Caribbean Curriculum
Vol. 19, 2012, 109–143.

SECONDARY SCHOOL ENTRANCE EXAMINATIONS IN THE CARIBBEAN: Legacy, Policy, and Evidence Within an Era of Seamless Education

Jerome De Lisle

Secondary school entrance examinations remain an important feature of education systems within the Anglophone Caribbean. This is at a time when many high-performing school systems have either diversified traditional test-based placement mechanisms or completely postponed early selection and placement. In contrast, high-stakes secondary school selection/placement examinations have persisted in Caribbean nation states, albeit under the guise of reform. Paradoxically, in the postcolonial era of seamless education, some form of test-based selection and/or placement continues, with newly added roles, refined purposes, and exotic new names. These high-stakes systems compete strongly with formative classroom assessment and large-scale learning assessments used for monitoring student achievement. This paper argues that the persistence of test-based early selection in the Caribbean points to a widespread and implicit belief in the infallibility of test scores. This naïve perception among different publics has remained, even in the face of evidence from early sociological studies demonstrating inequalities on the examination. In the era of seamless education reform, there is need for an explicit measurement focus to better judge fairness, validity, and equity. Unless a fledgling Caribbean measurement community can head in the direction of collecting credible evidence, abuses and test score misuse will continue.

Early Test-Based Selection in Trinidad and Tobago: Towards an Analytical Framework

Caribbean societies have traditionally placed a high value on education in general and on secondary school education in particular (Ellis, Ramsay, & Small, 2000). In Trinidad and Tobago, the value attached to secondary schooling was evident even prior to independence. In the colonial period, secondary education was considered to be very different from primary education and technical or craft training because it was

designed to provide "an academic education for the intellectual elite of the Colony" and was considered to be "an avenue for the few to reach the Universities and enter the learned professions" (Trinidad and Tobago [T&T]. Government, 1947, p. 18). Although places were restricted to a privileged few prior to the 1960s, secondary schooling remained in great demand across the wider population (Campbell, 1996, 1997). The limited numbers and high demand meant that from very early on, some method of selection was required. In the post-independence period, the demand for secondary education continued to rise, fuelled perhaps by the arguments of then premier, Eric Williams; the desire to emulate other successful locals; and the belief that schooling was the most productive route to social mobility (Alleyne, 1996; Brereton, 2007).

Secondary school entrance examinations are test-based selection and placement systems positioned at the transition between primary and secondary schooling. As part of a colonial legacy, these systems function as gatekeepers by determining entry and placement into what was once considered a cherished pathway. In most of the Anglophone Caribbean, secondary school examinations are still administered. Selection occurs at age 11 and above in most countries, except Belize, where the Primary School Examination (PSE) is administered at age 12+. Thus, secondary school entrance examinations result in a high-stakes selection process early in the schooling career of students. In the period just after independence, the primary purpose of secondary school entrance examinations was selection; but with universal secondary education instituted in many Caribbean islands during the first decade of the 21st century, that role has shifted towards placement into different types of secondary schools. This paper argues that the test-based selection system in Trinidad and Tobago, which developed during the colonial era, evolved unique characteristics aligned to function and purpose. It will show that many of these features have been repeated as a legacy during the post-independence era. The most important of these are (1) high societal legitimacy and valorisation, (2) overall administrative stability and persistence, and (3) experimentation and tinkering without evidence.

The analytical framework makes use of an explicit postcolonial probe to interrogate and deconstruct past policy and practice. London (2003) has noted the value of using such a frame when analysing educational issues in the Anglophone Caribbean:

> The postcolonial probe therefore provides some new ammunition and perspectives which might be used to investigate some old themes and, in so doing, helps to revise understanding of the "colonial." The approach has been credited with capability

> to impart new prominence to matters relating to colony and to independent statehood alike and in recent times has had a major impact on established modes of cultural analysis. The reason is that the post-colonial theoretic is an approach which has potential to bring to the forefront interconnections among issues relating to race, nation, empire, and cultural production, of which education and schooling are components. (p. 291)

Thus, postcolonial theory is a mechanism through which we might examine "development" or "change" over time, while at the same time gauging the persistent impact of colonial administrative structures and mechanisms (Shahjahan, 2011). Used in an abstract sense, the experiences of colonialism become an important referent for understanding the direction of, and processes associated with, assessment reform in the Caribbean (Tikly, 1999). The key analytical questions in such a historical analysis then become: (1) Are changes associated with secondary school entrance examinations really new or do they represent more of the same, a mirror from the colonial context? (2) Why have secondary school entrance examinations been retained in the post-independence era, with modern and expanded school systems?

The analytical framework also makes use of an explicit measurement perspective. There is no contradiction in using both postcolonial and measurement theory within a single analytical framework. Indeed, this paper argues that modern measurement theory has had very little impact upon the development of assessment systems in the Caribbean. However, measurement theory can highlight critical concerns about the efficiency and impact of an early selection/placement system, and therefore is able to extend the socio-historical postcolonial perspective by focusing upon the quality and utility of the selection instrument. Measurement theory challenges the implicit assumption that tests are simply neutral tools, and provides a framework for gathering credible evidence about psychometric functioning (Mislevy, 1996). Far from being blind to the issue of test use in society, measurement theorists consider fairness as central to their work, and this issue has been explicitly considered in the 1999 *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). Notably, then, Messick (1998) argued that "the concept of fairness is impossible to divorce from the concept of validity because the two share mutuality of meaning and import" (p. 12).

In this paper, the historical account is divided into two main periods separated by the watershed event of independence. The pre-

independence era includes the modern colonial era (pre-1950) and the era of growing nationalism and greater democracy starting in 1950. For convenience, the post-independence period has also been organized into two periods: 1960-1987 and 1988 to present. The year 1987 is a critical milestone because notable changes were made to the Common Entrance examination in 1987 (London, 1997; MacKenzie, 1989). Another possible milestone is the year 2001 when universal secondary education was instituted. Although Campbell (1996, 1997) has documented the history of education in Trinidad and Tobago, dealing extensively with the different examinations, he did not explicitly plot the course of assessment reform. As such, this paper focuses specifically upon the evolution of form and function in secondary school entrance examinations. The historical analysis is based on both secondary and primary documents. The latter include selected annual reports from the colonial era, along with policy papers, academic papers, and task force reports from the post-independence era. Attention is focused upon those documents that contain empirical data or which refer to significant policy decisions for the examination.

The Colonial Period: The College Exhibition

In colonial Trinidad and Tobago, the forerunner to the current system of secondary school selection was the College Exhibition Examination. This examination was established in 1879 for boys and girls under the age of 12. However, the first girl to actually present for the examination was in 1926 (T&T. Working Party on Education in Trinidad and Tobago [Working Party], 1954). Campbell (1983) wrote at length on the nature of the College Exhibition system, which stood shoulder to shoulder with fee payment for secondary school placement. Secondary school fees were established at 16 dollars per term, a goodly sum for the working class at that time. These fees meant that the College Exhibition system was the only viable mechanism for social promotion for the majority of locals. The increasing demand for places in secondary schools led to policymakers granting more and more College Exhibition awards. For example, in 1941, only 16 exhibitions were awarded, but by 1947 that number had increased to 55, and by 1950 to 100 annually (T&T. Government, 1947, 1950, 1951; Working Party, 1954). From this perspective, it is understandable that the disenfranchised local population would come to see a selective examination as valuable in its own right and advantageous compared with other alternatives.

Although the numbers of test takers earning places in secondary school were kept low for a time, one of the advantages of the College

Exhibition system was that the great majority perceived test-based selection to be a relatively transparent and "fair" mechanism (Olmedilla, 1992). In time, it was widely accepted as the primary gateway through which some African-Trinidadians and a few Indian-Trinidadians could earn a secondary school place and, most importantly, quite possibly an island scholarship. O'Callaghan (2009) observed that the College Exhibition was perhaps the only viable mechanism at the time for social mobility among locals such as Arthur McShine, who later qualified to become a doctor. She wrote:

> For Arthur Hutton McShine, more than a generation before, winning one of the few exhibitions was crucial to his future. He would get one of those exhibitions in spite of his mother's first begrudging the extra cost of electricity that his studies entailed. He would, like some other students of his time, cotton on to the very useful street light. He won the exhibition, went to QRC and eventually won that Island Scholarship to study at a University abroad, which Dr Patrick Solomon in his autobiography related as being all that stood between the failure which was the lower rungs of the Civil Service and the success of becoming a doctor or lawyer.

It is in this way that the examination-based selection system might have become legitimized, leading ultimately to intense valorisation by the public. This valorisation of successful test takers, their families, and institutions goes back to the very birth of high-stakes public examinations for the Chinese public service (Miyazaki, 1976). Likewise, in colonial Trinidad and Tobago there was no shortage of support for the College Exhibition, and some of the adherents were surprisingly successful locals who became the examination's most notable and trusted advocates. For example, Campbell (1996) reported H.O.B Wooding as bemoaning the actions of administrators trying to limit the College Exhibition as the proven pathway to success. His cry was in response to a suggestion that places were to be opened up in the intermediate schools for other high performers in the College Exhibition. Such strong advocacy for a selection examination was reasonable considering what success meant to the test taker. Inevitably, it was these successful test takers, a chosen few from the local populace, who would be inducted into the colonial administrative structure. The testimony to the success of the system as a vehicle of social mobility was perhaps best embodied in the first premier himself, who had successfully "earned" the ultimate reward of a university degree in the motherland. He returned to open the door into what was believed to be a new era—one of greater opportunity

and fairness. Such solidly presented myths associated with an examination are not easily erased from the minds of observers.

