

The Test-Takers' Perspectives on the Oral Proficiency Interview of a High-Stakes English Language Examination

Petros Papadopoulos

Military Language Testing Center, Training Center Palaskas, Skaramagkas, Greece
Email: pepapado@yahoo.gr, p.papadopoulos@hellenicnavy.gr

Abstract. This paper focuses on the oral proficiency interview that officers of the Hellenic Armed Forces are required to take in order to have their speaking proficiency certified at Level 3 (Professional) and sets out to investigate their familiarity with the NATO STANAG 6001 Level 3 Speaking Descriptors in an attempt to gather validity evidence for the oral proficiency interview from their perspective. In doing so, an online survey was conducted and fifty-five military officers participated in it followed up by five interviews conducted with five different officers (who had first completed the online questionnaire). A qualitative approach to the analysis of the findings was adopted which aimed at gaining insight into test takers' perceptions about the oral proficiency interview in relation to the three research questions that this study revolves around. The study was conducted as part of my dissertation submitted in partial fulfillment for the degree of M.A. in Language Testing at Lancaster University, U.K.

Keywords: oral proficiency interview, validity, high-stakes examination, NATO STANAG 6001, speaking descriptors, linguistic interoperability

INTRODUCTION

Due to the fact that speaking is the most salient of the four language skills (Bachman, 1988), “[t]he ability to speak in a foreign language is [considered to be] at the very heart of what it means to be able to use a foreign language” (Alderson & Bachman, 2004, p. ix). In fact, speaking in a foreign language is very challenging and developing competence in speaking is a time-consuming process, Alderson and Bachman (2004) further note. In addition, there seems to be a commonly held view among language testers that speaking is the most difficult of the four language skills to assess (Ginther, 2012), and, according to Bachman (1988), assessing a person’s speaking ability can be “problematic because of the complexity both of the skills and the context in which these skills are to be elicited and assessed” (p. 150).

The present study falls within the scope of language testing in the military. It focuses on the assessment of the speaking proficiency of the officers of the Hellenic Armed Forces through oral proficiency interviews in the form of a semi-structured interaction between the test taker (a

military officer) and the members of the examination committee. Following Ginther (2012), the oral proficiency interview is a direct method of assessing a person's speaking skills and abilities in actual performance involving "a series of warm up questions followed by a subsequent series of increasingly difficult questions where examinees are expected to display concomitantly increasing levels of complexity in their responses" (p. 1).

The oral proficiency interview, which is part of the English language examination that all officers of the Hellenic Armed Forces are required to take in order to be eligible for any post within the North Atlantic Treaty Organization (NATO), is conducted in accordance with the NATO STANDARDIZATION AGREEMENT (STANAG) 6001. The NATO STANAG 6001 describes six language proficiency levels coded 0 (No proficiency) through 5 (Highly-articulate native). Military officers are usually required to achieve Level 3 (Professional) in Speaking (as well as in the other three language skills). Following Berger (2015), the NATO STANAG 6001 is the *de facto* test construct.

It results then that the oral proficiency interview is part of a high-stakes examination. For Kunnan (2012), a test is considered high stakes when the test scores/results are used to make important career and life-changing decisions which can have serious consequences for test takers. In a similar vein, Shohamy (2001a) concedes that tests are powerful and play a major role as they can lead to far-reaching and high-stakes decisions and consequences about individuals and groups; "tests can create winners and losers, successes and failures, the rejected and the accepted" (p. 374).

Test takers have long been considered "important stakeholders" in language testing, perhaps the most important ones according to Rea-Dickins (1997), whose views "are among the most difficult to make sense of and to use" (p. 306). Based on Alderson and Clapham (1992), she further points out that "drawing conclusions about those views is not easy since often [test takers] contradict one another or may have no views at all". Brown (1993) and Wall, Clapham and Alderson (1994) are among the first who have practically demonstrated how and to what extent test takers' feedback and views can contribute to the development and validation of tests echoing Kenyon and Stansfield (1991, as cited in Brown, 1993) who support that test takers' comments can significantly help test developers improve the quality of the test product. Despite the inherent perplexities surrounding the interpretation of test takers' views, Rea-Dickins (1997) puts forward the case for a more active involvement of all interested parties in the process of test development and test use and calls for greater stakeholder involvement which will lead to fairer, more ethical and more democratic testing practices. This seems to signal the end of an era during which the test taker was not of major concern and consequently had no rights but "was viewed mostly as a block box ... [and] was important only insofar as computing the psychometric traits of the test" (Shohamy, 2001b, p. 155).

