

Hong Kong secondary students' perspectives on selecting test difficulty level and learner washback: Effects of a graded approach to assessment

Language Testing

1–27

© The Author(s) 2021



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/02655322211050600

journals.sagepub.com/home/ltj**Chi Lai Tsang** University College London, UK
St. Joseph's College, Hong Kong**Talia Isaacs**

University College London, UK

Abstract

This sequential mixed-methods study investigates washback on learning in a high-stakes school exit examination by examining learner perceptions and reported behaviours in relation to learners' beliefs and language learning experience, the role of other stakeholders in the washback mechanism, and socio-educational forces. The focus is the graded approach of the Hong Kong Diploma of Secondary Education English Language Examination (HKDSE-English), incorporated in 2012, that allows test-takers to choose between easier and more difficult sections for reading and listening-integrated skills papers. Inductive coding of focus groups involving 12 secondary students fed into the development of the Washback on Students' Learning questionnaire, which was administered to another 150 learners. Exploratory factor analyses of identified washback effects revealed four major types straddling different settings (classrooms, tutorial schools, learners' personal environment), and seven categories of mediating variables pertaining to learners themselves, other stakeholders, and societal influences. Simultaneous multiple regressions identified influential clusters of mediating variables and showed the strongest predictors for each macro-level washback type varied. At least one intrinsic and one extrinsic factor category significantly contributed to all types, reaffirming learner washback as a socially situated, negotiated construct. Implications related to the consequences, use, and fairness of the graded approach are discussed.

Keywords

Factor analysis, impact, multiple regression, second language learners, test preparation, test-taker perceptions, washback

Corresponding author:

Talia Isaacs, UCL Centre for Applied Linguistics, UCL Institute of Education, University College London, 20 Bedford Way, London, WC1H 0AL, UK.

Email: talia.isaacs@ucl.ac.uk

Washback has been examined in language testing research for over a quarter of a century, owing, in part, to a growing awareness of the consequential effects of high-stakes testing for stakeholders, including in perpetuating or reinforcing social inequality (Shohamy, 2014). The notion of washback has evolved from being conceived of as a strand of validity (Messick, 1996), to an interface between testing, teaching, and learning (Alderson & Hamp-Lyons, 1996), and, more recently, to a socially situated construct nested within intricate webs of agents, contexts, systems, and power (Booth, 2012). Whereas washback studies initially focused on describing how testing influences aspects of teaching, washback on learning (i.e., the effects of testing on aspects of students' learning) has been a subject of growing attention (e.g., Green, 2006; Xie & Andrews, 2012), paralleling the shift in focus to learning processes in learning-oriented assessment (Turner & Purpura, 2016). However, most existing learner washback studies are constrained by a narrow focus on observable washback effects and/or a limited coverage of the discrete mediating factors within and immediately surrounding individual test-takers largely in isolation of the wider socio-educational and sociocultural context. Little is known about how mediating factors, both within and beyond individual learners as test-takers and the local-level micro contexts they engage in, work collectively to shape such effects. To continue to elucidate this complex and underexplored construct, research needs to, as Saville (2010) suggested, locate washback within the superordinate notion of impact (i.e., effects and consequences of testing throughout society). There is a need to examine how students' learning is mediated by agents and factors in micro contexts related to their personal sphere (e.g., home environment, classes, courses, school) and the overarching macro context (i.e., the social reality where learners and the test are situated), and to identify the consequences of such effects for learners.

The present study responds to this need by examining an under-researched, high-stakes test as a case study: the Hong Kong Diploma of Secondary Education English Language Examination (HKDSE-English). Since 2012, HKDSE-English has adopted a graded approach in the reading paper and the listening-integrated skills paper. This requires test-takers to choose between an easier and a more difficult section within each of these two papers. The study aims to uncover (1) the washback effects learners identify following the introduction of this graded approach, (2) the network of underlying personal, familial, institutional, systemic, and societal mediating factors shaping such effects, and (3) the categories of mediating factors that predict each identified type of washback effect. Washback research is sorely needed in the context of the HKDSE testing system, which selects candidates for tertiary education, owing to its gatekeeping function in shaping test-takers' educational futures and life trajectory (Smart et al., 2014), and in distributing power and establishing social order. We investigated the different actors and forces at play through the lens of the test-taker, probing their perceptions of the introduction of the graded approach in HKDSE-English.

Literature review

Empirical studies exploring the effects of testing on learners and their learning, which were once "peripheral to the design [of washback research]" (Green, 2006, p. 114), began to come to light in the field of language testing since the mid-2000s (Cheng et al.,

2015). The centrality of learners and their learning is arguably now anchored in mainstream thinking on washback in the field. An overview of this developing body of work in the language testing field reveals two major strands: observable washback effects on learning and mediating factors underlying learner washback. These are foundational to the present study and covered in the next sections.

Observable washback effects

Underpinned by findings demonstrating misalignment between teachers and students regarding what should be taught, learned, and practised in test preparation (e.g., Alderson & Hamp-Lyons, 1996), early learner washback studies established that in response to testing learners “do things they would not necessarily otherwise do” (Messick, 1996, p. 241). To elaborate, in formal instructional contexts, learners showed eagerness for test-related instructions and practices, especially upon the introduction of a new test (e.g., Stoneman, 2006) or major changes to a pre-existing test (e.g., Andrews et al., 2002). Marked by a preference of tested over untested constructs (Ferman, 2004) and a reluctance to learn the untested ones (Qi, 2004), learners favoured, demanded, and engaged in test-oriented teaching and practices built specifically around the target test (Tsagari, 2009). This was particularly evident when learners identified a mismatch between the teaching they desired and that they received (Cheng, 1998).

Learners’ test-focused preparation was mirrored in their out-of-class learning as the rehearsal of the test-taking experience and participation in private tutoring. The former was characterized by learners’ completion of self-guided and/or teacher-prescribed exam-related materials and sample test papers (Gosa, 2004; Xie & Andrews 2012). Often with the goal of test familiarization and practice, such materials were practised intensively. However, these materials selected by learners and/or teachers based on their perceptions of the test may in fact deviate from the actual test construct (Zhan & Wan, 2016), as evidenced by the prevalence of decontextualized discrete-point grammar drills in learners’ preparation for communicatively oriented tests (Pan, 2014; Qi, 2004). Private tutoring, including lecture-type test-specific preparation courses (Yung, 2015) and one-to-one, small group tutoring (Ferman, 2004) was undertaken by instrumentally motivated learners in hopes of being spoon-fed test-taking strategies and developing test-wiseness strategies (Allen, 2016b). Learners in these studies believed these were insufficiently taught in mainstream education and constituted a gap that shadow education (i.e., tutoring outside of formal schooling) could fill. In addition, non-test-focused L2 skill development strategies such as extensive reading, movie watching, and interactions with native speakers (Mickan & Motteram, 2009; Stoneman, 2006; Xie, 2013) were noted as less common practices among some self-motivated learners. In sum, learner washback was shown to be predominately test-oriented, penetrate facets of learning, and lead to a narrowing of curriculum and superficial learning (Andrews et al., 2002; Xie, 2013), especially when learners lacked personal agency (Mickan & Motteram, 2009) and changed only what but not how they learnt (Cheng, 1998; Qi, 2004).

Taken together, early studies established washback on learning as a construct that differs fundamentally from washback on teaching. This is owing, in part, to teachers’ and learners’ different perspectives, their relative power and agency, the locus of their

instructional activity (which in learners' case, may not solely be confined to the school classroom), and so forth. The primary focus of this body of research is on describing observable aspects of washback (i.e., learners' actions). However, the constructs underlying these phenomena and how washback effects differentially affect individual learners (Alderson & Wall, 1993) were not directly investigated.

Mediating factors

Prompted by research demonstrating the variability of washback across learners (e.g., Andrews et al., 2002; Gosa, 2004), subsequent studies have explored intrinsic (i.e., test-taker level) and extrinsic (i.e., beyond individual test-taker level) mediating factors underlying learners' observable actions. One classic intrinsic factor was language proficiency. Findings of studies across first language (e.g., Fox & Cheng, 2007) and second language contexts (e.g., Shih, 2007) confirmed that learners had marked differences in the type, amount, and intensity of their test preparation as a function of proficiency level. However, some studies identified a positive relationship between high proficiency level and washback intensity (e.g., Cheng et al., 2011; Pan, 2014), whereas others revealed the opposite trend (e.g., Fox & Cheng, 2007). Ferman (2004) noted that above-average-proficiency learners were the most devoted to test preparation, possibly because of their potential for upward socioeconomic mobility given the stakes involved. This suggests the importance of context-specific factors in interpreting findings across studies.