Still, the records of the time showed that there were also several unintended and negative consequences associated with the College Exhibition system. Substantial evidence came from the Working Party (1954) report, which noted the distortions in the organization of the primary school and the promotion of intense examination coaching:

> When the examination was introduced, the papers were set in English, Arithmetic, Geography and Spelling; and there was a test of Reading. It seems that the purpose then was to secure a few brilliant children who would hold the honour of being 'scholars'. The result was that special classes were started in the schools into which were put only the children who appeared to have the best brains, to be given individual attention and to be kept apart from the rest of the school. Coaching then became intensive, was extended to out-of-school time, and has grown to such an extent that we must condemn it in the strongest terms. (p. 74)

Thus, from very early on, the competitive nature of the College Exhibition created a strong washback effect on both teaching and school organization. As defined in the modern literature, washback is the positive or negative effect of a high-stakes test on pedagogy, organization, and learning in schooling. The concept of washback is especially prominent in modern-day language testing (Alderson & Wall, 1993; Cheng, 1997; Cheng, Watanabe, & Curtis, 1994). The Working Party (1954) explored several dimensions of negative washback, including the reorientation of the purpose and nature of teaching in the primary school. Thus, the committee observed that "the rest of the school is often neglected in order to provide for the needs of the 'exhibition' class" (p. 70), with the syllabus truncated in line with examination demands.

Another significant adverse impact was the intense examination orientation promoted among test takers, teachers, and parents. Again, the Working Party noted that the College Examination tended to spawn additional internal examinations prepared by head teachers. These were designed to "weed out" unprepared students. Such actions were possibly prompted by the numerous reports on the unpreparedness of candidates. For example, in the 1928 Administration Report, the Director of Education had noted that:

> In my previous reports I have called attention to the undesirability of making children sit for an examination for which they are not prepared, and which they have not the smallest chance of passing, but in spite of these remarks, matters were even worse this year. (T&T. Government, 1928, p. 6)

This finding also alluded to the overall poor performance of candidates, which was accompanied by large urban-rural disparities in performance. Thus, both quality and equity were notable issues at that time.

Nevertheless, despite these adverse and unintended consequences (Broadfoot, 2002), there was also much public and administrative resistance to stopping the examination altogether, leading eventually to its retention over the years (Payne & Barker, 1986). Arguably, the system of test-based selection may have survived almost intact because of the high value and perceived legitimacy of the system. Paradoxically, however, this apparent stability did not prevent frequent modifications in the assessment design. Indeed, small-scale administrative tinkering by technocrats was a common feature of the College Exhibition. As defined here, such tinkering included minor changes to assessment content or format, scoring, and the number and value of awards. These changes were commonly made without the benefit of empirical evidence (T&T. Government, 1947, 1950, 1951). For example, in 1947, the redesign led to English and Arithmetic being retained and a third paper covering Geography, History, Hygiene, and Nature Study added. In 1953, another design was implemented so that the constructs then measured included English Language and Grammar, English Composition, Arithmetic, and Geography. The Working Party (1954) linked changes in assessment design to a need to avoid specialization in the primary school and achieve greater alignment with the curriculum.

Post-Independence: Changing Contexts, New Policies, and Rebirth

Three features of the colonial College Exhibition system have been retained in the post-independence secondary school entrance examination systems: (1) high societal legitimacy and valorisation, (2) overall administrative stability and persistence, and (3) experimentation and tinkering without evidence. However, residualism from the colonial era is both contradictory and ambiguous, especially for motives, purposes, and explanations for assessment changes (MacKenzie, 1989; Payne & Barker, 1986). This is because the processes associated with residualism lie outside the rhetoric of modern technical explanations for

education reform, such as efficiency or measurement-driven instruction, but operate at the much deeper level of value assumptions and belief patterns, which are the primary motivators for institutional change in postcolonial systems (Jones, 1975; Jones & Mills, 1976).

MacKenzie (1989) documented the several changes made to the Common Entrance examination in the 1970s and 1980s, most notably the inclusion of Social Studies and Science in the battery of tests, and the introduction of a written essay (T&T. Ministry of Education. Examination Review Committee [Examination Review Committee], 1983a[1], 1983b[2]). London (1997) had rationalized the inclusion of the essay component by pointing to the advantages of measurement-directed reform, whereas the Examination Review Committee (1983b) pointed to advice given by a foreign consultant[3]. Despite several further changes, however, the examination persisted even with the implementation of universal secondary education in 2001[4]. Indeed, some features, such as legitimacy and valorisation by the public have continued unabated. To illustrate, performance in the Secondary Entrance Assessment (SEA) is now associated with scholarships and awards for the best-performing students. The sources of these awards include the Ministry of Education, private sector institutions, and even the regional university (11 in Secondary Entrance Assessment, 2010; SECU awards, 2011; T&T. Government, 2011).

During the post-independence era, there were several proposals for managing adverse consequences arising from the selection and placement mechanism, the most notable being in the Moses and St. Clair King reports (T&T. Cabinet Appointed Committee [Moses], 1975; T&T. Committee Apppointed by Cabinet [King], 1982). However, the National Task Force on Education (T&T. National Task Force, 1994) was much more cautious in recommending a gradual change to the inclusion of scores from continuous assessment and "national tests" to determine selection and placement. There was intense opposition within the Ministry of Education to some of these minor changes, as seen in the case of the "St. Clair King" Working Committee, which had suggested reducing the number of attempts from two to one (King, 1982). This idea was strongly rejected by the Examination Review Committee, which raised concerns about the fairness of such a policy to test takers in general (Examination Review Committee, 1983a[5]). However, although the argument over restricting the number of chances might have centred on fairness, data were never presented by either of the opposing sides. In the modern era, the 1998 Task Force Report (T&T. Task Force) was the source of two notable changes in the examination—the first in 2001 was

a change in format and content, and the second in 2005 was a change in calculating the composite score and the essay rubric.

Most of the debate over these seemingly minor changes did not involve the general public, but was handled "in-house" by the technocracy. Although there was little stakeholder involvement in the 1980s debate over the number of repeats, there was some stakeholder voice in the proposed changes in format and content of 2001. The salient point, however, is that policymaking remained obscure and evidence was rarely used to guide these critical technical decisions. Instead, strongly held personal views or political realities tended to dominate both the dialogue and the policymaking. Sometimes, however, formal positions might be clothed with the logic of educational theory. Such an approach to policymaking appears to be especially common in the opinion-led societies of Latin American and the Caribbean (Cueto, 2005). In such contexts, evidence and indigenous knowledge are scarce, and movers and shakers are able to influence public opinion (Isaac, 2001).

Admittedly, this view of evidence-based policymaking runs counter to the postcolonial analysis of Shahjahan (2011), who considered the current evidence-based rhetoric promoted by the West to be simply another instance of neocolonialism. This paper argues, however, that colonial structures could not be evidence-based and, currently, true evidence-based systems are not easily built within postcolonial societies (Sutcliffe & Court, 2005). Indeed, a review of the history of measurement in the Western world suggests that the evolution of secondary school entrance examinations in the Anglophone Caribbean is very dissimilar to the development of selection testing in the West, especially with regards to use of evidence to inform policy and assessment design (Cole, 1973; Cole & Zieky, 2001; Linn, 1973; Linn & Werts, 1971).

Firstly, then, secondary school entrance examinations in the Caribbean remained relatively immune to evidence, with questions of selection bias (fallibility) infrequently addressed. Secondly, and more importantly, in the Western world, selection testing in education, which developed from intelligence testing for the military, remained intimately connected to a growing measurement community (McArthur, 1983). This academic measurement community retained a significant role in the theoretical development and debates of the 1960s and 1970s. In the Caribbean, however, no such measurement community existed in this period, although several prominent sociologists, educationists, and economists provided useful insight into the nature of selection through rigorous empirical studies (Cross & Schwartzbaum, 1969; Hamilton, 1979; Manley, 1963). Nevertheless, the influence of this scholarly

community on policymaking remained minimal. Thus, what emerged in the 1960s was a selection/placement system in the Caribbean that was relatively immune to evidence, able to perpetuate itself through the non-transparent policymaking cathedrals of Ministries of Education.

In the 1960s, the introduction of the secondary school entrance examinations across several Caribbean countries was associated with increasing numbers of students accessing "free" secondary school places (Alleyne, 1996). The systems in Jamaica, Barbados, and Trinidad and Tobago were the oldest, dating back to the period 1957 to 1962 for first implementation. In both Trinidad and Tobago and Jamaica, the numbers of test takers rose sharply after independence, becoming a significant source of failure for students from the lower socio-economic bracket and from rural areas (Cross & Schwartzbaum, 1969; Mahabir, 1973). An analysis of these studies suggested that student failure was associated with both poor performance and the restricted number of secondary school places. Still, the examination retained credibility in countries like Jamaica and Barbados, perhaps because the test batteries were constructed in England. Nevertheless, both Manley (1963) and the Barbados Committee to Review the Operations and Effects of the Common Entrance Examination (1974) pointed to the inappropriateness of many items. At that time, however, the level of measurement knowledge in the Caribbean did not facilitate an extensive consideration of cultural or any other type of test bias. On a more positive note, the report by the Barbados Committee (1974) included a supplemental statistical report. However, few other government reports of that time made adequate use of empirical evidence, including the 1998 Task Force for the Removal of the Common Entrance Examination in Trinidad and Tobago. This is indeed surprising given the potential access to examination data by Ministry of Education technocrats.

It could be that the examination's credibility remained despite the evidence because tests are widely perceived as neutral tools in the Caribbean, a perception that may be greatly facilitated by the general lack of transparency in the processes associated with public examinations. In the early period after independence, Cross and Schwartzbaum (1969) provided credible evidence highlighting the segregating impact of the examination. The authors analysed the composition of the Form 1 classes, focusing upon socio-economic status and ethnicity. Jules (1994) also studied multiple cohorts and came to a similar conclusion, but used data obtained directly from examination databases. That study was conducted on behalf of the Centre for Ethnic Studies in Trinidad and Tobago, a unit appointed to gather data on the impact of race and ethnicity in the social and educational systems of the

country. Unlike the work of Manley (1963) and Cross and Schwartzbaum (1969), however, Jules did not use an explicit sociological perspective but instead focused on attainment gaps for different target groups. The study disaggregated the data by ethnicity and race, along with other exogenous variables such as gender, socio-economic status, and geographic location. Jules also addressed less frequently studied issues such as transfers and geographic location.

