018). Assessment practices of teachers in selected primary and secondary schools in Jamaica. *Open Access Library Journal*, *5*(12), 1-25. https://doi.org/10.4236/oalib.1105038

Organisation for Economic Cooperation and Development (OECD) (2019). *OECD reviews of evaluation and assessment in education: Student assessment in Turkey*. Retrieved from https://www.oecdilibrary.org/sites/1807effcen/index.html?itemId=/content/component/1807effc-en

Popham, J.W. (2005). *Classroom assessment: What teachers need to know*. Pearson Education.

Popham, J.W. (2018). *Assessment literacy for educators in a hurry*. Alexandria. ASCD.

Saefurrohman. (2017). Indonesian EFL teachers' classroom assessment methods in reading. *Advances in Social Science, Education and Humanities Research (ASSEHR), 109.* 4th Asia Pacific Education Conference. https://doi.org/10.2991/aecon-17.2017.40

Saefurrohman, & Balinas, E. (2016). English Teachers Classroom Assessment Practices. *International Journal of Evaluation and Research in Education, 5*(1), 82 – 92.

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Sage.

Sewagegn, A.A. (2019). A study on the assessment methods and experiences of teachers at an Ethiopian university. *International Journal of Instruction*, *12*(2), 605-622. https://doi.org/10.29333/iji.2019.12238a

Statistical Institute of Jamaica (2017). Education Statistics. Retrieved from https://statinja.gov.jm/Demo_SocialStats/Education.aspx

Stiggins, R.J, & Conklin, N.F. (1992*). In teachers' hands: Investigating the practices of classroom assessment*. State University of New York Press.

Taber, K.S. (2018). The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Thomas, D.R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, *27*(2), 237-246. https://doi.org/10.1177/1098214005283748

Vlachou, M. (2018). Classroom assessment practices in middle school science lessons: A study among Greek science teachers. *Cogent Education*, *5*(1), Article: 1455633. https://doi.org/10.1080/2331186X.2018.1455633

Williams-McBean, C. (2021). Contextual considerations: Revision of the Wiliam and Thompson (2007) formative assessment framework in the Jamaican context. *The Qualitative Report*, *26*(9), 2943-2969. https://doi.org/10.46743/2160-3715/2021.4800

Wren, D., & Gareis, C.R. (2019). *Assessing deeper learning: Developing, implementing, and scoring performance tasks.* Rowman & Littlefield.

Yin, R.K. (2014). *Case study research design and methods* (5th ed.). Sage.

Zhang, Z., & Burry-Stock, J.A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, *16*(4), 323–342.

Zohrabi, M. (2013). Mixed method research: Instruments, validity, reliability and reporting the findings. *Theory and Practice in Language Studies, 3*(2), 254-263. http://dx.doi.org/10.4304/tpls.3.2.254-262