Other prominent intrinsic factors included learners' past test-taking experiences and perceptions of test design and use. The first related to learners' prior experience with test preparation methods, which they either rigidly adhered to (Pan, 2014; Stoneman, 2006; Zhan & Andrews, 2014) or strategically evaluated for usefulness and/or efficiency to inform future test preparation (Knoch et al., 2020; Sato, 2019). The effects of learners' perceived test design and use were explicated by Xie and Andrews (2012). Learners' positive endorsement of test design led to high evaluation of test importance, confidence towards test-taking, and intensive test preparation. Learners' intention of using the test for high-stakes purposes led to high value being attached to test-taking and ultimately test preparation. Xie (2015) subsequently added learners' favourable perceptions of test validity as a significant contributor to test preparation. Finally, Zhan and Wan (2016) elaborated on the consequences of learners' misinterpretation of the test construct, particularly underscoring forms of test preparation and test-taking behaviour unintended by the test developer. Together, these studies placed learners' perceived test design, value, use(s), and face validity amid major intrinsic mediating factors.

Compared with intrinsic factors, extrinsic factors have received little attention. Green's (2006) identification of a gradual convergence of learners' perceptions of course outcomes with their teachers' reported focus, and Zhan and Wan's (2016) documentation of teachers' tight control over learners, affirmed the pivotal role teachers played in shaping learner washback. Cheng et al. (2011) reported that parents' impressions of the test significantly predicted their support for their children's test-related learning. Parents' perceptions also directly related to children's perceptions of the test's impact on their motivation and L2 skill development. Delving into micro-contexts, Mickan and Motteram (2009) delineated how learners' life circumstances including living arrangements,

relationships, and work affected individual's preparation. Furthermore, Chik and Besser (2011) demonstrated that Hong Kong young learners were driven by schools, commercial language centres, and parents to undertake and drill international language tests because of a myriad of institution- and system-level extrinsic factors (e.g., language status, power and control in educational systems, institutional constraints, and parental and peer pressure). Their study has, therefore, brought mediating factors in the macro-context to the scene.

Despite this volume of research, reporting on both intrinsic and extrinsic factors within the same study is, as yet, scarce. Drawing on Dörnyei's (2005) motivational selves construct, Zhan and Andrews (2014) attributed differential washback to each test-taker's unique possible self. This possible self was comprised of their self-knowledge, beliefs regarding the exam, past learning and test-taking experience, others' test-taking experience, and their learning environment. Shih's (2007) case study unearthed more blocks of personal-, micro-, and macro-level mediating factors, including 12 test factors, three intrinsic factors, and four extrinsic factors. A number of these were mirrored in Booth's (2012) "test-taker, community & test complex" (p. 292), which posited that the factors residing in individual test-takers, the wider community and the test itself contributed to test-takers' actions and outcomes. Undertaking a more learner-oriented approach, several studies stressed learners' personal agency and strategic decision-making in the washback mechanism. Allen (2016a) ascribed learners' change in test preparation strategies across two consecutive IELTS tests to an array of factors based on interview data, including learners' perceptions of test difficulty, efficiency and effectiveness of test preparation methods, knowledge of how to improve, and assistance from others. Likewise, Sato's (2019) interview interpretation suggested that student views, influenced by examination, school, and examination-independent factors, shaped learners' test preparation methods the most. Lastly, Knoch et al. (2020) documented repeat test-takers' strategic transition in test preparation practices across multiple attempts at retaking the test. Their transition, from test familiarization and practice to test-wiseness to ultimately language learning, was found to be the result of their evaluation of score reports, perception of success with previous methods, and uptake of suggestions from peers, friends, and tutors.

In sum, these studies transcended descriptions of learners' observable reactions to testing by attempting to identify the constructs (i.e., mediating factors) underlying the differential washback across test-takers. However, this work does not sufficiently account for the complexity and social-situatedness of learner washback. Three gaps have yet to be addressed. First, the expanding repertoire of intrinsic and extrinsic mediating factors are categorized differently across researchers and contexts (e.g., Allen, 2016a; Shih, 2007; Zhan & Andrews, 2014). It would thus be useful to build upon these findings, elicit new ones, and observe whether the ways existing and new factors cluster in new contexts fit previous assumptions and categorizations. Second, with rare exceptions, studies either examine intrinsic factors independently in isolation of the wider socio-educational and sociocultural context (e.g., Fox & Cheng, 2007; Xie & Andrews, 2012), or provide limited coverage of intrinsic and extrinsic factors (e.g., Allen, 2016a; Sato, 2019; Knoch et al., 2020). Third, although some qualitative studies have examined the effects of selected mediating factors on learners' general washback

intensity (e.g., Sato, 2019) and selection of test-preparation strategies (Knoch et al., 2020), the question of which broad categories of intrinsic and extrinsic factors relate to each specific washback type and how they do so has not yet fully been answered. Studies adopting a more holistic approach and also incorporating an explanatory quantitative component could strengthen the existing evidence. Building on this preliminary work, we address these gaps using a socially situated and learner-oriented approach, which is grounded in contemporary socio-cognitive frameworks (e.g., Weir, 2005; Saville, 2010) within which learners' use of language, learning, test preparation, and the interpretation and use of their test score are all inherently social phenomena. This approach subsumes washback on learning within the superordinate construct of impact, and anticipates learner washback to straddle complex dynamic systems, sub-systems, and cultures, where the values and beliefs of the stakeholders involved come into play. Rooted in constructivism, which posits that reality is socially situated, the present study fits within the real world paradigm of research (Robson, 2002) and flexibly draws on both qualitative and quantitative sources of evidence. The overarching goal is to explicate how webs of intrinsic and extrinsic mediating factors interwoven within and beyond individual test-takers impact aspects of their second language learning and test-taking experience. We view this as a central contribution of our mixed methods study, which examines Hong Kong learners' reported responses to a novel test design feature in HKDSE-English, a test that has yet to be subject to rigorous external research and validation.

HKDSE-English and its graded approach

HKDSE is a battery of large-scale, high-stakes, criterion-referenced tests developed by the Hong Kong Examinations and Assessment Authority (HKEAA) in 2012 to replace two former senior-secondary public examinations (the Hong Kong Certificate of Education Examination and the Hong Kong Advanced Level Examination), following the 2009 reform that reduced the senior-secondary curriculum from four to three years. HKDSE-English, one of four core HKDSE subjects, is administered to 60,000 local secondary six (grade 12) students annually upon completion of their 12-year compulsory education (HKEAA, 2018). HKEAA's test specification for the general public frames HKDSE-English as an achievement test assessing students' performance in relation to communicatively oriented curricular targets (HKEAA, 2019). However, this is overshadowed by the test's power to select learners for post-secondary education, which is the major intended consequence of test use. To be admitted to undergraduate programmes offered by the eight local University-Grants-Committee-funded institutions or 280 overseas institutions recognizing HKDSE-English, a minimum of Level 3 on a five-level reporting scale, with Level 5 being the highest, is required. As for local sub-degree and higher diploma programmes, Level 2 or higher is expected. The consequences of test use, coupled with the examination-oriented culture in Hong Kong (Berry, 2011) and high status of English language (Poon, 2013), make HKDSE-English a powerful mechanism for maintaining the (im)balance of power in Hong Kong's educational system and in society. It is, thus, a fitting window through which the theoretical interests of this study could be explored.