The Jules (1994) study is one of the few empirical studies to have accessed actual examination data. Nevertheless, there were limitations to the methodological approach chosen. Firstly, multivariate analytic techniques, such as logistic regression, were not used. Secondly, interaction effects were largely ignored so that the subtle interactions among ethnicity, social class, and gender remained unclear. Thirdly, effect sizes or odds ratios were not reported, so that judgements cannot be made about the magnitude of the differences. Thus, it might well be that some of the differences found for the ethnic groups were practically small although statistically significant. Likewise, the conclusion that African-Trinidadian females of the lowest economic strata were the most disadvantaged remains questionable because the extent of this disadvantage was not directly measured.

After 1994, there were very few high-quality empirical studies of secondary school entrance examinations. Thus, "new" secondary school entrance examinations emerged in an evidence vacuum. It might be that by changing the name of the examination, some have been led to believe that the issues associated with high-stakes selection are no longer valid. Such a misconception appears to be evident in the report of the 1998 Trinidad and Tobago Task Force, whose terms of reference required them to eliminate the Common Entrance Examination. In the end, the Task Force chose to retain the examination as a placement tool and rename it as the Secondary Entrance Assessment (SEA). This government-commissioned report did little to eliminate local measurement misconceptions, such as the direction and size of the gender gap, fairness of ability testing, role of the essay component in weighting of the composite score, and the utility of standard scores for placement (De Lisle & Smith, 2004).

In Barbados, the name and role of the examination was also changed to the Secondary Schools' Entrance Examination (BSSEE); and in Jamaica, the Grade Six Achievement Test (GSAT) (Barbados. Ministry of Education, Youth Affairs and Culture, 2001; Jamaica. Task Force on Educational Reform, 2004). With hindsight, it would appear that the politics of educational policymaking in Caribbean states might also have been a critical factor. To illustrate, the 1998 Trinidad and Tobago Task

Force did not include Vena Jules, who four years earlier had accessed the Ministry of Education's confidential databases to study the sensitive issue of differentials across ethnic groups. Perhaps more significantly, the Task Force did not include an educational statistician to provide the necessary analysis for evidence-based decision making, as in the case of the Barbados Committee (1974). There is little doubt that such exclusionary practices have contributed to misdirected policymaking.

Another salient issue brought to the fore by the 1998 Trinidad and Tobago Task Force was the popular perception of emerging gender differences in achievement outcomes. The term *notable,* with reference to the size of the gender gap, is used cautiously here because, in reality, there was little large-scale empirical evidence at that time to prove whether or not females had a practically significant advantage in subject areas at the primary school level. Still, the absence of evidence has rarely stopped Caribbean policymakers from proposing drastic solutions to manage issues that they perceive as important, often without regard to the consequences that might accrue (De Lisle & Smith, 2004). The solution to the perceived issue proposed by the committee centred on test redesign for greater perceived gender fairness (Gipps & Murphy, 1994; Willingham & Cole, 1997). De Lisle and his collaborators examined the issue of a female advantage in secondary school entrance examinations, using cohort data from the Trinidad and Tobago Ministry of Education. In a series of published articles from 2004 to 2006, they highlighted the complex nature of the issue and the danger of rushing to judgement based on evidence gathered on small-scale or even single cohort studies (De Lisle, 2006; De Lisle & Smith, 2004; De Lisle, Smith, & Jules, 2005).

Since 2006, additional full cohort data have been made available, facilitating a better judgement on the magnitude of gender differences in the SEA. As shown in Figure 1, the evidence from these full cohort studies confirms the existence of frequent medium-sized gender differentials in Creative Writing and sometimes in Language. However, few Mathematics or composite score differentials reached Cohen's benchmark of 0.5 for a medium-sized effect from 1995 to 2005 (Cohen, 1988). Ironically, however, this data suggest that the change in assessment design proposed in 1998, and implemented in 2001, led directly to the current greater female advantage in Language and Mathematics. This change in assessment design included the use of constructed response instead of multiple-choice items, and the omission of Social Studies and Science measures from the battery of tests, where gender differences were usually negligible or small in the pre-2001 assessment design. The result of these two changes in assessment design

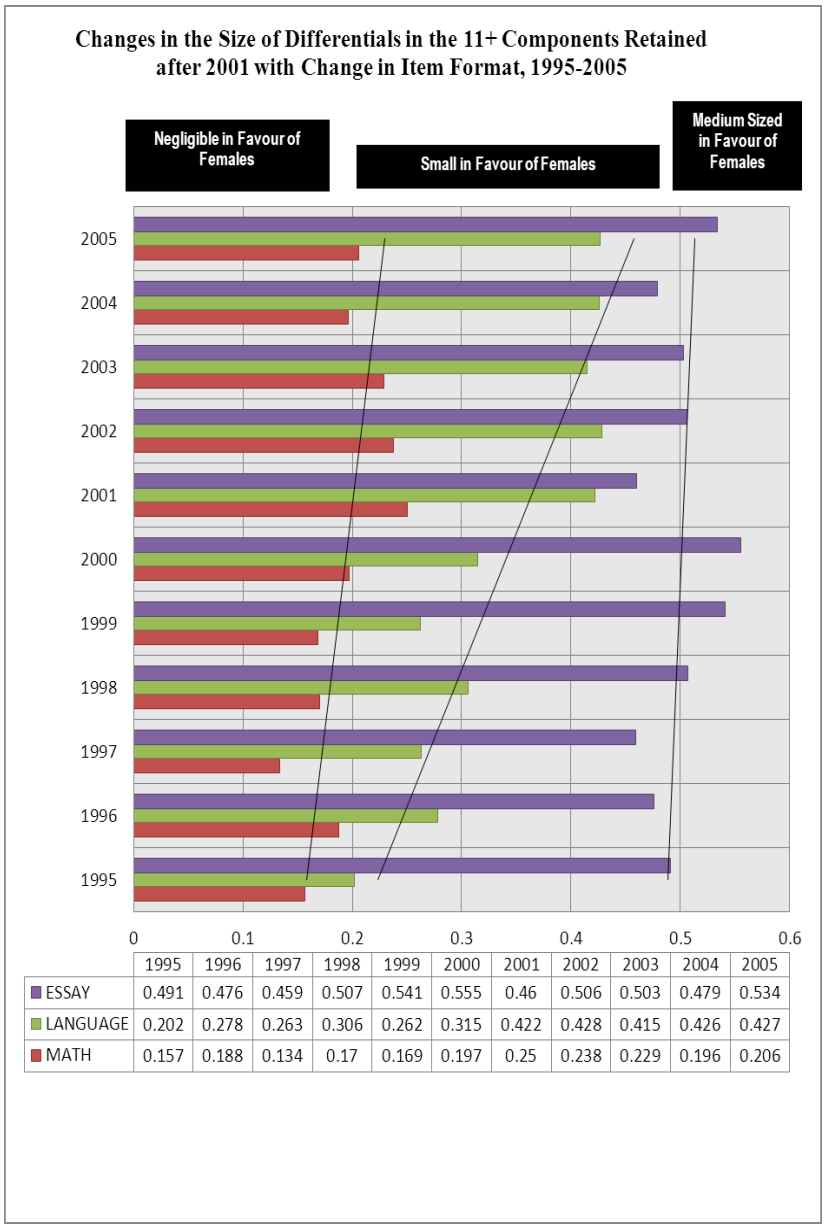
was that the composite score increased sharply, although it is below the benchmark of 0.5. This finding points to the danger of policymaking that excludes evidence (Willingham, 2002; Willingham, Cole, Lewis, & Leung, 1997). Policymaking based on expertise or intuition can lead to changes that worsen the problem they were trying to resolve.

It is clear from the analysis of policy changes in the post-independence era that opinion-led tinkering and experimentation on secondary school entrance examinations have continued. At the same time, there has been little consideration of the possible negative impact of early selection on system stratification or individual mobility. Moreover, to Caribbean policymakers, new education reform concepts like seamlessness are compatible with the retention of secondary school entrance examinations.

Some policymakers have even considered new uses for examinations positioned at the primary-secondary school interface. For example, policymakers from the Organisation of Eastern Caribbean States (OECS) have considered the possibility of regional harmonization (OECS, 2010a; 2010b). This reform being seen as modernist, rational, and perhaps even "decolonizing," as noted by the Head of the OECS Reform Unit:

> We believe that for purposes of comparison and efficiency, and generally for raising the standards, the region needs to institute one examination at the end of the primary level. We know that each country undertakes a very expensive system of developing items, storing items and preparing exams on a yearly basis and we believe that there could be greater efficiency gains if it is done by CXC for the entire region. It will also present the opportunity to compare performances among the different countries as well as establish a standard of assessment for the region at that level as to how our children are performing at the end of primary school. (OECS, 2010b)

From this perspective, then, it appears that current policymakers in Trinidad and Tobago and the Caribbean intend to make the secondary school entrance examinations a permanent and critical part of future schooling. Such a perspective is both naïve and simplistic, because it ignores the measurement limitations of high-stakes achievement tests, the threat of unintended consequences, and the inequalities in opportunities to learn that persist throughout the system. Such proposals position early selection and placement as a useful educational strategy in Caribbean societies, even in the modern era of globalization, and new roles can be successfully attached to the basic sorting function, including regional benchmarking.



	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
ESSAY	0.491	0.476	0.459	0.507	0.541	0.555	0.46	0.506	0.503	0.479	0.534
LANGUAGE	0.202	0.278	0.263	0.306	0.262	0.315	0.422	0.428	0.415	0.426	0.427
MATH	0.157	0.188	0.134	0.17	0.169	0.197	0.25	0.238	0.229	0.196	0.206

Figure 1. Changes in the size of the component scores in the Secondary Entrance Assessment due to changes in assessment design.