This study aspires to contribute to the emerging body of knowledge in which the emphasis is on the test taker in an attempt to "broaden the scope of inquiry and contexts that inform knowledge about language assessment" (Cumming, 2004, p. 5). In the context of language testing in the military, test takers are viewed as the key stakeholders as decisions based on test scores/results can profoundly affect their careers and lives. In addition, this study is the first of its kind in the field of language testing in the Hellenic Armed Forces and seeks to foster a better understanding of the nature of oral language proficiency assessment at military command level and establish a more consistent channel of communication between language testers and those responsible for the implementation of military language testing policy (Crossey, 2009).

In my capacity as one of the oral proficiency interview interlocutors and raters, I have often felt that test takers (i.e. military officers) are not (adequately) familiar with Level 3 Speaking Descriptors of the NATO STANAG 6001 prior to taking the oral proficiency interview. It is believed that this can give rise to a number of misconceptions which are reflected in test takers' views on the oral proficiency interview, their performance on it, and what being a Level 3 speaker actually means. In order to investigate the "observed problem" further, a qualitatively-oriented inquiry into test takers' perceptions and beliefs has been conducted. Cumming (2004, p. 9) suggests that "[s]uch inquiry is indispensable for understanding why people perform the way they do in language assessments, and thus necessary for validation".

PART ONE: THEORETICAL BACKGROUND

A PROPOSED FRAMEWORK FOR GATHERING VALIDITY EVIDENCE

Shaw, Crisp and Johnson (2012) acknowledge the importance of achieving validity with relation to the construct assessed and propose a framework for evidencing validity in assessments. Their framework, which is influenced primarily by Kane (2006) but also by Chapelle (2008), has been adopted in this study because it provides a practical approach to gathering validity evidence in a structured and systematic way. More specifically, Shaw et al.'s framework is based around a chain of inferences which "flow" from the task to the test performance (construct representation), from the test performance to the test score (scoring), from test score to test competence (generalisation), from test competence to domain competence (extrapolation), and from domain competence to trait competence (decision making). For each inference, an associated warrant sets out a statement that is claimed to be true. In practice, the inferences and warrants make up the interpretive argument. The validity argument is set out in the form of five validation questions which map onto the inferences. The validation questions seek to establish whether the tasks elicit performances reflecting the intended construct, whether the scores and grades are dependable measures of the construct, whether the tasks adequately sample the intended construct, whether the sampled construct is representative of wider competence in the domain, and whether guidance is available on the meaning of scores and how they should be used, and "are intended to guide the assessment professional through the validation effort in relation to data collection methods and analyses, whilst maintaining a strong link between the evidence collected and the inference which this evidence seeks to validate" (p. 166). To this end, Crisp et al. present a proposed set of data sources and methods of data collection that can provide evidence for each of the validation questions. Analysis of the evidence collected will provide evidence for validity (i.e. support for validity), and may suggest potential threats to validity, which, in turn, can indicate areas for improvement and further work.

Interpretive argument		Validity argument		
Inference	Warrant justifying the inference	Validation question	Evidence for validity	Threats to validity
Construct representation	Tasks elicit performances that represent the intended constructs	1. Do the tasks elicit performances that reflect the intended constructs?		
Scoring	Scores/grades reflect the quality of performances on the assessment tasks	2. Are the scores/grades dependable measures of the intended constructs?		
Generalisation	Scores/grades reflect likely performances on all possible relevant tasks	3. Do the tasks adequately sample the constructs that are set out as important within the syllabus?		
Extrapolation	Scores/grades reflect likely wider performance in the domain	4. Are the constructs sampled representative of competence in the wider subject domain?		
Decision-making	Appropriate uses of scores/grades are clear	5. Is guidance in place so that stakeholders know what scores/grades mean and how the outcomes should be used?		