HKDSE-English comprises four papers: reading, writing, listening-integrated skills, speaking, and a school-based component. In the reading paper and the listening-integrated skills paper, which together constitute half of the total score, a *graded approach* is used. The graded approach requires test-takers to choose which section to take within each of these two papers, either Part B1, the easier section, or Part B2, the more difficult section, after completing Part A. Students and all stakeholder groups are known to generally understand what it means to take Part B1 or Part B2, and test-takers generally approach the exam knowing which section they will select well ahead of time. Test-takers attempting Part B2 are scored using the full range of the five-level scale, whereas those attempting Part B1 have their marks capped at Level 4. That is, according to HKEAA's (2013b) benchmarking study, the highest attainable level for Part B1 is Level 4 (IELTS 6.90), and Level 5 (IELTS 7.64) for Part B2. The fact that HKEAA has kept all technical details in-house, including the theoretical and empirical basis of the pedagogical model upon which the test is based (i.e., the senior-secondary curriculum), the validity and reliability of the test design, and the exact procedure for item selection and development in Part A, Part B1, and Part B2, means that there is no official information in the public domain regarding the graded approach except its general scoring mechanism and objectives. According to Smart et al. (2014) from HKEAA, the graded approach adopts a nonequivalent groups with anchor test (NEAT) design, where Part B1 is linked to Part A (anchor) using data from candidates taking Part A and Part B1, and then Part A is linked to Part B2 using data from candidates taking Part A and Part B2. Thus, by means of equipercentile equating, during the scoring procedure scores on Part B1 are converted to the scale used for Part B2 using the scores on the compulsory Part A as a mediator. Each test-taker's mark of Part A is then added to their mark of Part B2 (for Part-B2-takers) or mark equivalent to Part B2 (for Part-B1-takers) to give the total score, which is ultimately converted back to a level via an expert judgement procedure. With this graded approach, HKEAA aims to "give candidates a choice of which optional part of the paper best matches their ability" and thus "efficiently test candidates with different abilities" (HKEAA, 2013a, p. 1). As Smart et al. (2014) noted, the intended washback is the "promotion of classroom activities and assessments which are less focused on replicating the public examination" (p. 269). It is common for local schools to prescribe senior-secondary classes either a Part-B1- or Part-B2-focused curriculum based on school banding or students' performance on the school-based secondary three (grade nine) school examination, or both. Unlike HKDSE-English, the secondary three school examination is at least partly textbook-based (with sections on textbook grammar and vocabulary) and happens three years prior to HKDSE-English. However, no studies have, as yet, examined the washback, impact, consequences, and fairness of the graded approach. The present study addresses these gaps through the following research questions.

1. Which washback effects do Hong Kong secondary school learners identify following the introduction of the graded approach in HKDSE-English?
2. Which intrinsic and extrinsic mediating factors shape learners' perceived washback effects?
3. Which categories of mediating factors predict each type of washback effect identified?

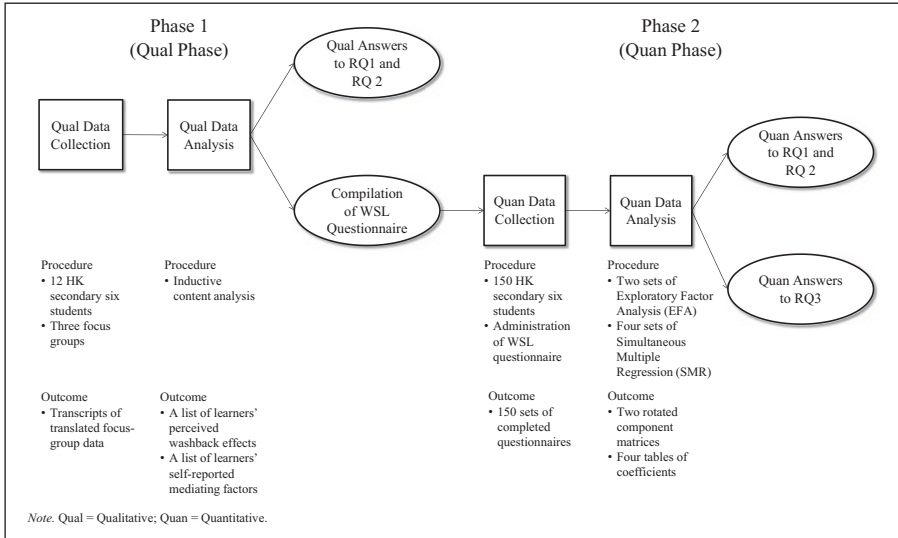


Figure 1. Research design.

Methods

Research design

The descriptive, exploratory, and explanatory nature of the research questions (1, 2, and 3 respectively) led to an exploratory sequential mixed methods research design (QUAL → QUAN). This design integrated the strengths of qualitative and quantitative approaches, particularly regarding the richness and generalizability of data (Creswell, 2015). The study, schematized in Figure 1, commenced with a qualitative phase, in which focus groups were conducted. We used focus groups because of their potential to capture rich information about washback effects and mediating factors through the student's lens (Patton, 2014), with the small group environment enabling them to build on each other's comments while sharing their individual experience. The resulting data, which addressed the first two research questions qualitatively, informed the items included in a questionnaire administered to another 150 learners. Finally, in the quantitative phase, sets of exploratory factor analysis (EFA) and simultaneous multiple regression (SMR) were performed on closed-ended questionnaire data to identify major types of washback effects, broad categories of mediating factors, and the mediating factors categories predicting each washback type.

Participants

Twelve Hong Kong secondary six students (9 male, 3 female; $M_{age} = 18.33$, $SD = 0.65$) participated in the qualitative phase. They were the part of the fifth cohort to be taking HKDSE-English since its inception in 2017 and had registered to sit the test three months

from the date of data collection. All participants' selection of which examination section (i.e., either Part B1 or Part B2) to take on test day was made prior to the test (and the focus groups) at various time points as early as the beginning of Secondary 4, when they were first prescribed a section-focused curriculum. Four (S5, S9, S10, and S12) had selected Part B1 and eight (S1–4, S6–8, and S11) had chosen Part B2. Maximum variation sampling was adopted to maximize the exploratory power. Participants were from three local government or government-aided secondary schools across three school districts. These three schools corresponded to the three academically streamed school bandings in Hong Kong (Bands 1, 2, and 3, with Band 1 having the highest academic abilities and Band 3 the lowest). S1 to S4 studied at a Band-1 English Medium Instruction (EMI) boys' school, S5 to S8 were from a Band-2 Chinese Medium Instruction (CMI) co-educational school, and S9 to S12 were from a Band-3 co-educational CMI school. Within each school, one top-scoring (90th percentile), one high-scoring (75th percentile), one average-scoring (50th percentile), and one low-scoring (25th percentile) student were selected to ensure the inclusion of participants at varying L2 proficiency levels, as informed by scores on in-house HKDSE-English mock examinations administered at each school.

The 150 (97 male, 53 female; $M_{age} = 18.13$, $SD = 0.47$) participants in the quantitative phase were secondary six students randomly selected from the same three schools using quota sampling: 50 from each of the Band 1, 2, and 3 schools. As in the previous phase, this stratified sample represented considerable variability in school bandings, academic abilities, and L2 proficiency.

Instruments

The development of the focus groups probes drew on Shih's (2007) overarching framework of washback effects as changes in content, time allocation, strategy use, motivation, and anxiety; and representation of mediating factors at personal, familial, institutional, and societal levels. This is the most extensive repertoire of washback effects and mediating factors identified in a similar context (L2 learners in Taiwan) and so guided the development of the focus group prompts in this study. After adapting the general washback effects to fit the study context and removing or replacing less relevant mediating factors (e.g., colleagues; girlfriends/boyfriends) with context-relevant ones (e.g., social media posts), the draft focus group prompts underwent two rounds of piloting with eight students from the Band-1 school. The resulting three-part bilingual (English-Chinese) discussion guide elicited learners' perceptions of the graded approach, any specific learning practices that they attributed to the approach, and their self-reported factors leading to such practices. After the main themes were introduced (e.g., perceptions of Part B1 and B2), probes were used to elicit learners' feelings, experience, behaviour, opinion, or values, making the focus group discussions open, adaptive, yet systematic (see Tsang, 2017, for the full research instrument).

The Washback on Students' Learning (WSL) questionnaire was developed by integrating insights generated from analyses of focus group data with two established questionnaires from studies that investigated learner washback in high-stakes testing. The first, Xie's (2013) Test Preparation Questionnaire, was developed to examine Chinese

College English Test candidates' perceived test-preparation practices. The second, Purpura's (1999) Cognitive and Metacognitive Strategy Questionnaire, was developed to elicit European First Certificate in English test-takers' self-reported cognitive and metacognitive strategy use. We incorporated items from these instruments that applied to and matched the perceived washback effects that our focus group participants had identified, making minor contextual adaptations where necessary (e.g., *I drilled on my reading comprehension skills* [Xie, 2013] was recontextualized as *I drill only on the skills specific to my preferred section [e.g., Reading Part B2 – referencing, inferencing questions and long questions]*). Part A of the WSL questionnaire related to washback effects on learning, whereas Part B focused more specifically on identifying potential influences and mediating factors on learners' behaviour. The WSL questionnaire was then translated and presented bilingually to participants (English-Chinese) on a printed page to enhance its accessibility, and instructions were added at the beginning of each section.

The WSL questionnaire was then refined through two rounds of piloting with learners from the Band-1 school using think-aloud protocols ($n = 4$) and post-questionnaire completion debriefs ($n = 30$) prior to its implementation. The revisions, including rewording ambiguous items, removing the neutral response choice, and paraphrasing syntactically similar items to avoid biased clustering in statistical analyses, led to a two-part bilingual (English-Chinese) WSL questionnaire, available open access on the IRIS digital repository (Tsang, 2020). Part A (15 items; Cronbach's $\alpha = 0.72$) constituted direct declarative statements covering aspects of attitude, motivation, and behaviour regarding students' preparatory work for the graded approach, which they rated on a four-point Likert scale (*strongly disagree/strongly agree*). Part B (30 items; Cronbach's $\alpha = 0.86$) probed potential influences on students' views or behaviours, as rated on a four-point Likert-type scale (*not influential/very influential*).