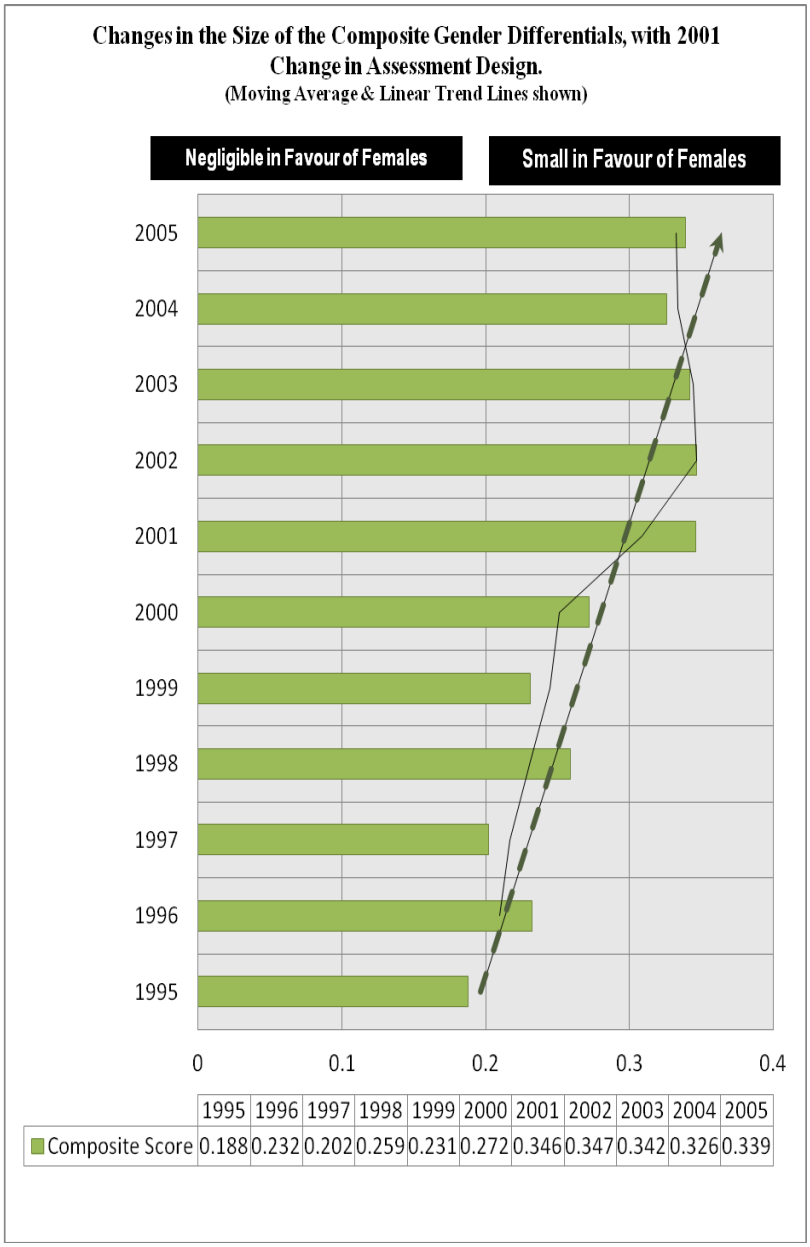


Figure 2. Changes in the size of the composite score used to place students in the Secondary Entrance Assessment due to changes in assessment design.

Judging a Legacy: Evidence for Validity, Fairness, and Equity

Fifty years on and well into the postcolonial era, the question becomes: *How do we judge the legacy of secondary school entrance examinations? What evidence do we solicit for making that judgement*? Evaluation is the process of judging the worth, merit, or quality of an educational innovation using credible evidence. There is little work on evaluating examination systems in the Caribbean, with testing agencies and universities remaining mostly consumers of examination protocols and procedures from elsewhere, rather than inventors or producers of technical knowledge about examinations (De Lisle, 2009). Judging a legacy requires a system to gather credible evidence using a strong evaluative framework. This framework must make use of measurement theory, possibly supported by a socio-historical analysis to illuminate the context. Three broad areas in a proposed evaluative framework are (1) test validity, (2) test fairness, and (3) assessment design. These issues must extend from use of test scores to the study of placement protocols.

In the early 1980s, some proponents of selection testing in developing countries touted measurement-driven instruction as a theoretical rationale to retain these examinations (Chapman & Snyder, 2000; Eisemon, 1990; Heyneman, 1987). For example, Chapman and Snyder outlined five advantages when implementing high-stakes testing:

1. Policymakers can use test scores to target educational resources for low-achieving schools and regions.
2. Testing can be used to shape and "pull" teachers' pedagogical practices in desirable ways.
3. Testing can be used to motivate teachers to improve their teaching.
4. Testing gives the teachers information with which they can target remediation.
5. Assessments can support cross-national comparisons.

The problem with these recommendations is that they do not acknowledge the unintended, negative consequences associated with high stakes (Broadfoot, 2002). Such consequences can nullify or make contentious some of the propositions stated above. The current evidence from studies on washback certainly does not support propositions 2 and 3 (Wall, 2005). These propositions also ignore the issues of appropriate test use and test validity (Bachman, 2005), which are critical issues in educational testing (Messick, 1989).

The assumption that scores from secondary school entrance examinations can facilitate efficient and appropriate placement of students in different types of secondary schools is really a question of test validity, defined as "the overall plausibility of a proposed interpretation or use of test scores" (Kane, 2001, p. 328). However, there are several other concerns to be addressed before validity is fully investigated. One such concern is identifying the construct being tested. This relates to whether the examination measures anything apart from elements associated with socio-economic status (London, 1989).

Thus, secondary school entrance examinations present quite a challenge to developing an argument-based approach to validation because the first step requires clear documentation of the claims made about the examination scores (Kane, 2011). However, the usefulness of the scores for placing students into different school types is not the only claim made about this examination. Increasingly, scores from secondary school entrance examinations are also used for additional purposes such as monitoring standards, and judging quality and equity in the system (Planning Institute of Jamaica, 2009). In Trinidad and Tobago, the accountability function is evident in the use of the 30% capricious cutscore to determine which schools are failing, and assigning students under this score to remediation status.

Each of these new purposes is a documented claim and must be validated along with the primary placement function. Indeed, an argument-based validation framework for the secondary school entrance examination would consider each claim, providing warrants, rebuttals, and evidence (Chapelle, Enright, & Jamieson, 2008; Kane, 1992, 2002, 2004, 2006). Bachman (2005) has argued that a claim might centre on what the test performance is supposed to measure and the decision made in response to the assessment-based information. Consider Figure 3, which provides a sample argument structure for the primary claim found in the 1998 Task Force for the Removal of the Common Entrance Examination, namely, that the examination can place students of different abilities within schools of different quality. Unpacking the argument shows that there are really two claims being made: (1) that the examination can measure ability, and (2) that measuring ability or ability by proxy allows us to make efficient school placement decisions.

The argument structure in Figure 3 is based on the work of Llosa (2008) and Chapelle, Enright, and Jamieson (2010). The claim is the intended interpretation about the student and of test use; the grounds are the basis for making the claim; the warrant is the proposition used to justify the inference from the grounds to the claim; the rebuttal is an alternative explanation; and the backing is the evidence to support the

warrant. In Figure 3, both an anecdotal and an evidence-based backing for the warrants and rebuttals are provided. Anecdotal backings may be important in opinion-based postcolonial societies. Presenting the entire validity argument in this way guides the systematic selection of evidence in validation (Llosa, 2008). Although the claims, grounds, and backing appear to be coherent, the rebuttal suggests that the original claims might not be solid. Moreover, presenting criterion-related validity evidence will be insufficient to support the claim, since the rebuttal demands exploration of the processes occurring within each school type. Put simply, saying that students who attend different secondary school types are less likely to do well is not an adequate justification for the placement process, because an alternative explanation is that teacher expectations and behaviour are the causal antecedents. Thus, credible evidence for the claim requires analysis of data from qualitative or mixed methods studies on equal opportunities to learn before and after the secondary school entrance examination.

An issue closely connected to test validity and fundamental to efficient selection within a multicultural society is test fairness (Stobart, 2005). London (1997) has noted that both policymakers and the public have promoted this aspect of examination-based selection in their rhetoric. Thus he observed that:

> The Common Entrance Examination is a key component in attempts to establish an egalitarian society in Trinidad and Tobago. In this connection state managers have agreed that "all children pass through the same door to life" (quoted in Garcia 1985, 6), and that through its principles for allocating secondary school places, the Common Entrance Examination offers some degree of fairness in achieving this ideal. Support for the examination also comes from parents. They believe that despite the social difficulties that the test might pose, the device serves as an opportunity for their children to escape poverty and class prejudice which in the past hindered socioeconomic advancement of the poor and the hapless. "The examination is not perfect," argues a leading daily newspaper, "but no-one has suggested a better system of selection." (p. 67)

