Universitatea de Vest din Timișoara
Școala Doctorală de Științe Umaniste
Domeniul Filologie

INVESTIGATING WASHBACK EFFECTS OF CAMBRIDGE B2+ EXAMS ON STUDENTS. A UNIVERSITY EDUCATION CASE STUDY

Doctorand: Diana Ramona Botoșan

Conducător științific: Prof. dr. habil. Codruța Goșa

2023

Table of Contents

List of Figures / 6

Introduction / 7

Chapter 1: Washback in Language Testing / 13
1.1. Introduction to the chapter / 13
1.2. Principles in the design of language tests / 13
1.3. Washback / 18
1.3.1. What is washback? / 18
1.3.2. Labels used for washback / 21
1.3.3. Forms of washback / 22
1.3.3.1. Positive and negative washback / 22
1.3.3.2. Overt and covert washback / 24
1.4. Can washback be a criterion in test design? / 25
1.5. Operationalising washback / 28
1.6. Does washback exist? / 32
1.7. Conclusion to the chapter / 36

Chapter 2: Cambridge English Tests and the Romanian Tertiary Education System / 38
2.1. Introduction to the chapter / 38
2.2. Cambridge Assessment English / 38
2.2.1 Range of Cambridge exams / 39
2.2.2. Cambridge First and Advanced / 42
2.3. The place and face of university education in the Romanian education system / 43
2.3.1. The system of education in Romania in a nutshell / 43
2.3.2. A brief history of tertiary education in Romania / 44
2.3.2.1. The 20th century and its communist legacy / 46
2.3.2.2. New horizons / 50
2.3.3. The current status of university education in Romania / 55
2.3.4. English and other foreign language studies in Romania and the role of Cambridge exams / 56
2.4. Conclusion to the chapter / 61

Chapter 3: Methodological Underpinnings / 62
3.1. Introduction to the chapter / 62
3.2. General remarks on research methodology / 62
3.2.1. The qualitative-quantitative continuum / 63
3.3. Qualitative research / 65
3.3.1. Forms of qualitative research / 65
3.3.2. Research instruments in qualitative research / 66
3.3.3. Qualitative data analysis / 67
3.4. Quantitative research / 67
3.4.1. Quantitative research instruments / 68
3.4.2. Quantitative data analysis / 69
3.5. Mixed-method research / 69
3.6. Situating my research / 70
3.6.1. Focusing in on the approaches of my research / 71
3.6.1.1. Approaches / 71
3.6.1.1.1. Case study research / 71
3.6.1.1.2. Ethnographic research / 73
3.6.1.1.3. Diary studies research / 74
3.6.1.1.3.1. Criticism of diary studies / 77
3.6.2. Instruments I use in my research / 77
3.7. Ethics in research / 79
3.8. Conclusion to the chapter / 81

Chapter 4: The Questionnaire-Based Survey on Students' Views / 83
4.1. Introduction to the chapter / 83
4.2. The questionnaire-based survey / 84
4.2.1. Good practice in questionnaire design / 87
4.3. The respondent profile and administration of the questionnaire / 91
4.4.The findings and their interpretation / 92
4.4.1. Summary of the findings / 104
4.5.Conclusion to the chapter / 105

Chapter 5: The Focus Group. Analysis and Interpretation / 107
5.1. Introduction to the chapter / 107

5.2. On focus groups / 107
5.2.1. Defining focus groups / 108
5.2.1.1. Online focus groups / 109
5.2.2. Matters to consider / 110
5.2.1.1. Design / 110
5.2.2.2. Organisation / 111
5.2.1.3. Moderating / 111
5.2.2. The process of analysis / 112
5.2.3. Criticism of focus groups / 113
5.2.4. Transcribing focus group data / 114
5.3. My focus group / 115
5.3.1. Aspects related to the administration / 115
5.3.2. Findings and their interpretation / 116
5.3.2.1. Interaction / 116
5.3.2.2. Signs of washback / 119
5.4. Conclusion to the chapter / 123

Chapter 6: The Diary Study. Analysis and Interpretation / 124
6.1. Introduction to the chapter / 124
6.2. Diary studies as data collection instrument / 124
6.3. My diary study / 126
6.3.1. Phases of diary keeping in my diary study / 126
6.4. The findings and their interpretation / 129
6.4.1. Topic related / 129
6.4.1.1. Overt washback of Cambridge exams / 129
6.4.1.2. Covert washback of Cambridge exams / 132
6.4.1.3 Positive attitudes to washback / 136
6.4.1.2. Topic unrelated / 139
6.4.3. Feelings-related language / 141
6.5. Conclusion to the chapter / 144