Data analysis

The focus group data were analysed inductively in an iterative process. Following orthographic transcription, the first author translated the learners' responses from Cantonese into English, the accuracy of which was confirmed by a third independent Cantonese-English bilingual researcher. Cantonese was the primary language participants used, with occasional Cantonese-English code-switching involving common adjectives, nouns, and formulaic phrases. The first author then identified the following broad emergent categories: washback in classrooms, washback in personal environment, washback across venues, intrinsic factors, and extrinsic factors. These initial categories were deconstructed after preliminary coding and multiple rounds of discussion with the second author. For example, washback in personal environment was differentiated into formal and informal ways of learning, and extrinsic factors were further categorized into personal, familial, school, and societal levels. Once these narrower categories had been established, the first author and the third independent researcher independently (re)coded the entire corpus. The agreement level was 80.9% (161/199 codes), with differences in opinion, mostly on the personal-level intrinsic mediating factors (e.g., exam knowledge versus perceptions of the examination sections), resolved through discussion. Participants'

verbatim comments were then mapped onto the resulting lists of perceived washback effects and mediating factors.

The questionnaire data were analysed using SPSS 24.0. Preliminary analyses of Part A and Part B data ensured sampling adequacy (Kaiser-Meyer-Olkin values $> .65$; Bartlett's test of sphericity $< .05$) and ruled out multicollinearity (determinant of item correlation matrices $> .00001$). Sets of EFA were then conducted using principal axis factoring with Promax rotation to investigate the number and types/categories of washback effects and mediating factors, particularly those brought about by the graded approach. Items that cross-loaded (i.e., loading difference between the primary and alternative factors $< .2$) and/or had a loading below a conservative $.4$ (Howard, 2016) were dropped. Factors with two items were retained only when the items were highly correlated (i.e., $> .7$; Worthington & Whittaker, 2006). Assumptions for the follow-up SMR (e.g., linearity, independent errors, homoscedasticity, no multicollinearity, normality of errors) were met. Therefore, sets of SMR were performed to investigate the categories of mediating factors that predict each type of washback effect. Simultaneous entry was adopted as a result of little prior knowledge regarding the effects of the categories of mediating factors on the types of washback effects identified in this study. The alpha level was set at $.05$ for all statistical tests.

Results

Learners' perceptions of washback effects brought about by the graded approach

To address research question one on the washback effects learners identify following the introduction of the graded approach, EFA was performed on learners' perceptions of their attitude, motivation, and behaviour regarding their preparatory work for the graded approach. The point of inflection in the scree plot was at five factors. Four factors had eigenvalues over Kaiser's (1974) criterion of 1, collectively explaining 66.67% of the variance. Thus, four factors were extracted.

Table 1 shows the factor loadings for predicting observed variables (i.e., 13 Part A questionnaire items) from the four underlying factors, which we interpreted as four types of learners' perceived washback effects, in addition to eigenvalues, proportion of variance explained, and reliability of each factor. Factor 1 represents informal ways of training for the preferred HKDSE-English section outside the classroom. Factor 2 represents selective attention in English language learning (i.e., some areas focused on, others ignored). Factor 3 represents intensive paper-and-pencil drills on the preferred HKDSE-English section. Finally, factor 4 represents enrolment in private section-focused HKDSE-English tutorial classes. Note that readers may view the item correlation matrices of Part A and Part B in the online supplementary file that appears next to this article on the *Language Testing* website. Likewise, the correlation matrix of the four washback types and that of the eight mediating factor categories appears in the online supplementary file.

The first type of learners' perceived washback effects was informal ways of training for the preferred HKDSE-English section outside the classroom. The items with the largest factor loadings related to deliberate exposure to media or books and memorizing

Table 1. Four-factor rotated component matrix for WSL questionnaire part A: Perceived washback effects.

Item	Factor			
	1	2	3	4
Reading newspaper articles relevant to the preferred section	.78			
Memorizing words related to the preferred section	.77			
Reading books at a difficulty level approximating the preferred section	.72			
Listening to radio programmes at a level similar to the preferred section	.67			
Neglecting language skills specific to the dispreferred section		.93		
Reluctance to drill the dispreferred section		.69		
Reluctance to learn about the dispreferred section		.47		
Eagerness towards learning language skills specific to the preferred section		.42		
Drilling the preferred section of mock papers in school lessons			.91	
Drilling the preferred section of past papers in school lessons			.75	
Completing out-of-class exercises beneficial to the preferred section			.43	
Continuous enrolment in private section-focused tutorial class				.84
selecting private tutorial class targeting at the preferred section				.84
Eigenvalue	3.06	2.30	1.88	1.43
Variance explained (%)	23.50	17.72	14.43	11.02
Cronbach's α	.82	.72	.71	.83

Note: Factor loadings < .3 not shown.

lexical items related to the preferred section in terms of content or difficulty level. Learners' comments corroborated this.

"I read SCMP [South China Morning Post] to look for potential B2 topics." (S3)

"I listen to RTHK [Radio Television Hong Kong] because their broadcasts are appropriately difficult, and B2 topics are always social issues." (S7)

"I read novels at a B2 level of difficulty. . . I time myself to read faster." (S11)

"I walk down the streets scanning for difficult B2 words. . . They are everywhere. . . I try to memorize all." (S11)

The examples featured in these quotes suggest that reading and listening outside the classroom were not for leisure but for section-focused test preparation, with all quotes explicitly referring to the target proficiency level. Printed and aural materials were utilized selectively, purposefully, and strategically as informal means to achieve aims such as increasing reading speed, predicting exam topics, or identifying and memorizing words relevant to the preferred section.

The second type of learners' perceived washback effects was selective attention in English language learning. In general, learners embraced section-focused teaching and some viewed this positively:

"I love our teacher teaching us B2 skills such as writing with appropriate tone and register." (S4)

". . . we value our teacher's step-by step guide to answering B1-type reading questions – multiple choice, ordering, and those [item types] whose answer can be copied directly." (S9)

In contrast, with regards to the dispreferred HKDSE-English section, learners expressed reluctance to learn or engage with material relevant to a different level:

". . . our teacher once introduced B2 . . . nobody listened." (S12)

"Learning, not least practising, B1 is a waste of time." (S2)

"B1 is all about understanding stated information, which is irrelevant in our B2." (S4)

"We forgo B2 long questions because they aren't in B1 . . . we lack the ability to construct such sophisticated responses." (S10)

The third type of learners' perceived washback effects was intensive paper-and-pencil drills and teaching to the test activities on the preferred HKDSE-English section. Within classrooms, drilling the preferred section of past and mock HKDSE-English papers was reportedly "an everyday ritual throughout the senior-secondary years" (S9). Learners noted that English lessons typically followed the standard structure of paper-and-pencil B1/B2 drills, answer checking, and evaluation, and some expressed eagerly awaiting teacher's feedback on the preferred section:

"B2 marking requires professional judgement because many questions are subjective . . . We are dying for teachers' feedback." (S2)

Completing homework and optional out-of-class exercises assigned by the teacher relating to the preferred section was also common:

"At home I spend hours completing additional B1 exercises from my teacher." (S10)

These quotes illustrate that exercises and activities directly targeting the preferred section tended to penetrate classrooms, infiltrate learners' personal environment, and shape their L2 learning experience.

Finally, the fourth type was enrolment in private tutorial schools' section-focused HKDSE-English classes targeting the preferred section. These commercialized private classes were in addition to section-focused school teaching, with one learner describing them as a "necessity" (S3) and others similarly underscoring their importance:

“. . . courses in tutorial schools are classified into elite and standard streams with differing content, focus, and length. . . our desire to succeed in B2 drives us towards the elite stream.” (S4)

“I will never drop a tutorial course when it teaches me useful B1 skills.” (S12)

Learners' self-reported mediating factors

A separate EFA was conducted to address research question two on the mediating factors that learners perceive to have shaped the washback effects they identified (i.e., reported influences on their attitudes and behaviour). The scree plot levelled off at eight factors and seven had eigenvalues over 1, together accounting for 68.62% of the variance. Therefore, seven factors were retained. Table 2 shows the factor loadings for predicting observed variables (i.e., 24 Part B questionnaire items) from the seven underlying factors, which we interpreted as seven categories of learners' self-reported mediating factors, alongside eigenvalues, proportion of variance explained, and reliability of each factor. Factor 1 represents teachers' evaluations. Factor 2 represents experience at private tutorial schools. Factor 3 represents language proficiency. Factor 4 represents peer influence. Factor 5 represents influence of personal contacts. Factor 6 represents socio-educational forces. Factor 7 represents exam knowledge.