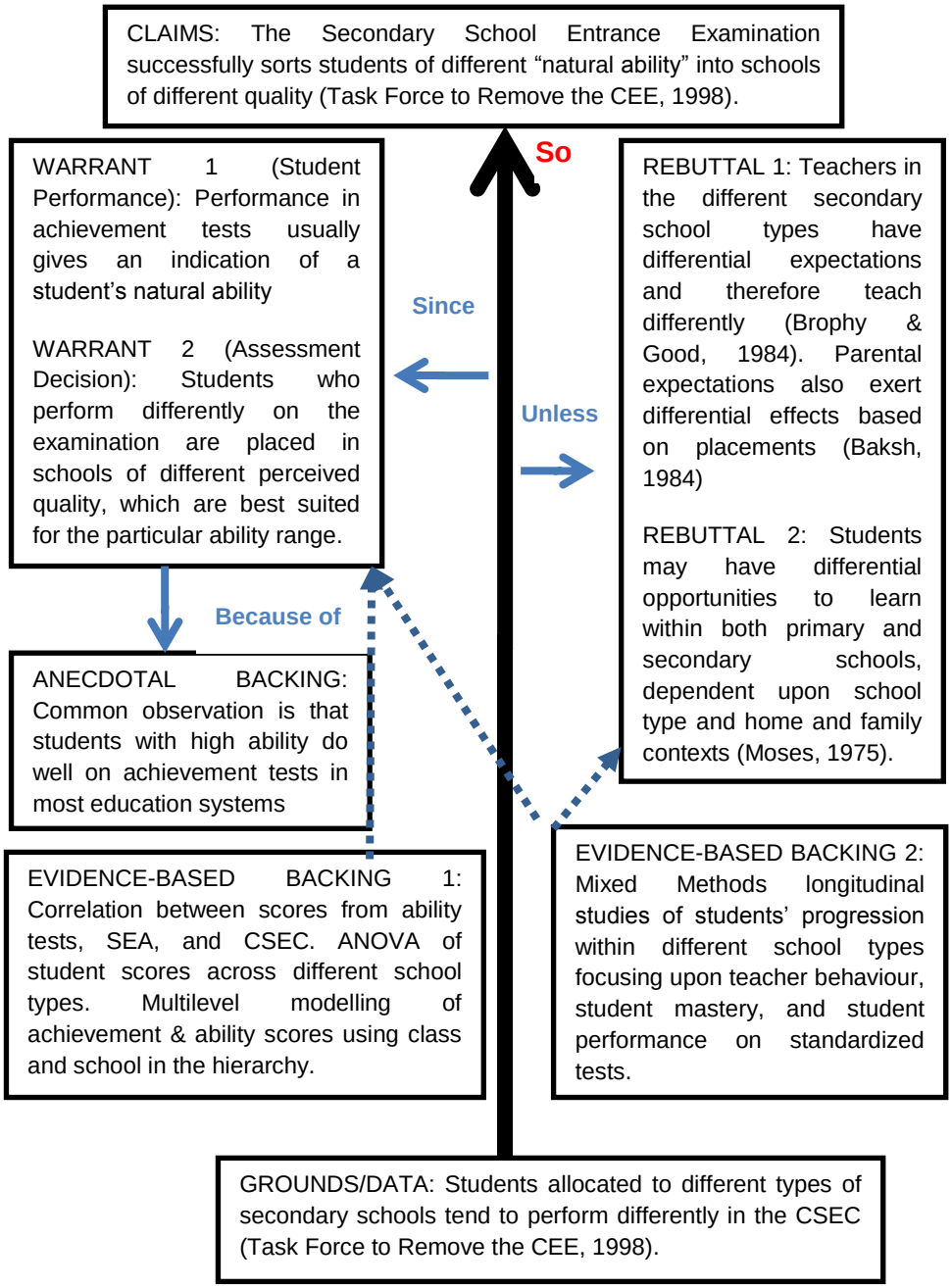


Figure 3. Illustration of a validity argument for the Secondary School Entrance Examination.

Such claims are widespread in the current societal setting, but to be validated, evidence must be gathered using an explicit argument-based validity framework or a separate fairness-based approach (Willingham, 1999). There are three possible relationships between validity and fairness. As proposed by Xi (2010), there might be no direct relationship between validity and fairness; a second sees fairness as an all-encompassing characteristic subsuming validity; and the third considers validity as the fundamental test characteristic with fairness subsumed under it. I would argue that fairness considerations cannot fully be captured by a pure validity framework, because there are dimensions of fairness not directly related to the psychometric properties of the test. Indeed, in a multicultural society like Trinidad and Tobago, fairness is critical and may be considered as the overarching feature of any selection system (Stobart, 2005). The fairness of a test-based selection system must draw on measurement theory as well as constructs in sociology and economics, with moral, societal, and political dimensions made clear.

Fairness is more than the actual test (as in bias) or its administration (as in equal treatment of test takers), and extends to opportunities to learn prior to the test (AERA, APA, & NCME, 1999). For test-based selection systems, fairness may also be regarded as equitable treatment of relevant groups, with no group having an advantage based on extraneous characteristics related to the group's identity (Paes de Barros, Ferreira, Molinas Vega, & Saavedra Chanduvi, 2009). Cole and Zieky (2001) recounted the development of fairness concerns among the measurement community in the United States, noting a period of little interest prior to the civil rights era of the 1960s. However, in the Anglophone Caribbean, fairness considerations were critical even in colonial times (Campbell, 1996). Such interest might have been fuelled by the segregated nature of the society. Indeed, examinations such as the College Exhibition were really responses to the demands for greater equity and fairness in these societies. Debate on fairness was especially intense during the period of nationalism in the 1950s and post-independence in the 1960s (Campbell, 1997). In these periods, the theme was even broached in the political campaigns. However, the absence of a Caribbean measurement community meant that these concerns were not supported by theory or evidence.

In the past, some have considered the secondary school entrance examinations to be fair, simply because the examination agency was external to the country—Education Testing Services in the past and now the Caribbean Examinations Council (Examination Review Committee, 1983b). However, the perception that the examination is defensible and impartial merely because it is scored by an apparently neutral agency is

surely flawed. Surprisingly, unfailing trust in the test scores and placement process has remained despite known imperfections in the 20% allocation and transfers (London, 1989). There is also a persistent belief that the examination is able to measure "merit" precisely. However, even if merit could be defined and measured, and there were equal opportunities to learn, a test-based system would still be fallible (Gardner & Cowan, 2005). Fairness theory helps to explain the continued perceived credibility of secondary school entrance examinations in the Caribbean. This subset of justice theory posits that for discrete events, individuals first gauge how that particular decision event affects them. If the decision is negative, these individuals will first try to blame decision makers, if they can hold them accountable. Individuals form "would," "could," and "should" counterfactuals in the process of assigning this blame (LaHuis, MacLane, & Schlessman, 2007). The tendency to form all three counterfactuals might be low in the current context because of prevailing perceptions of infallibility and system-wide lack of transparency. Thus, stakeholders might be prone to assign blame to the test takers' ability rather than to the examination process or agency.

In judging the fairness of test-based selection systems, both distributive and procedural justice principles must be considered. There is little empirical support for believing that test-based selection can be neutral to group differences (Cross & Schwartzbaum, 1969; De Lisle, Smith, Keller, & Jules, 2012; Jules, 1994; Manley, 1963). In a region already distracted by the size of differentials favouring females, implementing a test design that will further exacerbate differences might be considered foolhardy. Indeed, Willingham and Cole (1997) suggested that investigating alternative test designs provides a unique opportunity to use evidence when considering comparative gender fairness. An important consideration for procedural fairness is the operation of the school choice system. Both the test and the choice/placement systems are intimately intertwined in several other Caribbean countries, and the functioning of these different systems and their rules must be fully explored.

The decision to choose a secondary school is multi-staged and complex; not easily defined by the simple rules constructed by the Ministry of Education to guide decision-making (De Lisle, Keller, Jules, & Smith, 2009). Family choice of secondary school is also dynamic, responsive to the changing perception of the value of institutions and their physical location. It must be noted, however, that in the end, most students are assigned by the Ministry of Education to a secondary school and, therefore, to a large extent, the availability of choice is more imaginary than real, especially for students of low economic status. This

reality is shown clearly in Figure 4. The neighbourhoods chosen for this quantitative case study are located within the Diego Martin Administrative Region and are all in close proximity. The neighbourhoods are ranked by the Basic Needs Index (BNI), a measure that captures the economic and social conditions of the inhabitants (Kairi Consultants Ltd., 2007). Qualitatively, there are five high-income neighbourhoods, one middle class neighbourhood, and three low-income neighbourhoods—Carenage, La Puerta, and Point Cumana. In these highly populated low-income communities, the numbers receiving their first choice in school ranged from 3.6% to 16.1% compared to 21.4% to 35.0% placed directly by the Ministry of Education.

How is fairness related to equity—a moral and political concept? Equity implies fair treatment of different groups within a society. Critical to such fair treatment is not just administering the test, but ensuring equal opportunity to learn the content. At the same time, equity is not just a matter of numbers. The judgement of inequity assumes the operation of an extraneous characteristic like gender, ethnicity or religion.

At the same time, however, large inequalities (magnitude only) in a society might also be considered as morally unacceptable. Levin (2003) thus sees two broad issues in equity: (1) whether overall levels of provision are sufficient, and (2) whether they are of the right kind. These are broadly equivalent to the key sociological themes of access, opportunity, and attainment. These issues are embedded within the test-based selection system because of variability in the quality of education at the primary school level. Although early system reforms focused on improving and equalizing access and equalizing opportunity, dissimilar opportunities for learning remain even at the basic level of schooling, contributing to differences in attainment on the test.

So how rigorous and precise is the current secondary school entrance examination system? The level of precision is a good litmus test for judging fallibility, pointing perhaps to measurement error. Kane (2011) reminded us that large or random errors will challenge interpretations and inferences based on the test scores. In Trinidad and Tobago, a notable challenge to precision was raised in 2003 when several families requested a remark of the SEA. The issue was brought forward to the Trinidad and Tobago House of Representatives and is recorded in the Hansard (T&T. House of Representatives, 2003).