Figure 1: Shaw, Crisp and Johnson's (2012) proposed validation framework

PART TWO: RESEARCH CONTEXT

LINGUISTIC INTEROPERABILITY

Although both English and French are officially recognized as the two working languages at NATO, English is the *de facto* operational language within NATO, and linguistic interoperability has become a high priority for both Allied and Partner countries especially due to an ever increasing number of peace-support operations (Crossey, 2005; Dubeau, 2006).

Crossey goes as far as to suggest that linguistic misunderstandings arising from inadequate language skills of non-native speakers of English on peace-support operations risk leading to mistakes, which may prove fatal and result in casualties. He states unequivocally that “linguistic interoperability is as important to ensuring that countries are able to participate effectively in both NATO and wider Alliance activities as any other form of interoperability”. In a similar vein, Adubato and Efthymiopoulos (2014, p. 27) support that good language skills are conducive to a successful administrative and operational environment as they help avoid “possible misunderstandings or misinterpretations across the whole spectrum of military activities and ... duplication of efforts in NATO operations and/or NATO working environments”. It results, then, that language testing in the NATO context has taken on added importance and carries high stakes not only for individual military members but for participating countries as well (Dubeau, 2006).

THE NATO STANAG 6001

The Bureau for International Language Co-ordination (BILC) was established in 1966 within the NATO Training Group (NTG)/Joint Services Subgroup (JSSG) as a consultative and advisory body for language training matters in NATO (Dubeau, 2006). Its mission, as stated on its website (www.natobilc.org), is “[t]o promote and foster interoperability among NATO and PfP [Partnership for Peace] nations by furthering standardization of language training and testing, and harmonizing language policy”. This is mainly accomplished through a series of BILC initiatives, such as an annual conference and language testing seminars and workshops, which contribute to a mutual understanding of testing practices and developments among participating countries.

During the 1970s, the BILC developed a set of language proficiency levels which were derived from the ILR scale and in 1976 were adopted by NATO as the NATO STANDARDIZATION AGREEMENT (STANAG) 6001: Language Proficiency Levels (Green & Wall, 2005) reflecting a NATO-wide need to define language proficiency and promote a shared understanding of the language proficiency of its members (Dubeau, 2006). Currently in its fifth edition (published in May 2016), the NATO STANAG 6001 (hereafter referred to as “the STANAG”) is “the common standard (construct) for language curriculum and test development” throughout NATO responding thus to interoperability requirements. In essence, it is used as the *de facto* test construct (Berger, 2015). To this end, the STANAG provides NATO forces with a table describing proficiency levels in each of the four language skills broken down into six levels coded 0 through 5, as follows: Level 0 – No proficiency, Level 1 – Survival, Level 2 – Functional, Level 3 – Professional, Level 4 – Expert, Level 5 – Highly-articulate native.

There are also five “plus levels” in between the base levels 0 to 4 which are to be used when a candidate’s language performance exceeds a base level but does not successfully meet all of the criteria to be awarded the next higher base level. For each NATO post there is a four-digit Standardized Language Profile (SLP) indicating the required level for each language skill in the following order: listening, speaking, reading, and writing. In order to be considered eligible for any NATO post, candidates have to achieve the SLP required by the specific post.

Dubeau (2006) maintains that although the STANAG testing system presents itself as standardized, in practice each country develops its own tests and establishes its own procedures for measuring language proficiency in accordance with the STANAG (Green & Wall, 2005).

Dubeau points out that in fact “there are no standardized testing instruments administered throughout countries ... [only] common testing practices are suggested” (p. 3). Although the BILC, as the custodian of the STANAG, provides testing training and consultation to national STANAG testers and testing teams with a view to helping them develop a better understanding of their own testing practices and better monitor their testing outcomes, in the absence of an effective mechanism in place which would ensure that the SLPs reported from one country are equivalent to the scores from another (Dubeau, 2006), the tests developed by each country run the risk of being viewed merely as “national interpretations of the language levels outlined in STANAG 6001 and often one nation’s Level 2 is another nation’s Level 3” (Aduabato & Efthymiopoulos, 2014, p. 29).