The first category of learners' self-reported mediating factors, under which six mediating factors were grouped, was teachers' evaluations. To begin with, teachers' selection of course focus, which was essentially a class-level evaluation made with reference to the school banding and reflected in materials used in class, influenced learners' attitudes and test preparation:

“My teacher prefers teaching us B2. Therefore, we are brainwashed to practise just that.” (S3)

“In a Band-3 school, the idea taught and shared among us is to drill B1 and secure a pass.” (S10)

“. . . we are conditioned to practise B1, as everything our teacher exposes us to is B1.” (S9)

These school- and class-level evaluations discouraged and limited learners' access to the untaught section. This was particularly evident at Band-2 and Band-3 schools, where some learners believed that, regardless of their abilities, they were being taught and pushed towards one HKDSE-English section, which benefited their teacher (and school) the most. B2 tended to be targeted at Band-1 schools so that “more scores in the upper range (i.e., Level 4 and above) can be recorded” (S5), whereas B1 tended to be targeted for Band-2 and Band-3 schools so that “a higher overall pass rate can be achieved” (S11). Thus, school streaming was perceived to dictate B1 or B2 level selection. Furthermore, teacher's personal-level evaluations, including their assessment of individual students' English proficiency, expectations, advice, and comparisons with other students, also played a role:

“My teacher affirms my English is not up to B2 level, so I practise only B1.” (S10)

“I practise B2 in-and-out of classrooms, as I don't want to disappoint my teacher.” (S1)

Table 2. Seven-factor rotated component matrix for WSL questionnaire, part B: Self-reported mediating factors.

Item	Factor						
	1	2	3	4	5	6	7
Teachers' assessment of English proficiency	.78						
Teachers' advice	.74						
Teachers' expectation	.71						
Teachers' selection of course focus	.65						
Comparisons made by the teacher	.54						
School banding	.47						
Tutorial school's advice		.93					
Test-wiseness strategies and language skills learnt in tutorial class		.84					
Advertisements from tutorial schools		.56					
Performance on school quizzes			.80				
Performance on school mock exam			.72				
Self-perceived English proficiency			.58				
Priority of English in relation to other subjects			.43				
Classmates' preparatory work				.88			
Classmates' selection of exam section				.74			
Classmates' performance on school exams				.70			
Siblings' advice					.88		
Parents' expectations					.78		
Posts on social media and online forums					.42		
Importance of HKDSE-English in HK						.93	
Examination-oriented culture in HK						.81	
Ability to resist external influences						.42	
Perceptions of the two examination sections							.82
Knowledge about the examination							.78
Eigenvalue	5.97	3.00	2.05	1.82	1.40	1.17	1.05
Variance explained (%)	24.85	12.51	8.55	7.58	5.85	4.90	4.39
Cronbach's α	.82	.81	.74	.82	.79	.71	.83

Note: Factor loadings < .3 not shown.

"Sometimes my teacher says some of us are B2-ready. . . implying others are not so they should practise B2 harder, which I am doing now." (S7)

Teacher-, school-, class-, and personal-level evaluations thus constituted a powerful category mediating the type, nature, and intensity of learner washback.

The second category highlighted advice, teaching of test-wiseness strategies and selected language skills, and advertising/marketing from private HKDSE-English

tutorial schools where the learners were registered, as being influential in their test preparation:

“Tutorial school advertisements are everywhere. We are told level 5 is possible only if we take their B2-focused classes.” (S2)

“Tutorial schools show how easy B2 is if we master the right skills. . . so eventually we move away from B1.” (S10)

The next category was language proficiency, which comprised four mediating factors. Among these, performance on school mock examinations or in-class quizzes in some cases spurred learners into action:

“Unsatisfactory mock exam result is the final warning that I should take B2 courses.” (S11)

“Scoring well in quizzes consistently affirms I should go above B1 and practise B2.” (S11)

Continuous assessment led some learners to develop a firm view about their own language proficiency:

“As an EMI student who knows English well, I despise B1.” (S1)

It was also instrumental in determining how much to prioritize English in relation to other subjects:

“English is my strongest subject which I must excel in. . . so I practise B2 intensively.” (S1)

The fourth category was peer influence, with classmates’ examination section selection, preparatory work, and school examinations performance affecting the nature and intensity of learners’ own test preparation.

“When all of my classmates select B1, I don’t dare to touch B2.” (S10)

“I print whatever my classmates print. I attend whatever classes my classmates attend.” (S6)

“When a weaker classmate scores higher than me in school exams, I panic and drill harder.” (S4)

The fifth category was influence of personal contacts, agents surrounding learners’ immediate learning contexts who exert a direct or indirect influence on their learning. The following quotes illustrate how parents’ expectations, older siblings’ advice on HKDSE-English test preparation, and other test-takers’ posts on social media and online forum shaped learners’ washback:

“My parents expect me to pass. . . so I practise B1 to get a safe pass.” (S9)

“My brother high-achieving advises me to take B2 tutorial classes like he did. . . I follow suit to keep up with or even surpass him.” (S4)

“The popular Facebook page ‘Secrets of Prestigious Schools’ is flooded with posts about how Band-1 students prepare frantically for B2. . . this suggests B2 is for Band 1 but not 2 or 3.” (S3)

The sixth category was socio-educational forces, which encompassed broader mediating factors that underscored the importance of HKDSE-English against the backdrop of Hong Kong’s examination-oriented culture.

“Level 5 is useful in Hong Kong in whatever discipline so I practise B2 hard.” (S7)

“. . . everyone who wants a successful career must drill B2 so they excel in HKDSE-English, which leads them to university and then a promising career.” (S3)

These characterizations of the consequences of taking the different sections appeared to prompt learners to learn and practise selectively. Finally, learners’ (in)ability to resist external influences and go with the majority view also appeared to play a part, as reflected by the comment, “those who lack confidence follow the mainstream.” (S6)

Exam knowledge, the last category, included two mediating factors: perception of the examination sections and knowledge about the examination. The former concerned learners’ beliefs regarding B1 and B2; the latter concerned how much they knew about HKDSE-English, particularly regarding its capping policy.

“B2 is for high-achievers, whereas B1 is for those who struggle with English, lack exposure, and hope for a pass. Therefore, I haven’t touched B1.” (S1)

“Full marks in B1 gives only level 4 so practising B2 is essential.” (S10)

Predictors of each washback type

Based on the EFA results that, together, had identified four types of washback effects and seven categories of mediating factors, we conducted four sets of SMR to address research question three. In each set, the same seven broad categories of learners’ self-reported mediating factors – teachers’ evaluations, experience at private tutorial schools, language proficiency, peer influence, influence of personal contacts, socio-educational forces, and exam knowledge – were computed as variables predicting each of our four major types of learners’ perceived washback (see Table 3). SMR Set 1 was performed to predict learners’ informal ways of training for the preferred HKDSE-English section outside the classroom. Influence of personal contacts and language proficiency were significant predictors. SMR Set 2 was conducted to predict learners’ selective attention in English language learning. Exam knowledge, language proficiency, and teachers’ evaluations had significant effects. SMR Set 3 was run to predict learners’ intensive paper-and-pencil drills on the preferred HKDSE-English section. The significant predictors were exam knowledge, teachers’ evaluations, and peer influence. Finally, SMR Set 4 was computed to predict learners’ enrolment in private section-focused HKDSE-English tutorial classes. Exam knowledge, experience at private tutorial schools, and peer influence were significant predictors.

Table 3. Results of SMR sets.

	Sets							
	1		2		3		4	
	β	p	β	p	β	p	β	p
Teachers' evaluations	.087	.327	.200*	.026	.199*	.025	-.124	.143
Experience at private tutorial schools	-.061	.524	-.067	.489	.141	.143	.200*	.030
Language proficiency	-.195*	.033	-.237*	.010	-.025	.781	-.070	.418
Peer influence	.154	.103	.063	.509	-.195*	.040	.194*	.032
Influence of personal contacts	.205*	.045	-.033	.743	-.123	.224	-.020	.835
Socio-educational forces	-.037	.693	-.024	.800	.001	.994	.060	.504
Exam knowledge	.116	.180	.243*	.006	.200*	.022	.266*	.001

Note: Adjusted R^2 : Set 1 = .066 ($p = .018$); Set 2 = .132 ($p = .029$); Set 3 = .072 ($p = .013$); Set 4 = .155 ($p = .000$). β values are asterisked where $p \leq .05$.