Chapter 7: Putting It All Together / 146
7.1. Introduction to the chapter / 146
7.2. Revisit of the research questions and summary of the findings / 146

7.2.1 Answers to the research questions / 147
7.2.2. Summary of the findings by the instruments used / 150
7.3. Implications and recommendations suggested by the study / 151
7.4. Limitations of the study / 153
7.5. Further lines of research / 155
7.6. Conclusion to the chapter / 156

Conclusions / 157

Works Cited / 160

Appendices / 173
Appendix 1. The Questionnaire / 173
Appendix 2. The Focus Group Transcript / 170
Appendix 3. The Diary Data / 183

Abstract

Key words: washback, Cambridge exams, students, survey, focus group, diary studies

We live in an exams-dominated world, whether we like it or not, whether these exams are standardised or not, or whether they are constructed by observing good practice in their design, as the theoreticians and practitioners in testing and assessment advise us to do, especially when it comes to highstakes exams. Highstakes exams or tests are those which have a major impact on the stakeholders, be they students, teachers, administrators, markers, parents, or decision-makers alike. This is why exams or tests, especially highstakes ones, are believed (and a number of research studies have shown it) that they have the power to produce changes to the way teachers teach and learners learn. This effect is generally known as washback. Washback is expected to manifest itself wherever highstakes exams (to be defined in more detail later on in this thesis) appear in certain contexts, be it secondary or tertiary level education. It is an object of study in language testing and, as such, it is an applied linguistics rooted phenomenon. It is research into washback at tertiary level that the study presented in this thesis sets out to do.

Washback, as mentioned above, an applied linguistics phenomenon, studied under the umbrella of language testing, has been rather extensively and intensively investigated ever since mid-1990s after the publication of Alderson and Wall (1993) seminal article *Does washback exist?* Due to the popular, among testers, belief that it can be used as a "lever for change" (Pearson, 1988) it has been taken for granted and evoked frequently. However, research so far has shown that it is a complex phenomenon which does not appear automatically and straightforwardly as was claimed. Moreover, it cannot be used as a means of creating the desired change in the curriculum and the teaching and learning strategies adopted by stakeholders. For achieving these changes, research has shown the need to adopt successful models of washback and consider the many factors involved and levels at which it may appear. Consequently, due to its potential power to produce change, besides the interest provoked, many primary, applied,

empirical, studies have so far been conducted. Only in the last decade, as shown in this study, a plethora of various studies on washback have been done, covering most continents.

So far as research on washback in the Romanian context is concerned on the other hand, it does not seem to join this trend. Barely a couple of such research could be discovered when searching for the relevant literature. Both studies discovered are mentioned in the research report instantiated in this thesis and they refer to washback at upper secondary school level. Thus, to the best of my knowledge, this is a first attempt to study washback at university level. As for the highstakes exam whose washback was under scrutiny, they represent external, international exams developed by the University of Cambridge, its Cambridge Assessment English Branch, and they have become quite popular in Romania, the majority of the test-takers being high school students. Cambridge B2+ exams (First and Advanced) were the tests and their washback focused on in this study. They were considered highstakes due to the impact on academic and job trajectories of the students. Thus, in my study, I set out to investigate washback to university students, namely, how Cambridge B2+ exams affect the way they are taught and learn. Moreover, the aim of the present study is not just to ascertain the existence of washback effects of a particular highstakes exam (Cambridge B2+ suite of English as a Foreign Language exams) in a particular context (a Romanian state university), but it also proposes a deeper exploration of the phenomenon for adding up to the existing knowledge into its nature and operation. The three broad research questions I asked are both of a more general nature, such as the first research question referring to Cambridge exams as being or not highstakes exams in the Romanian context, or the last one, methodology related, and of a topic-related nature. As such, the remaining overarching question, with its five sub-questions, refers directly to the presence, forms, and levels of washback.

As far as the methodology employed is concerned, I use a mixed method paradigm and a multi-method strategy through which I focus on the student-respondent perspectives. As such, though my approach is mainly qualitative, being a case study conducted in a certain university, quantitative approaches in a questionnaire-based small-scale survey are also adopted. The qualitative data was constructed by using a focus group and a diary study. It is important to highlight that in the design and administration of the entire process generally, and specifically of its instruments, I tried to observe what is regarded, by theoreticians and practitioners alike, as good practice. I was similarly careful to observe good ethical practices in research, such as voluntariness and confidentiality.

The presentation of the study is shaped in seven core chapters, preceded by an introduction, where I briefly present the topic and its rationale, my research questions and I announce the structure of the thesis. It is succeeded by a conclusion where I very briefly summarise the entire endeavour. Three of the core chapters discuss the study from a theoretical point of view, the remaining four chapters being dedicated to the primary, applied study, revolving around the three instruments used for the data collection, the questionnaire, the focus group, and the diary study. In the final core chapter, I bring together the strings of the multi method analysis. The core chapter are presented in more detail as follows.