THE ORAL PROFICIENCY INTERVIEW IN FOCUS

The oral proficiency interview (hereafter referred to as “the interview”) is part of the English language examination that all officers of the Hellenic Armed Forces are required to take in order to be eligible for any post within NATO. It is conducted on the same day as the other three sections of the examination at the Military Language Testing Center, which is the official national NATO STANAG 6001 Testing Center, established in 2008, and is located at Training Center Palaskas, Skaramagkas, Athens, Greece.

The interviews are conducted and scored by the members of the examination committee, who are in their majority military officers from the three branches of the Hellenic Armed Forces. The examination committee typically consists of seven members but not all of them are required to be present (and are not usually present due to other professional commitments) on the day of the interviews; the committee consists minimally of four members, that is the Head, a senior officer at the rank of Army or Air Force Colonel or Navy Captain, and three other members.

The interview is in the form of a ten- to fifteen-minute long semi-structured interaction between the candidate¹ (i.e. the military officer) and the examiners about a topic of general interest with social and/or professional implications. In this context, the interview seeks to assess an individual’s unrehearsed, general proficiency in speaking as this is outlined in the STANAG’s Speaking Descriptors. Military officers are usually required to achieve Level 3 (Professional) in Speaking (https://www.natobilc.org/en/products/stanag-60011142_stanag-6001/).

The questions posed in the duration of the interview, which require different levels of cognitive engagement on the part of the candidate and seek to establish how well the candidate can elaborate on the topic under discussion, serve as *level checks* and *probes* and help determine the candidate’s level. Typically, at the beginning of the interview, candidates are asked some *Warm-up* questions which act as “ice-breakers” while the interview is usually brought to an end through a short *Wind-down* phase. Candidates are given a holistic rating based on four broad assessment criteria: content (what the test taker can talk about), tasks (what the test taker can do with the language), accuracy (how well the test taker can use the language), and the length and organization of the text produced.

¹ The terms “test taker” and “candidate” are used interchangeably in this study.

RESEARCH QUESTIONS

In an attempt to demystify the construct of the interview, I will rely on Shaw et al.'s (2012) framework to gather validity evidence solely from the test takers' perspective, who are the key stakeholders as the interview (and by extension the English language examination) carries high stakes for both their professional and personal lives. The research questions and the respective inferences that are being addressed are as follows:

Construct representation

1. What do test takers think of the content of the interview and the way it is conducted?

Generalisation

2. Do test takers think that the interview can effectively assess their speaking proficiency in English for interoperability purposes at Level 3?

Extrapolation

3. Do test takers think that their scores on the interview reflect their linguistic readiness for NATO posts requiring Level 3 in Speaking?

The “Scoring” and “Decision-making” inferences have not been addressed in this study due to the fact that the former requires expert judgment while as far as the latter is considered it is well established in the context of the specific testing situation that the scores on the interview are to be used for staff appointments within NATO only.

PART THREE: METHODOLOGY

TYPE OF DATA, PARTICIPANTS & DATA COLLECTION INSTRUMENTS

This study employs a mixed methods research design, which Dörnyei (2007, p. 44) defines it in the most straightforward manner as “some sort of a combination of qualitative and quantitative methods within a single research project”. Indeed, both quantitative and qualitative data are used in this study for the purpose of data triangulation. According to Dörnyei (2007), triangulation effectively reduces the chance of systematic bias through the corroboration of evidence coming from different data collection/analysis methods and ultimately contributes to strong validity evidence. However, in this study the emphasis is on the qualitative end of the quantitative-qualitative continuum. Following Johnson and Christensen (2004, as cited in Dörnyei, 2007), the sequence and the dominance of the method constituents of this study can be represented as *quan* → *QUAL* indicating that the qualitative component is dominant and is preceded by a quantitative phase. In this way, quantitative data can be utilized in a qualitative analysis legitimizing thus the research findings, what Dörnyei calls “qualitizing the data”.

The bulk of the data was collected through an online questionnaire which was administered to officers of the Hellenic Armed Forces (Army, Navy, Air Force) who took the

interview during the academic year 2015-2016. I opted for a non-probability convenience sample for reasons of accessibility and representativeness (Dörnyei, 2007; Dörnyei & Taguchi, 2010). Subsequently, a series of semi-structured interviews were conducted in order to elaborate on emerging patterns in the questionnaire responses which are of particular interest to the purposes of this study.