Discussion

This mixed methods study has positioned washback within the broader notion of impact, operationalizing washback on learning as a socially situated construct to examine stakeholder, classroom, and societal repercussions of the implementation of HKDSE-English's graded approach. We investigated learners' perceived washback effects, self-reported mediating factors, and the predictors of each identified washback type. First, we found that the 13 learners' perceived washback effects fall under four macro-level categories: informal ways of training for the preferred HKDSE-English section outside the classroom, selective attention in English language learning, intensive paper-and-pencil drills on the preferred HKDSE-English section, and enrolment in private section-focused HKDSE-English tutorial classes. The last three of these effects corroborate findings from previous studies. For instance, selective attention and drills were noted in Qi (2004) and Tsagari (2009), and participation in private tutoring was identified in Allen (2016b) and Yung (2015). By establishing and documenting their presence in a new testing context, the current study reaffirms them as key indicators of learner washback. We demonstrated how in Hong Kong, HKDSE-English test-takers' preference of tested over untested constructs was (re)contextualized as examination-section-oriented (i.e., either Part B1 or Part B2) learning and drills. We also further delineated the adaptive nature of Hong Kong's shadow education, which is characterized by streamed section-focused, teacher-fronted, test preparation courses within which learners engage in response to changes to a high-stakes examination. The remaining effect, informal ways of training for the preferred section outside class, is, however, far less elaborated in the literature. Despite cursory mention in a handful of studies (e.g., Cheng, 1998; Mickan & Motteram, 2009; Pan, 2014), detailed accounts of why these less direct forms of test preparation count as evidence of learner washback are scarce. This may be because this washback type lies outside the realm of formal instruction and, hence, may be integrated with and appear as everyday activities in learners' personal environment (e.g., reading books, listening to

podcasts). This obscures the identification of such washback effects, whose link to high-stakes testing is less than apparent on the surface. The present study, thus, advances knowledge by enlisting the washback effects within this under-researched type, bringing these previously discrete, isolated effects into one coherent, overarching type alongside more direct washback effects (e.g., intensive drills). We discussed the motives for such informal washback effects, illustrating how they may, in contrary to previous conceptions of them as general language enhancement strategies (e.g., Stoneman, 2006; Xie, 2013), in fact be strategic, test/purpose-driven ways of learning in learners' personal environment in response to high-stakes testing.

Viewed more holistically, these four washback effect types demonstrate a progressive transmission of washback effects across settings: from classrooms, to tutorial schools, to learners' personal environment. This finding aligns with the current understanding of washback on learning (Cheng et al., 2015), which posits that learner washback penetrates contexts and facets of learning, including the type(s) of test-preparation practices learners engage in and their focus and attention.

The second major finding relates to the 24 learners' reported mediating factors, which fall into seven broad categories (2 intrinsic, 5 extrinsic) and generally substantiate Shih's (2007) categorization. The two intrinsic factor categories, language proficiency and exam knowledge, are well-established in the washback literature (e.g., Fox & Cheng, 2007; Pan, 2014), with the current study's findings explicating and confirming the constituents of these two categories. In particular, we illustrated how each of these intrinsic factor categories has an objective component concerning facts and knowledge (e.g., performance on school exams; knowledge about the examination), and a subjective component rested upon learners' perceptions and beliefs (e.g., self-perceived language; perception of examination sections), thereby grounding these intrinsic mediating factor categories in learners' thinking and the information available to them. The five categories of extrinsic mediating factors are teachers' evaluations, experience at private tutorial schools, peer influence, influence of personal contacts, and socio-educational forces. This shows a wide range of extrinsic factors spanning school, familial, institutional, and societal levels, echoing previous washback studies (e.g., Allen, 2016a; Sato, 2019; Shih, 2007). Some factors such as teachers' advice and parents' expectations have been discussed in previous work (e.g., Cheng et al., 2011; Green, 2006; Zhan & Wan, 2016). However, other extrinsic factors, and particularly those at the broader societal level (e.g., posts on social media and online forums, examination-oriented culture) have received relatively less attention. One possible explanation is that socio-educational and sociocultural factors are not the main focus of earlier washback studies, which confine the locus of washback to instructional settings only. Another possibility is that some studies attributed such factors to individual test-takers and the agents surrounding them rather than being drivers of the system that overarch individuals' actions and perceptions. For example, Zhan and Andrews (2014) characterized examination-oriented culture as learners' urge to address their learning weaknesses, their rich past test-taking experiences, and commercial publishers' extensive test preparation materials, instead of a systemic-level societal factor that is itself also a washback effect (see also Cheng, 1998; Stoneman, 2006). Therefore, by further elaborating both known and less-known categories of extrinsic factors (e.g., comparisons made by the teacher, advertisements from tutorial

schools, siblings' advice), the present study elucidates the scope and categorization of extrinsic mediating factors. Furthermore, the nature of these factors suggests that extrinsic factors must encompass not only human agents in learners' immediate formal learning environment, but also the wider social realities in which they are situated.

The last group of findings addresses the research gap of identifying the predictors of each washback type using clusters of both intrinsic and extrinsic mediating variables. Results of the four SMR sets reveal patterns of relationships corroborating and extending previous research findings. For instance, two mediating factor categories, influence of personal contacts and language proficiency, significantly predicted informal ways of training for the preferred HKDSE-English section outside the classroom. However, only language proficiency, alongside interest towards the language, has been directly related to this washback type in earlier work (e.g., Allen, 2016a; Pan, 2014; Sato, 2019). By way of another example, exam knowledge and teachers' evaluations predicted intensive paper-and-pencil drills on the preferred HKDSE-English section, confirming previous research findings (e.g., Green, 2006; Qi, 2004; Xie & Andrews, 2012; Zhan & Wan, 2016). However, in the current study, peer influence also predicted section-focused drills. This extrinsic mediating factor category is commonly referred to in literature (e.g., Allen, 2016a; Shih, 2007) but has seldom been ascribed to this particular washback type. Lastly, enrolment in private section-focused HKDSE-English tutorial classes is the only washback type predicted by mediating factor categories (i.e., exam knowledge, experience at private tutorial schools, peer influence) whose link to private tutoring has been documented (Allen, 2016b, Yung, 2015). In sum, the SMR sets complement existing research in confirming, identifying, and explaining the sources of influences that shape each of learners' specific washback types.

Having conducted these four SMR sets and viewed them holistically, we were able to identify influential mediating factor categories that significantly relate to multiple washback types. This included exam knowledge, language proficiency, teachers' evaluations, and peer influence. For example, exam knowledge significantly predicted three out of four washback types. This suggests that both learners' perceptions of the stakes and consequences of examination section selection, and their analysis of the language skills and knowledge assessed in their preferred and dispreferred section, guided their HKDSE-English test preparation. Through these analyses, we were able to extrapolate and reaffirm three properties of the washback on learning construct, which have been noted in previous studies. First, given that each washback type was significantly predicted by at least one intrinsic and one extrinsic factor, washback on learning is driven by an array of intertwining forces within and beyond learners' locus of control (e.g., Cheng et al., 2015). Second, because categories of intrinsic and extrinsic factors underlie every washback type and mediate the nature and intensity of washback, washback is the product of learners' strategic negotiation between sometimes conflicting mediating factors (e.g., Allen, 2016a; Sato, 2019). Further, the differential washback across learners suggests that learners themselves play a pivotal role in determining the influence of each mediating factor, either by prioritizing or downplaying it, with the ultimate aim of maximizing the likelihood of achieving their desired outcome as a result of test performance. Thus, washback on learning is essentially a dynamic negotiation performed by learners between competing factors (e.g., Knoch et al., 2020) that vary in perspective, intention(s), and

strength. Third, the agents in the context where the consequent washback effect occurs (e.g., teachers, peers) often have the power to influence learner washback. Therefore, one way to manipulate learner washback could be to alter the interaction between learners and these agents.