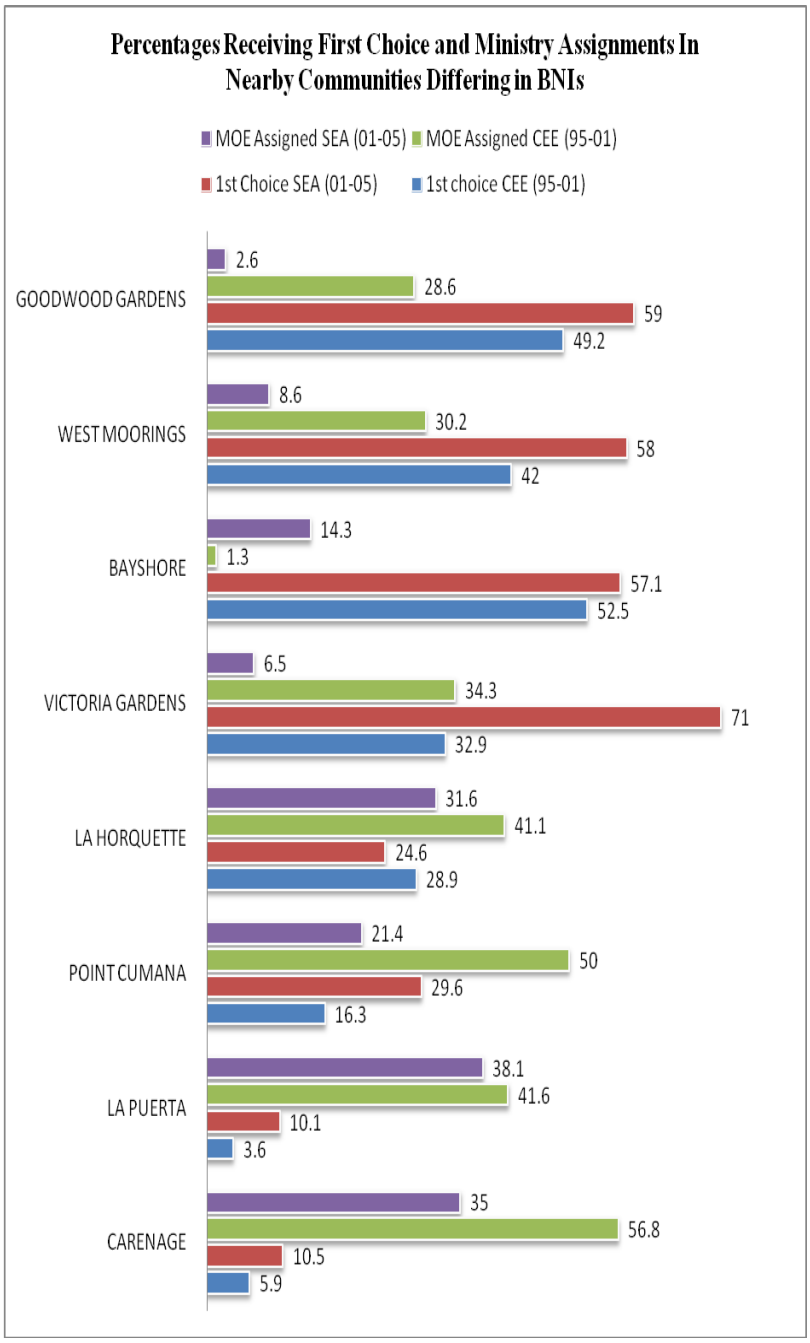


Figure 4. Percentages of students receiving their first choice in nearby communities in the Diego Martin Administrative Region (1995-2005).

The specific question raised was:

> Would the Hon. Minister of Education provide a list of the names of the students who were upgraded following complaints by their parents to the placement of their children following the 2002 SEA examination and identify the secondary schools involved? (p. 46)

A written response was provided, which described the procedure for remarking used by the testing agency. In 2002, there were 212 requests for reviews, and as many as 56 students were reassigned based on a new score in the remarking (26%) (T&T. House of Representatives, 2003). Figure 5 maps the approved changes in placement. As shown, the majority of the students were reassigned from 5-year denominational and new sector schools to 7-year denominational schools. However, there appears to be a tiered system operating, with students moving from the new sector schools to the 5-year schools, and students from the 5-year schools transferring to the 7-year schools. The number of successful transfers in this instance suggests that the test-based placement system is not as precise as many people believe. Interestingly seven students who were remarked and awarded new scores were not reassigned, including three for whom no spaces could be found. How fair can that be?

In Search of Evidence: Evaluation and Benchmarking for Future Change

Thus, the future of policy on the use of secondary school entrance examinations lies not in further tinkering or experimentation without evidence, but in developing a robust ongoing monitoring and evaluation system to provide information on examination processes and outcomes. A system providing evidence to guide policymakers and facilitating the evaluation of alternatives is the essence of empowerment and rational thought that is embodied in progressive postcolonial thought. In reality, then, using evidence to make decisions in the post-independence era is a critical element in the decolonization process. Arguably, then, the construction of high-quality systems for generating and using evidence might be the key to helping postcolonial societies break free from colonial structures and processes. Segone (2009) and Sutcliffe and Court (2005, 2006) have documented the value of an evidence-based approach to policymaking in developed countries. Schleicher (2009) confirmed that international benchmarking evidence has the capacity to shed light on education issues and solutions that would otherwise be hidden without the data. Indigenous evidence will certainly allow postcolonial societies

to achieve more effective context-relevant implementation and change (Louisy, 2004).

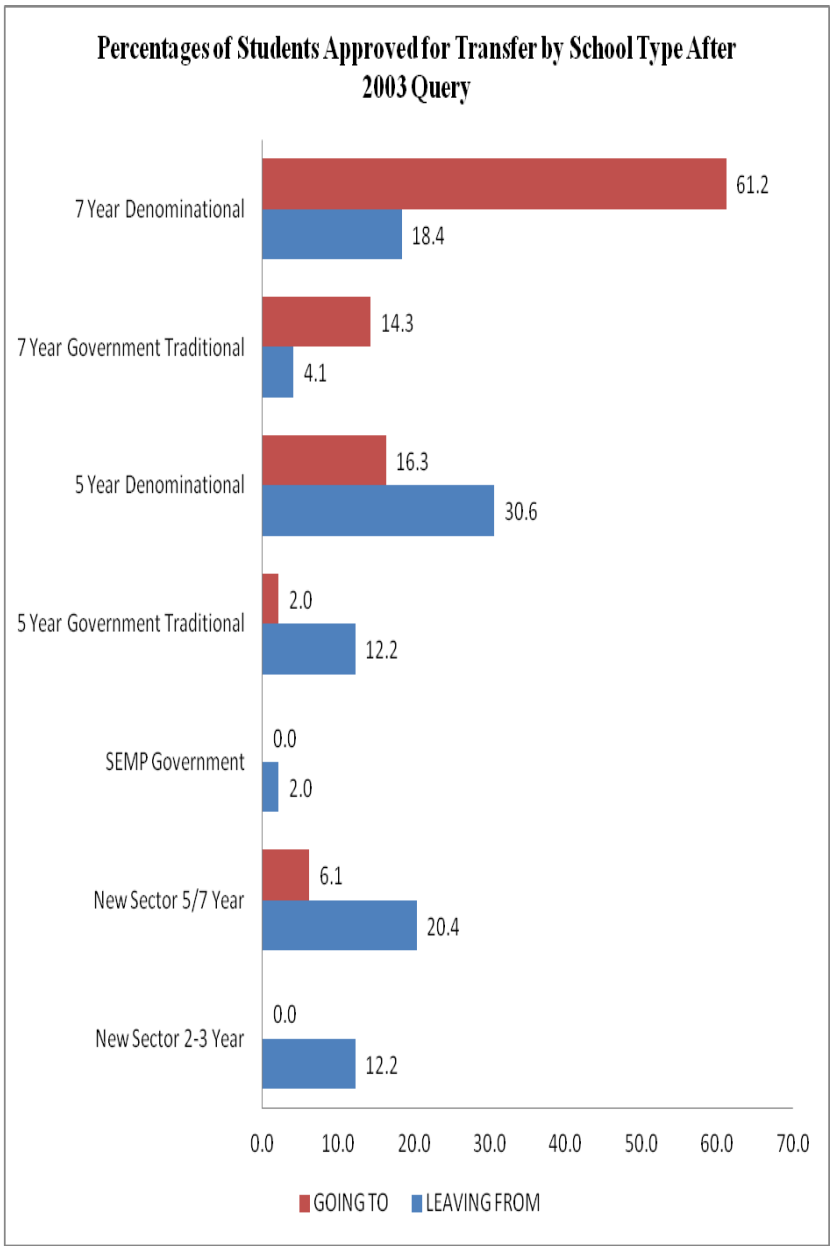


Figure 5. Percentages of students reassigned after remarking in the 2003 query issue recorded in the Hansard. (T&T. House of Representatives, 2003)

Of course, the capacity to provide evidence is one area in which Caribbean systems have traditionally been very deficient (Crossley, 2008; Crossley & Holmes, 2001; Lewis & Simmons, 2010). This is true even for established research or examination centres such as The University if the West Indies (UWI) and regional agencies like CARICOM and the Caribbean Examinations Council (CXC). For example, even now, there is limited published work on regional secondary school entrance examinations even though they are managed by a single regional agency. Yet, information on the functioning of examinations is needed to move the system forward. Information in a transparent system can create disequilibrium, a condition needed to motivate stakeholders to seek change. Caribbean agencies cannot remain consumers of information from foreign sources because most education systems are fundamentally different.

However, Caribbean education systems do not have to reinvent the wheel in pursuing assessment reform. As argued by Schleicher (2009), much can be learnt by benchmarking best practice in high-performing systems and recontextualizing policy guided by indigenous knowledge (Luke, 2011). Thus, the Caribbean can learn much from the trajectory of reform in the test-oriented societies of Shanghai, Singapore, and Hong Kong. In the case of Hong Kong, a decision was taken in the 1970s to replace the Secondary School Entrance Examination (SSEE). Likewise, Shanghai has recently abandoned early selection, although the rest of China continues to employ it. In Hong Kong, the General Schools Circular 10/76 first replaced the Hong Kong SSEE with internal assessments moderated by an Academic Aptitude Test (AAT) (Tam, 1977). Of course, this solution involved two contentious issues: (1) whether aptitude tests can measure future ability, and (2) whether school-based assessments can be conducted with fidelity when high stakes are attached. Since then, however, several additional evidence-based reforms have helped Hong Kong to construct an improved system. It would seem, then, that the key to successful assessment change is comprehensive reform that includes structural and policy changes aligned to examination reform.

This is not to say that removing early selection was done without much public dissent in Hong Kong. Cheung (1979), for example, observed that:

> The outcries for the resumption of the S.S.E.E. are mostly based on selfish reasons. Since it is free junior secondary education for all and most of the prestigious schools in which parents wish to have their children enrolled are publicly financed, children of all

> categories and abilities should have the privilege to study in these schools. To maintain the S.S.E.E. and its bad effects on teaching and learning simply to enable some children to get into the secondary schools of their choice is too selfish and short-sighted. In the long run, the new system of allocation will even up the standard of schools in Hong Kong at both the primary and the secondary levels. If education is not for the elite only, it really should not matter that much where a child receives his/her schooling. (p. 145)

Cheung (1979) correctly positioned the need for the reform on the future performance of the education system. The problem in the Anglophone Caribbean is that most policymakers have not considered reform from the perspective of future global competitiveness. Thus, they have been at great pains to retain placement systems, regardless of the merits or demerits of the new assessment designs.