The first chapter represents the literature review on the way in which the phenomenon of washback was contextualised, its theories and models, and I overview the latest applied studies into its manifestation and function. To this end, a number of theoretical studies are evaluated from a chronological perspective, but also by synthesising the most important findings. It is in the same first chapter that 18 applied studies on washback of English language exams, conducted in the last decade, are overviewed and their most important and relevant findings for my analysis are summarised and highlighted. All these theoretical and empirical studies, though leading to a better understanding of the reasons why washback seems to occur, as well as working towards clarifying the way in which tests in general, and highstakes test in particular, can be designed and administered as to work towards desired washback, they also reveal how complex the phenomenon of washback is and the many factors and levels surrounding its appearance and operation. Consequently, if language testers in general and English language tester in particular want to ascertain that such tests can truly cause desired washback effects, many more studies are needed for understanding washback in its entirety, for making this possible.

The second chapter has two main parts dedicated to the contextualisation of my research. The first one puts forward an analysis of Cambridge exams, including their role and status. Cambridge exams are, indeed, designed and administered by carefully observing good practice in test design. They are aligned with the Common European Framework of Reference, issued by the Council of Europe to ensure standardisation. More than that, they are pre-tested and then analysed by item before the actual administration. There is a wide range of exams offered, having a wide range of test-takers in view, however some of the most popular around the world, including Romania, being B2+ exams. In the second part of this chapter, the Romanian educational context is scrutinised diachronically, and its present status being also focused upon.

The role of Cambridge exams in the secondary and tertiary education was also the object of my scrutiny. As such, I conclude that the Soviet and Ceausescu's totalitarian regimes have left deep scars on the system of education in Romania and after the rather turbulent 1990s, the new millennium has brought about several beneficial changes and reforms, the closest in relevance for this study being the status of Cambridge exams in the Romanian context and the recognition of Cambridge exams for admittance, graduation, and mobility purposes, which turn them into highstakes exams, thus enabling them to produce washback effects.

In the third chapter I discuss aspects related to the methodology of research relevant for the present study, from a theoretical point of view, as can be found in the literature. It starts from a more general perspective, such as the qualitative-quantitative paradigms conceptualised as a continuum, the approaches available to a researcher, as identified in the literature with their general and more specific characteristics, as well as the methods relevant for my research. Then I also very briefly mention the instruments I use for data collection, introducing their most important features: questionnaires, focus groups and diaries, all used for eliciting the students' views. Finally, in this chapter I speak about the importance of ethics in research, and I discuss, in short, the core principles.

In chapter four, the results of the questionnaire-based survey are examined. First, I try to show how I took into account good practice in the designing and administration of the questionnaire, which I also briefly refer to. The google form questionnaire addressed 238 students whose teacher/instructor I was, and after receiving 127 completed forms, I analysed the answers received and present them in the form of graphs which I then interpret. The quantitative analysis showed that the majority of the student-respondents were familiar with Cambridge exams, they found them useful and trustworthy, and believe that the teaching of English that their teacher/instructor uses influences her teaching, as they expected it to be. Thus, at least as far as the general view is concerned, the student-respondents' perceptions of washback effects of Cambridge First and Advanced exams are extensively present in the survey data. However, the results of the survey cannot be generalised, due to the nature of case-study research. The only extrapolation possible is that to the students targeted by the case study.

In chapter five, I deal with the focus group I conducted and moderated. The features of focus groups are first discussed, then I present the analysis, results, and drawbacks. Nine student-respondents agreed to participate in the online focus-group, which I then transcribed. The data-driven content analysis revealed two themes, one referring to the interaction between the

participants and the moderator, the second one referring to the presence of washback. The analysis of the focus group in terms of interaction showed that the interaction was not appropriate to a focus group, taking the shape of a group interview. The data was not as rich as it was expected, the participants were rather reluctant to intervene, the moderator resorting to frequent prompts, many of which remaining unanswered. This unnatural interaction was due to the moderator's lack of experience in facilitating discussions, the participants' lack of experience of participating in focus groups and, most probably, due to the online format of the focus group. This finding has a methodological value and should be further investigated. Anyway, it is indicative that, in such cases, professional and experienced moderators are to be preferred, even if not being the researchers themselves, and even if the research is ethnographic. Similarly, more, and better preparation of the focus group is needed, to make the participants feel at ease. As for the presence of washback, the participants definitely felt it as being positive and rather intense, affecting both their classroom and their private studies, especially the content. More than that, they showed they believe that washback is to be desired and, therefore, expected.