Implications

This study has implications for both stakeholders embedded within the Hong Kong educational system and society at large. The result that learners reportedly focus only on their preferred section in test preparation reaffirms that they undertake narrow test preparation according to their or other agents' interpretations of the test construct (Green, 2006; Xie & Andrews, 2012; Zhan & Wan, 2016). Regardless of the instructional focus, test-takers' selective attention results in a narrowing of curriculum, which is common when a high-stakes test is administered upon the completion of a formal curriculum (e.g., Cheng, 1998; Gosa, 2004; Qi, 2004). However, unlike most other testing systems, HKDSE-English's graded approach inherently streams learners' language proficiency assessment and/or achievement of the curriculum into two distinct pre-defined sections (i.e., Part B1 and Part B2) within one examination. Consequently, as our data show, test-takers' learning is geared toward a particular section of the test. This is concerning in the case of HKDSE-English's graded approach because HKEAA does not appear to have a theoretical or an empirical basis to support claims about curriculum sequencing and test section difficulty. For example, the senior-secondary curriculum does not appear to have been informed by research or theory on order or sequence of acquisition, nor benchmarked to a common set of standards that chart incremental increases in learner ability and task difficulty (e.g., Common European Framework of Reference for Languages; Council of Europe, 2001). There is also no available information in the public domain about the extent to which the two examination sections sample the relevant curricular content. It is also unclear how much, or even if, the section-focused learning and teaching align with HKEAA's intended Part B1 and Part B2 test construct. In the context of this dearth of substantive backing for curricular and test sequencing and design features, it is difficult for students to be able to envisage and for teachers to advise what language skills ought to be developed and focused on in an informed way. For example, if HKEAA's model of curriculum sequencing or theory of content difficulty is not empirically or theoretically supported and the presumably more difficult Part B2 turns out to cover mostly lower-order language skills, then suggesting that Part-B2-takers study purportedly more difficult material would be poor advice. Furthermore, even if the assumptions of sound theory of content difficulty and high content validity are met, students' section-focused learning could still undermine the senior-secondary English curriculum in several ways. For instance, students choosing B2 might overestimate their ability to do well on the test and overlook fundamental language skills or content that they disregard but have not yet mastered (Trofimovich et al., 2016). In this respect, our data echo previous studies suggesting that in testing systems offering tests at different L2 proficiency levels (e.g., Cambridge English qualifications), test-takers may be intrinsically and/or extrinsically motivated to select more advanced levels (Chik & Besser, 2011). Preparing for a test beyond their ability may also lead learners to perform more poorly (Gu &

Saville, 2016). Conversely, students choosing B1 might have a pre-set internal ceiling and filter out any forms of learning that they consider out-of-syllabus (i.e., beyond B1-level), even those that are meaningful and are pitched at their actual ability level. In HKDSE-English, their scores are capped due to their taking the lower-level section, which is a practical explanation for their reluctance to learn beyond the tested knowledge. The detrimental effects of the resulting narrowing of the curriculum may extend beyond students' language learning to their wellbeing, as studies have found narrow test preparation to be a reason for heightened test anxiety and even mental breakdown (e.g., Fox & Cheng, 2007; Tzagari, 2009). Such effects would work against HKEAA's (2013a, 2019) intended assessment and curricular objectives.

Another finding relevant to Hong Kong's Education Bureau, HKEAA, and school personnel is that HKDSE-English's graded approach has differential impact on groups. While a high-stakes test should empower all test-takers (Shohamy, 2006), the graded approach attributes unequal power to the two sections through the capping policy. Learners' impressions and the school-level mediating factors to which they are exposed (e.g., teachers' selection of course focus, school banding) suggest that the power and control embedded in test design may have motivated schools to prescribe to their students' section-oriented teaching. This selection is likely structured around the section most beneficial to the schools themselves for accountability purposes. This could lead to B1-focused teaching at academically weaker Band-2 and Band-3 schools to boost pass rates, and B2-focused teaching at high-achieving Band-1 schools to maximize the number of Level 5 learners, thereby perpetuating cycles of high school achievement. In other words, based on the school banding hierarchy, students, and particularly those in lower bandings, who might not be able to accurately self-assess (Trofimovich et al., 2016), may be coerced into taking a particular section. Hence, learners' autonomy in deciding which section to select and what form(s) of test preparation to take may be appropriated by their school and teachers, making learners' free will and empowerment in this decision-making an illusion. This runs the risk of leading to instances of learners' underachievement especially among those capable Band-2 and Band-3 learners who are beyond Level 4 but are designated the self-limiting B1, thus reinforcing the power imbalance of Hong Kong's academically streamed educational system. Furthermore, given that the nature, quality, and quantity of surrounding extrinsic mediating factors vary markedly across individual learners and that learners from disadvantaged backgrounds might lack access to certain resources (e.g., input from tutorial schools; advice from educated parents/siblings), the graded approach potentially perpetuates social inequalities and limits opportunities for young people.

Given the stakes of HKDSE-English and negative systemic effects of the graded approach, we argue for the adoption of fairer, more scientific ways to inform test-takers' decision-making that take into account individual differences and seek to empower (not subjugate) test-takers' voices. For example, each year before HKDSE-English is administered, at least one free standardized territory-wide proficiency/achievement/diagnostic test, which appraises learners of their current and projected attainment level on HKDSE-English's five-level scale and area(s) of improvement, should be incorporated into the senior-secondary curriculum. In this regard, an extension of HKEAA's existing Territory-wide System Assessment, a standardized four-skill achievement test administered

annually to all primary three, primary six and secondary three students, to secondary four/five/six levels using a range of HKDSE-English item types could address the need for a free standardized four-skill test in students' senior-secondary years, on the condition that results of the test are not to be used beyond its intended low-stakes purposes. Results of the test would reduce learners' reliance on the judgements of their school and teachers to inform more fairly how many B1- and B2-oriented classes should be offered at each school and/or whether individual learners should receive B1- or B2-oriented teaching irrespective of school banding. Next, greater transparency and rigour in test development and validation could help educational stakeholders better understand the constructs being measured, how test content aligns with curricular content, how claims about test difficulty levels are buttressed, and so forth, to guide more informed real-world decision making. Finally, better construct representation could attenuate the effects of a narrowed curriculum (e.g., the inclusion of more inferencing and summarizing questions pitched at the B1-level in B1 reading). These suggestions could, in a small part, help make HKDSE-English a better engine for generating beneficial learner washback, which, in turn, could result in a fairer mechanism for shaping learners' educational futures and life trajectory.

Limitations and future research

The present study is subject to several limitations, a few of which we acknowledge here. First, our findings are constrained by the items drawn from Shih's (2007) framework, which we adapted for the Hong Kong context. However, the use of a pre-existing instrument shaped what we asked participants and the resulting data, which constrained the variables that we ultimately examined. Second, the fact that HKDSE-English's graded approach is embedded in a large-scale, high-stakes test means that some of the insights drawn in the present study might have more to do with the power of the test than the actual design of the graded approach. This is likely given that the relationships between test design features, power and control of a test, and washback have been shown to inextricably intertwine in high-stakes testing (e.g., Gu & Saville, 2016; Qi, 2004). Third, this study did not collect baseline data before HKDSE-English's graded approach was introduced and only captured learners' self-report data of how the graded approach affected their behaviour after the implementation had already taken place. Longitudinal studies that examine behaviour both before and after test reform, including those that corroborate self-report data from questionnaires or interviews with an observational component, such as diary analysis (e.g., Gosa, 2004; Tsagari, 2009) or classroom observation (e.g., Alderson & Hamp-Lyons, 1996), would have more robustly delineated how this graded approach has affected aspects of students' learning over time. Finally, the findings of our study must be viewed through the prism of students' perceptions. For a more comprehensive understanding of the multidimensional connections in washback, teaching and learning washback need to continue to be investigated together, ideally triangulating student perceptual data with other sources of evidence. In the case of HKDSE-English's graded approach, insights on how the test has impacted teachers' and school's decisions would seem to be necessary to better understand how these, in turn, affect learners' decisions. Therefore, future research needs to consider a range of stakeholder perspectives as part of the backdrop for interpreting students' perceptions.

Acknowledgements

We thank our participants, Yik Hong Terence Wong and other personnel who helped with recruitment, Qin Xie for assisting with data analysis, and Siyan Tan for her contributions, including serving as an independent coder.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Chi Lai Tsang  <https://orcid.org/0000-0002-1160-2686>

Open Practice



The data and materials from the present experiment are available are publicly available at the Open Science Framework website: <https://osf.io/y6qbv/>

Supplemental material

Supplemental material for this article is available online.