Can Caribbean societies successfully confront an issue so deeply rooted within the postcolonial culture and mindset? Radical education change and bold decision making is needed, as seen in the case of both Hong Kong and Shanghai:

> In both Shanghai and Hong Kong, deep cultural influences in values surrounding education (such as the emphasis on exams) have been perceived as problems and have provoked a reaction in order to modernise the system: moving from elite to massive popular education, from emphasis on teaching to emphasis on learning, from fact memorisation to development of learning capacities, and from economic needs to individual needs. In both cases, the change in the nature and orientation of the entire education system involves struggles against the culture. (Organisation for Economic Co-operation and Development [OECD], 2011, p. 106)

From this perspective, there is need for a clear coherent unified policy on the direction of education and the role of assessment in the Anglophone Caribbean. One might postulate that differences in Caribbean economies might lead to the question:

> *To what extent can the faster growing, more industrialized Caribbean economies adequately stand on the platform of "harmonized" secondary school entrance examinations?*

The evidence from international assessments suggests that these higher-performing economies will not progress much further, nor adapt

sufficiently to the demands of a global economy, if they choose to retain the colonial impediments of early selection and highly differentiated school systems (Buchmann & Hanuman, 2001; Demeuse & Baye, 2008).

Nevertheless, there might also be alternative pathways to assessment reform, considering the experiences of other high-performing school systems such as Singapore and Germany. Singapore has continued to administer the high-stakes Primary School Leaving Examination (PSLE) even while maintaining high performance in international assessments. However, since 1997, it has also vigorously pursued a policy of formative classroom assessment under the successive national school philosophies of Thinking Schools, Learning Nations (TSLN), and, later, Teach Less, Learn More (TLLM) (Tan, 2011). In terms of administrative reform of the placement system, some students are now allowed direct entry into some secondary schools. Likewise, the early selection system of the three-tier German system is now being diversified in different independent provinces (OECD, 2011). Therefore, there is no one best solution to the issue of reform when considering secondary school entrance examinations. What is required is full stakeholder participation, with dialogue informed by the data, an approach proposed in the 1993-2003 Task Force report (T&T. National Task Force, 1994).

Notes

1. The First Report of the Examination Review Committee recommended that the Common Entrance, in its present structure, be modified to broaden its content by incorporating General Science and Social Studies into the existing scheme of the examination.
2. The Fourth Report of the Examination Review Committee recommended that an "essay-writing" component be introduced into the scheme of the Common Entrance Examination to replace the existing Section III multiple-choice "writing skills" component, and that the Educational Testing Service of Princeton, New Jersey carry out the essay exercise—production, printing, scoring, processing detailed in paragraph 21 (1 to 6 initially).
3. The consultant was Dr. R.T. Green, an American who wrote a memorandum back in 1966.
4. The 30% cutscore on the examination is still used to retain some students in the primary school.
5. The Second Report of the Examination Review Committee recommended the rejection of the St Clair King Working Committee recommendation that "the two chances at the Common Entrance Examination be abolished" and replaced by a single chance, and made counter proposals.

References

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115–129.

Alleyne, M. H. McD. (1996). *Nationhood from the schoolbag: A historical analysis of the development of secondary education in Trinidad and Tobago*. Washington, DC: OAS.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing* (2nd ed.). Washington, DC: American Educational Research Association.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly, 2*(1), 1–34.

Baksh, I. J. (1984). Factors influencing occupational expectations of secondary school students in Trinidad and Tobago. *Social and Economic Studies, 33*(3), 1–29.

Barbados. Committee to Review the Operations and Effects of the Common Entrance Examination. (1974). *Report: Transfer from primary to secondary education in Barbados*. Bridgetown, Barbados: Ministry of Education. Chairman: Leonard L. Shorey.

Barbados. Ministry of Education, Youth Affairs and Culture. (2001). *The development of education: National report of Barbados 2001*. Bridgetown, Barbados: Author.

Brereton, B. (2007). Contesting the past: Narratives of Trinidad & Tobago history. *New West Indian Guide, 81*(3 & 4), 169–196.

Broadfoot, P. (2002). Beware the consequences of assessment! *Assessment in Education: Principles, Policy and Practice, 9*(3), 285–288.

Brophy, J., & Good, L. E. (1984). *Teacher behaviour and student achievement* (Occasional Paper No. 73). East Lansing, MI: Institute for Research on Teaching, Michigan State University.

Buchmann, C., & Hannum, E. (2001). Education and stratification in developing countries: A review of theories and research. *Annual Review of Sociology, 27*, 77–102.

Campbell, C. C. (1983). The College Exhibition in Trinidad and Tobago 1872–1938. *History Teachers' Journal, 2*, 43–52.

Campbell, C. C. (1996). *The young colonials: A social history of education in Trinidad and Tobago 1834–1939*. Mona, Jamaica: The Press UWI.

Campbell, C. C. (1997). *Endless education: Main currents in the education system of modern Trinidad and Tobago 1939–1986*. Mona, Jamaica: The Press UWI.

Chapelle, C. A., Enright, M. K., & Jamieson, J. (Eds.). (2008). *Building a validity argument for the Test of English as a Foreign Language*. London, UK: Routledge.

Chapelle, C. A., Enright, M. K., Jamieson, J. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues & Practice, 29*(1), 3–13.

Chapman, D. W., & Snyder Jr., C. W. (2000). Can high stakes national testing improve instruction: Reexamining conventional wisdom. *International Journal of Educational Development, 20*(6), 457–474.

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language & Education, 11*, 38–54.

Cheng, L., Watanabe, Y., & Curtis, A. (Eds.). (1994). *Washback in language testing*. Mahwah, NJ: Lawrence Erlbaum.

Cheung, D. (1979). A few thoughts on free junior secondary education for all. *Education Journal, 7*, 145–148.

Cohen, J. *(*1988*). Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cole, N. S. (1973). Bias in selection. *Journal of Educational Measurement, 10*(4), 237–255.

Cole, N. S., & Zieky, M. J. (2001). The new faces of fairness. *Journal of Educational Measurement, 38*(4), 369–382.

Cross, M., & Schwartzbaum, A. M. (1969). Social mobility and secondary school selection in Trinidad and Tobago. *Social and Economic Studies, 18*(2), 189–207.

Crossley, M. (2008). The advancement of educational research in small states. *Comparative Education, 44*(2), 247–254.

Crossley, M., & Holmes, K. (2001). Challenges for educational research: International development, partnerships and capacity building in small states. *Oxford Review of Education, 27*(3), 395–409.

Cueto, S. (2005). *Empirical information and the development of educational policies in Latin America.* Paper presented at the Meeting of the Regional Dialogue on Education of the Inter-American Development Bank. Retrieved from www.iadb.org/document.cfm?id=644579

Demeuse, M., & Baye, A. (2008). Measuring and comparing the equity of education systems in Europe. In N.C. Soguel & P. Jaccard (Eds.), *Governance and performance of education systems* (pp. 85–106). Dordrecht, Netherlands: Springer.

De Lisle, J. (2006). Dragging eleven-plus measurement practice into the fourth quadrant: The Trinidad and Tobago SEA as a gendered sieve. *Caribbean Curriculum, 13*, 89–101.

De Lisle, J. (2009). External examinations beyond national boundaries: Trinidad and Tobago and the Caribbean Examinations Council. In B. Vlaardingerbroek & N. Taylor (Eds.), *Secondary school external examination systems: Reliability, robustness and resilience* (pp. 265–290). New York, NY: Cambria Press.

De Lisle, J., Keller, C., Jules, V., & Smith, P. (2009). When choosing might mean losing: A mixed method study of secondary school choice in the Republic of Trinidad and Tobago. *Caribbean Curriculum, 16*(1), 131–176.

De Lisle, J., & Smith, P. (2004). Reconsidering the consequences: Gender differentials in performance and placement in the 2001 SEA. *Caribbean Curriculum, 11*, 23–55.

De Lisle, J., Smith, P., &. Jules, V. (2005). Which males or females are most at risk and on what? An analysis of gender differentials within the primary school system of Trinidad and Tobago. *Educational Studies 31*(4), 393–418.

De Lisle, J., Smith, P., Keller, C., & Jules, V. (2012). Differential outcomes in high-stakes eleven plus testing: The role of gender, geography, and assessment design in Trinidad and Tobago. *Assessment in Education: Principles, Policy, & Practice, 19*(1), 45–64.

Eisemon, T. O. (1990). Examination policies to strengthen primary schooling in African countries. *International Journal of Educational Development, 10*(1), 69–82.

11 in Secondary Entrance Assessment (SEA) top ten (2010, October 10). *UWI Today*. Retrieved from http://ufdcimages.uflib.ufl.edu/UF/00/09/41/80/00024/00010-2010_Page_04.jpg

Ellis, P., Ramsay, A., & Small, S. (2000). *Adult education in the Caribbean at the turn of the century.* Kingston, Jamaica: Office of the Unesco Representative in the Caribbean.

Gardner, J., & Cowan, P. (2005). The fallibility of high stakes '11-plus' testing in Northern Ireland. *Assessment in Education, 12*(2), 145–165.

Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement, and equity*. Buckingham, UK: Open University.

Hamilton, M. A. (1979). An investigation of the relationship between social class, success in the Common Entrance Examination and performance in the GCE `O' level examinations, Jamaica, 1975. *Social and Economic Studies, 28*(3), 499–515.

Heyneman, S. P. (1987). *Uses of examinations in developing countries: Selection, research, and education sector management*. Washington, DC: World Bank.

Isaac, A. A. (2001). *Education reform in the Eastern Caribbean: Implications of a policy and decision making program by an external donor* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. NQ75645)

Jamaica. Task Force on Educational Reform. (2004). *Report: A transformed education system*. Kingston, Jamaica: Author. Chairman: Dr. Rae Davis.

Jones, E. (1975). Tendencies and change in Caribbean administrative systems. *Social and Economic Studies, 24*(2), 239–246.

Jones, E., & Mills, G. E. (1976). Institutional innovation and change in the Commonwealth Caribbean. *Social and Economic Studies, 25*, 323–346.

Jules, V. (1994). *A study of the secondary school population in Trinidad and Tobago: Placement patterns and practices – A research report*. St. Augustine, Trinidad: Centre for Ethnic Studies, UWI.