Chapter six represents the last data analysis chapter of the thesis, dedicated to the diary study. Before discussing the results, I begin by briefly revisiting diary studies as data collection instrument and their characteristics, with direct reference to my diary study. Then I present the results of the diary data analysis and the interpretation of the findings. The analysis was done qualitatively and as such it was inductive, and data-driven, however also having in view the theory and practice of washback studies. Though no hypotheses were formulated, and as researcher I tried to keep an open mind in the process of analysis, no researcher could claim a blank mind. The review of the relevant literature and the stance s/he adopts, as well as a certain degree of researcher-related subjectivity play a part in the way the data is analysed. The diary data displayed rather strong signs of washback, both overt and covert, both positive and negative. After all, as Prodromou (1995) claims, the very presence of overt washback is a form of negative washback, and it indicates a narrowing down of the curriculum, a finding also present in my data. The washback effects appeared at the level of the skills practiced, the types of tasks and the materials used, both in the classroom and in private studies. Positive attitudes towards Cambridge exams which were found both in the analysis of the questionnaire and the focus group data were also found in the diary data. The positive attitudes were predominant. However, negative attitudes were also present, to a much smaller extent, and point to the direction mentioned above - the narrowing down of the curriculum. A salient, only one

occurrence finding suggests a new level at which washback might appear. This is a level which is not mentioned in the literature and which I would call washback of the highstakes exam to the classroom/mid/end term (university) exams. I believe that this should be further explored. Added to that, the data driven analysis reveals that the diaries contain quite a number of references which were not related to the topic (such as other courses at the university, family, hobbies, etc.). They also contain references to feelings expressed through semantically loaded words and other rhetoric devices. All these have a methodological value connected to the nature of diary studies themselves. As the theory and practice of diary studies show, the nature and strength of diary studies lies in their capacity to produce rich data and reveal attitudes and feelings which can be further explored and reveal new research paths. They give access to the minds of the respondents, and elicit a unique kind of data, in a particular kind of way, prompting candidness and openness.

In chapter seven, attempts are made to pull the strings together of the mixed method and multi-instrument-based analysis proposed in this thesis. An overarching, higher level interpretation of the results is similarly attempted. Additionally, the limitations are also discussed in this chapter, as well as suggestions for further research of the phenomenon of washback, grounded in the findings of my study. To this end, I have tried to show the way in which the results obtained through the mixed and multiple instruments method used feed into and complement each other by shedding s bit more light on the complexity of the phenomenon of washback which also led to a number of implications directed at stakeholders, such as the need to research awareness of washback as part of assessment literacy in general and at the tertiary level in particular. It is also in this final chapter that I offer a reflection on the way in which the entire endeavour has functioned, particularly on its limitations and drawbacks, such as resorting to a limited perspective (that of the students), as well as on discovering lines of research which would somehow mitigate these drawbacks and contribute further to the understanding of the phenomenon of washback, to the construction of a model through which washback might lead to produce beneficial changes to the teaching, learning and testing of English as a foreign language.

As mentioned before, a brief conclusion ends the content-based part of my thesis which aimed at highlighting once more the most important and relevant aspects of the entire endeavour. The last two sections list the works cited and the details regarding the instruments used: the template

of the questionnaire, the transcript of the focus group and the student diaries processed verbatim in their entirety.

To sum up, as far as the findings of this endeavour go, data analysis and interpretation revealed the existence of washback, in the student-respondents' views, covert and overt, quite an intensive presence which triggered positive attitudes. Mostly affected by washback was the content of teaching, especially the materials and strategies used, both when it comes to the teaching and to the learning. However, negative washback was also perceived by the students hinting at the narrowing down of the curriculum. The study also revealed that the use of open-ended instruments does not automatically lead to the obtaining of rich data which can be analysed in a data-driven way, even aiming at grounded theory results. At least, in the case of conducting focus-groups, the moderator has to be quite experienced to ensure appropriate interaction, especially in the case of online focus groups. Similarly, the diary data, may turn out to be rich, nevertheless, much of this richness may not be topic related. For this reason, there were not many new elements referring to washback to be found through the data-driven analysis. Nevertheless, a serendipity finding the analysis did reveal was of an unexpected nature, related to the way in which (mid)exams appear to be influenced by Cambridge exams. This could be considered as a potential new level at which washback can appear and thus be added to the models of washback proposed so far, such as Alderson and Wall's 1993 *Washback Hypotheses* model. Insofar, this might be considered a modest contribution to the knowledge only if, after being turned into a hypothesis, it could be confirmed by further studies.