References

- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280–297. <https://doi.org/10.1177/026553229601300304>
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115–129. <https://doi.org/10.1093/applin/14.2.115>
- Allen, D. (2016a). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia*, 6(7), 1–20. <https://doi.org/10.1186/s40468-016-0030-z>
- Allen, D. (2016b). Japanese cram schools and entrance exam washback. *The Asian Journal of Applied Linguistics*, 3(1), 54–67. <https://caes.hku.hk/ajal/index.php/ajal/article/view/338/412>
- Andrews, S. J., Fullilove, J., & Wong, Y. (2002). Targeting washback: A case study. *System*, 30(2), 207–233. [https://doi.org/10.1016/S0346-251X\(02\)00005-2](https://doi.org/10.1016/S0346-251X(02)00005-2)
- Berry, R. (2011). Assessment trends in Hong Kong: Seeking to establish formative assessment in an examination culture. *Assessment in Education: Principles, Policy and Practice*, 18(2), 199–211. <https://doi.org/10.1080/0969594X.2010.527701>
- Booth, D. (2012). *Exploring the washback of TOEIC in South Korea* [Unpublished PhD thesis]. University of Auckland.
- Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*, 24(3), 279–301. [https://doi.org/10.1016/S0191-491X\(98\)00018-2](https://doi.org/10.1016/S0191-491X(98)00018-2)

- Cheng, L., Andrews, S., & Yu, Y. (2011). Impact and consequences of school-based assessment in Hong Kong: Views from students and their parents. *Language Testing* 28(2), 221–250. <https://doi.org/10.1177/0265532210384253>
- Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, 48(4), 436–470. <https://doi.org/10.1017/S0261444815000233>
- Chik, A., & Besser, S. (2011). International language test taking among young learners: A Hong Kong case study. *Language Assessment Quarterly*, 8(1), 73–91. <https://doi.org/10.1080/15434303.2010.537417>
- Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press. <https://rm.coe.int/1680459f97>
- Creswell, J. W. (2015). *A concise introduction to mixed methods research*. SAGE Publications.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum Associates.
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 191–210). Lawrence Erlbaum Associates.
- Fox, J., & Cheng, L. (2007). Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first and second language test-takers. *Assessment in Education: Principles, Policy and Practice*, 14(1), 9–26. <https://doi.org/10.1080/09695940701272773>
- Gosa, C. M. C. (2004). *Investigating washback: A case study using student diaries* [Unpublished PhD thesis]. Lancaster University.
- Green, A. (2006). Washback to the learner: Learner and teacher perspectives on IELTS preparation course expectations and outcomes. *Assessing Writing*, 11(2), 113–134. <https://doi.org/10.1016/j.asw.2006.07.002>
- Gu, X., & Saville, N. (2016). Twenty years of Cambridge English examinations in China: Investigating impact from the test-takers' perspectives. In G. Yu & J. Yan (Eds.), *Assessing Chinese learners of English* (pp. 287–310). Palgrave Macmillan.
- Hong Kong Examinations and Assessment Authority. (2013a). *The graded approach*. http://www.hkeaa.edu.hk/DocLibrary/HKDSE/Subject_Information/eng_lang/ENG-Grading_Approach_1113.pdf
- Hong Kong Examinations and Assessment Authority. (2013b). *Results of the benchmarking study between IELTS and HKDSE English Language Examination*. http://www.hkeaa.edu.hk/DocLibrary/MainNews/press_20130430_eng.pdf
- Hong Kong Examinations and Assessment Authority. (2018). *HKDSE statistics overview*. http://www.hkeaa.edu.hk/DocLibrary/HKDSE/Exam_Report/Examination_Statistics/dseexam-stat18_1.pdf
- Hong Kong Examinations and Assessment Authority. (2019). *HKDSE English language assessment framework*. http://www.hkeaa.edu.hk/DocLibrary/HKDSE/Subject_Information/eng_lang/2021hkde-e-elang.pdf
- Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, 32(1), 51–62. <https://doi.org/10.1080/10447318.2015.1087664>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36. <https://doi.org/10.1007/BF02291575>
- Knoch, U., Huisman, A., Elder, C., Kong, X., & McKenna, A. (2020). Drawing on repeat test takers to study test preparation practices and their links to score gains. *Language Testing*, 37(4), 550–572. <https://doi.org/10.1177/0265532220927407>
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 243–256. <https://doi.org/10.1177/026553229601300302>

- Mickan, P., & Motteram, J. (2009). The preparation practices of IELTS candidates: Case studies. In J. Osborne (Ed.), *IELTS Research Report vol. 10* (pp. 223–262). IELTS Australia; British Council. https://www.ielts.org/-/media/research-reports/ielts_rr_volume10_report5.ashx
- Pan, Y. C. (2014). Learner washback variability in standardized exit tests. *The Electronic Journal for English as a Second Language*, 18(2), 1–30. <https://files.eric.ed.gov/fulltext/EJ1045139.pdf>
- Patton, M. Q. (2014). *Qualitative research and evaluation methods: Integrating theory and practice*. SAGE Publications.
- Poon, A. Y. (2013). Will the new fine-tuning medium-of-instruction policy alleviate the threats of dominance of English-medium instruction in Hong Kong? *Current Issues in Language Planning*, 14(1), 34–51. <https://doi.org/10.1080/14664208.2013.791223>
- Purpura, J. E. (1999). *Learner strategy use and performance on language tests: A structural equation modeling approach*. Cambridge University Press.
- Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 171–190). Lawrence Erlbaum Associates.
- Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers*. Wiley-Blackwell.
- Sato, T. (2019). An investigation of factors involved in Japanese students' English learning behavior during test preparation. *Papers in Language Testing and Assessment*, 8(1), 69–95. https://arts.unimelb.edu.au/_data/assets/pdf_file/0003/3060417/8_1_S4_Sato.pdf
- Saville, N. (2010). Developing a model for investigating the impact of language assessment. *Research Notes*, 42, 2–8. <https://www.cambridgeenglish.org/Images/23160-research-notes-42.pdf>
- Shih, C. M. (2007). A new washback model of students' learning. *Canadian Modern Language Review*, 64(1), 135–162. <https://doi.org/10.3138/cmlr.64.1.135>
- Shohamy, E. (2006). *Language policy: Hidden agendas and new approaches*. Routledge.
- Shohamy, E. (2014). *The power of tests: A critical perspective on the uses of language tests*. Routledge.
- Smart, C., Drave, N., & Shiu, J. (2014). Implementing innovation: A graded approach to English language testing in Hong Kong. In D. Coniam (Ed.), *English language education and assessment* (pp. 257–273). Springer.
- Stoneman, B. W. (2006). *The impact of an exit English test on Hong Kong undergraduates: A study investigating the effects of test status on students' test preparation behaviours* [Unpublished PhD thesis]. The Hong Kong Polytechnic University. <http://hdl.handle.net/10397/2720>
- Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self-and other-perception of second language speech. *Bilingualism: Language and Cognition*, 19(1), 122–140. <https://doi.org/10.1017/s1366728914000832>
- Tsagari, D. (2009). Revisiting the concept of test washback: Investigating FCE in Greek language schools. *Cambridge ESOL Research Notes*, 35, 5–10. <https://www.cambridgeenglish.org/images/23154-research-notes-35.pdf>
- Tsang, C. L. (2017). *Examining washback on learning from a sociocultural perspective: The case of a graded approach to English language testing in Hong Kong* [Master's thesis, University College London]. British Council. https://www.teachingenglish.org.uk/sites/teacheng/files/chi_lai_tsang_ucl_dissertation.pdf
- Tsang, C. L. (2020). *Washback on Student Learning (WSL) Questionnaire* [Measurement instrument]. IRIS Digital Repository. Available at: <https://www.iris-database.org/iris/app/home/detail?id=york:937987>

- Turner, C. E., & Purpura, J. E. (2016). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 255–274). De Gruyter Mouton.
- Weir, C. J. (2005) *Language testing and validation: An evidence-based approach*. Palgrave Macmillan.
- Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806–838. <https://doi.org/10.1177/0011000006288127>
- Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly*, *10*(2), 196–218. <https://doi.org/10.1080/15434303.2012.721423>
- Xie, Q. (2015). Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism. *System*, *50*, 56–68. <https://doi.org/10.1016/j.system.2015.03.002>
- Xie, Q., & Andrews, S. (2012). Do test design and uses influence test preparation? Testing a model of washback with structural equation modeling. *Language Testing*, *30*(1), 49–70. <https://doi.org/10.1177/0265532212442634>
- Yung, K. W. H. (2015). Learning English in the shadows: Understanding Chinese learners' experiences of private tutoring. *TESOL Quarterly*, *49*(4), 707–732. <https://doi.org/10.1002/tesq.193>
- Zhan, Y., & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: Insights from possible self theories. *Assessment in Education: Principles, Policy & Practice*, *21*(1), 71–89. <https://doi.org/10.1080/0969594X.2012.757546>
- Zhan, Y., & Wan, Z. H. (2016). Test takers' beliefs and experiences of a high-stakes computer-based English listening and speaking test. *RELC Journal*, *47*(3), 363–376. <https://doi.org/10.1177/0033688216631174>