Kairi Consultants Ltd. (2007). *Analysis of the 2005 survey of living conditions in Trinidad and Tobago*. Port of Spain, Trinidad: Ministry of Social Development.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*(3), 527–535.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*(4), 319–342.

Kane, M. T. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues & Practice, 21*(1), 31–41.

Kane, M. T. (2004). Certification testing as an illustration of argument based validation. *Measurement, 2*(3), 135–170.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: Greenwood Publishing.

Kane, M. T. (2011). The errors of our ways. *Journal of Educational Measurement, 48*, 12–30.

LaHuis, D. M., MacLane, C. N., & Schlessman, B. R. (2007). Do applicants' perceptions matter? Investigating reapplication behavior using fairness theory. *International Journal of Selection and Assessment, 15*(4), 383–393.

Levin, B. (2003). *Approaches to equity in lifelong learning*. Paper commissioned by the Education and Training Policy Division, OECD, for the Equity in Education Thematic Review. Paris, France: Organisation for Economic Co-operation and Development.

Lewis, T., & Simmons, L. (2010). Creating research culture in Caribbean universities. *International Journal of Educational Development, 30*(4), 337–344.

Linn, R. L. (1973). Fair test use in selection. *Review of Educational Research, 43*, 139–161.

Linn, R. L., & Werts, C. E. (1971). Considerations for studies of test bias. *Journal of Educational Measurement, 8*, 1–4.

Llosa, L. (2008). Building and supporting a validity argument for a standards-based classroom assessment of English proficiency based on teacher judgments. *Educational Measurement: Issues & Practice, 27*(3), 32–42.

London, N. A. (1989). Selecting students for secondary education in a developing society: The case of Trinidad and Tobago. *McGill Journal of Education, 24*(3), 281–291.

London, N. A. (1997). Socio-politics in effective curriculum change in a less-developed country: Trinidad and Tobago. *Curriculum Inquiry 27*(1), 63–80.

London, N. A. (2003). Ideology and politics in English-language education in Trinidad and Tobago: The colonial experience and a postcolonial critique. *Comparative Education Review, 47*(3), 287–320.

Louisy, P. (2004). Whose context for what quality? Informing education strategies for the Caribbean. *Compare, 34*(3), 285–292.

Luke, A. (2011). Generalizing across borders: Policy and the limits of educational science. *Educational Researcher, 40*(8), 367–377.

MacKenzie, C. G. (1989). The eleven-plus examination in developing countries: A case study. *Educational Studies, 15*(3), 281–300.

Mahabir, H. G. (1973). *A study of elementary students coming from varying socio-economic backgrounds in urban and rural areas of Trinidad and Tobago and the effect these backgrounds have on performance in the Eleven Plus (11+) examinations.* Unpublished doctoral dissertation, Boston University, Boston.

Manley, D. R. (1963). Mental ability in Jamaica: An examination of the performance of children in the Jamaica Common Entrance Examination, 1959. *Social and Economic Studies, 12*(1), 51–71.

McArthur, D. L. (1983). *Educational testing and measurement: A brief history* (CSE Report No. 126). Los Angeles, CA: Center for the Study of Evaluation, UCLA.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Washington, DC: American Council on Education.

Messick, S. (1998). *Consequences of test interpretation and use: The fusion of validity and values in psychological assessment* (RR-989-4). Princeton, NJ: Educational Testing Service.

Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement, 33*, 379–416.

Miyazaki, I. (1976). *China's examination hell: The Civil Service examinations of Imperial China*. New York, NY: Weatherhill.

O'Callaghan, M. (2009, December 28). The McShines and rise of middle class. *Trinidad Newsday*. Retrieved from http://www.newsday.co.tt/commentary/0,113262.html

Olmedilla, J. M. M. (1992). Tradition and change in national examination systems: A comparison of Mediterranean and Anglo-Saxon countries. In M. A. Eckstein & H. J. Noah (Eds.), *Comparative and international studies* (pp. 135–147). Oxford, UK: Pergamon.

Organisation for Economic Co-operation and Development. (2011). *Strong performers and successful reformers in education: Lessons from PISA for the United States*. Paris, France: Author.

Organisation of Eastern Caribbean States. (2010a). *Harmonized Common Entrance Examination to be discussed by OECS Ministers for Education* [Press Release]. Retrieved from http://www.caribbeanpressreleases.com/articles/7422/1/Harmonized-Common-Entrance-Examination-to-be-discussed-by-OECS-Ministers-for-Education-/Page1.html

Organisation of Eastern Caribbean States. (2010b). *New exam for primary school leavers in the OECS to take place in 2012* [Press Release]. Retrieved from http://www.caribbeanpressreleases.com/articles/7506/1/New-Exam-for-Primary-School-leavers-in-the-OECS-to-take-place-in-2012-/Page1.html

Paes de Barros, R., Ferreira, F. H. G., Molinas Vega, J. R., & Saavedra Chanduvi, J. (Eds.). (2009). *Measuring inequality of opportunities in Latin America and the Caribbean.* Washington, DC: World Bank.

Payne, M. A., & Barker, D. (1986). Still preparing children for the 11+: Perceptions of parental behaviour in Barbados. *Educational Studies, 12*(3), 313–325.

Planning Institute of Jamaica. (2009). *National report of Jamaica on millennium development goals for the UN Economic and Social Council annual ministerial review.* Kingston, Jamaica: Author.

Schleicher, A. (2009). International benchmarking as a lever for policy reform. In A. Hargreaves & M. Fullan (Eds.), *Change wars* (pp. 97–115). Bloomington, IN: Solution Tree.

SECU awards excellence in SEA and CAPE (2011, October 4). *Trinidad Guardian.* Retrieved from http://www.guardian.co.tt/lifestyle/2011/10/04/secu-awards-excellence-sea-and-cape

Segone, M. (2009). Enhancing evidence-based policy making through country-led monitoring and evaluation systems. In M. Segone (Ed.), *Country-led monitoring and evaluation systems. Better evidence, better policies, better development results* (pp. 17–33). Geneva, Switzerland: UNICEF.

Shahjahan, R. A. (2011). Decolonizing the evidence-based education and policy movement: Revealing the colonial vestiges in policy research, and neoliberal reform. *Journal of Educational Policy, 26*(2), 181–206.

Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education: Principles, Policy, & Practice, 12*(3), 275–287.

Sutcliffe, S., & Court, J. (2005). *Evidence-based policymaking: What is it? How does it work? What relevance for developing countries?* London, UK: Overseas Development Institute.

Sutcliffe, S., & Court, J. A. (2006). *Toolkit for progressive policymakers in developing countries*. London, UK: Overseas Development Institute.

Tam, P. T. K. (1977). The moderation of internal assessments for the award of grades in public examinations in Hong Kong. *Education Journal, 6*(1), 75–91.

Tan, K. H. K. (2011). Assessment for learning reform in Singapore: Quality, sustainable or threshold? In R. Berry & B. Adamson (Eds.), *Assessment reform in education: Policy and practice* (pp. 75–88). London, UK: Springer.

Tikly, L. (1999). Postcolonialism and comparative education. *International Review of Education, 45*(5-6), 603–621.

Trinidad and Tobago. Cabinet Appointed Committee to Consider Measures to Alleviate Problems of the Shift System at Schools. (1975). *Report*. Port of Spain, Trinidad: Ministry of Education. Chairman: Edward A. Moses.

Trinidad and Tobago. Committee Appointed by Cabinet to Consider the Feasibility of Converting Junior Secondary and Senior Comprehensive Schools into Full Time Five Year Schools. (1982). *Report*. Port of Spain: Trinidad: Ministry of Education. (Chairman: St. Clair King).

Trinidad and Tobago. Task Force for the Removal of the Common Entrance Examination. (1998). *Report.* Port of Spain: Ministry of Education.

Trinidad and Tobago. Government. (1928). *Administration report of the Director of Education for the year 1928.* Port of Spain, Trinidad: Author.

Trinidad and Tobago. Government. (1947). *Annual education report*. Port of Spain, Trinidad: Author.

Trinidad and Tobago. Government. (1950). *Annual education report*. Port of Spain, Trinidad: Author.

Trinidad and Tobago. Government. (1951). *Annual education report*. Port of Spain, Trinidad: Author.

Trinidad and Tobago. Government. (2011, June 30). Chaguanas Government Primary school tops SEA. *Trinidad and Tobago Government News*. Retrieved from http://www.news.gov.tt/index.php?news=8333

Trinidad and Tobago. House of Representatives. (2003). *House of Representatives Hansard* (debate of Friday, June 27, 2003). Port of Spain, Trinidad: Author.

Trinidad and Tobago. Ministry of Education. Examination Review Committee (1983a). *First report (June 1980); Second report (November 1982); Third report (December 1983)* (Historic Education Documents, Vol. 9). Port of Spain: Trinidad: Author.

Trinidad and Tobago. Ministry of Education. Examination Review Committee (1983b). *Fourth Report: Subject: The Common Entrance Examination* (Historic Education Documents, Vol. 15). Port of Spain: Trinidad: Author.

Trinidad and Tobago. National Task Force on Education. (1994). *Education policy paper (1993–2003) (White Paper)*. Port of Spain, Trinidad: Ministry of Education.

Trinidad and Tobago. Working Party on Education in Trinidad and Tobago. (1954). *Report.* Port of Spain, Trinidad: Government Printing Office. Chairman: Leslie R. Missen.

Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. New York, NY: Cambridge University Press.

Willingham, W. W. (1999). A systemic view of test fairness. In S. J. Messick (Ed.), *Assessment in higher education: Issues of access, quality, student development, and public policy* (pp. 213–242). Mahwah, NJ: Lawrence Erlbaum.

Willingham, W. W. (2002). Seeking fair alternatives in construct design. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 193–206). Mahwah, NJ: Lawrence Erlbaum.

Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Willingham, W. W., Cole, N. S., Lewis, C., & Leung, S. W. (1997). Test performance. In W. Willingham & N. S. Cole (Eds.), *Gender and fair assessment* (pp. 56–126). Mahwah, NJ: Lawrence Erlbaum.

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing, 27*(2), 147–170.