The questionnaire consists of a brief introduction/welcome message and twenty-eight questions in English. Questions one to seven ask for background information on the respondents (factual questions), while questions eight to twenty-eight ask for the respondents' opinion on the interview and other interview-related issues (attitudinal questions). The attitudinal questions have been designed with a view to providing evidence to support the inferences and answer the respective validation questions in Shaw et al.'s (2012) framework. In this way, the resulting data will help me answer my research questions.

The questionnaire items are all closed-ended questions with the exception of the last one which is an open-ended question asking for the participants' comments (for the reasons explained above, the open-ended question was optional; therefore the respondents could skip it). The majority of the closed-ended questions are multiple-choice items and there are only two rating scale items (a *Yes-No-Probably-Don't Remember* item and a Likert-type scale item).

An email inviting the recipients to take part in the online survey was sent out to 135 officers of the Hellenic Armed Forces (67 Navy officers, 63 Air Force officers, 5 Army officers). These military officers had voluntarily provided me with their email addresses after I had briefly explained to them the purpose of my study and their role in it at the end of various examination sessions throughout the year. The disproportionately small number of Army officers can be attributed to the fact that the examinations for the Army officers took place mainly between September and November 2015 when the conception of this study was at an early stage.

The second stage in the data collection process was to conduct a series of semi-structured interviews with a sub-set of the questionnaire respondents in order to "to probe some [of the questionnaire] themes in greater depth and detail" (Banerjee, 2004, p. 39). I interviewed five military officers (3 Navy officers, 1 Army officer, 1 Air Force officer) who expressed their interest in my project either in an email they sent me after they had completed the questionnaire or after I contacted them by phone and asked them to participate in a follow-up interview. Furthermore, I made sure that I interviewed at least one officer from each military branch in an attempt to make the sample as representative as possible and encourage officers from the three military branches to voice their opinions on the interview. The interviews were face-to-face sessions and were conducted in English. They lasted from 20 to 30 minutes and were recorded with the interviewees' consent.

The interviews evolved around an "interview guide" (Dörnyei, 2007), that is a list of questions which were prepared in advance and used during the interview in the form of prompts and probes. However, the exact sequence and wording of the questions was not followed with each interviewee. The interviews were not fully transcribed. Initially, I carried out an analysis of the audio data which involved taking notes while listening to the recordings, marking parts of the data that were most relevant to the purposes of the analysis. Subsequently, I carried out a partial transcription of the most interesting and relevant sections of each interview.

PART FOUR: RESULTS

QUESTIONNAIRE RESULTS

Fifty-five military officers (36 Navy officers, 16 Air Force officers, and 3 Army officers) completed the online survey. The number of the respondents is in line with both Hatch and Lazaraton (1991, as cited in Dörnyei, 2007), who recommend a minimum sample size of 30 people for L2 research, as well as with Dörnyei and Taguchi (2010), who maintain that in order for the results to be statistically significant, a sample size of around 50 respondents is required. Furthermore, it is worth noting that the response rate (41%) exceeds the thirty per cent response rate that questionnaires typically attract (Gillham, 2008). All respondents are male, with the exception of one female officer. They are all graduates of the respective Military Academies in Greece (equivalent to state universities). In addition, eight of them hold an undergraduate degree from a Greek university, twenty-six hold an MA either from a Greek university (13) or from a university in the UK or USA (13), and one holds a PhD from a Greek university. Nineteen officers have graduated from the Supreme Joint War College of Greece while three others have graduated from a Defence College abroad (these Colleges provide senior officers with education and training in military issues on a graduate level). Interestingly, more than half of the officers (58%) who completed the questionnaire stated that they have previous experience of working in NATO or in a NATO-related environment. Regarding their English language proficiency, thirty-two respondents (63%) hold a C2-level certificate, eight hold a C1-level certificate, seven hold a B2-level certificate and four hold a B1-level certificate, while four respondents indicated that they do not hold any English language certification in accordance with the Common European Framework of Reference for Languages (CEFR). Commenting on their overall speaking proficiency, twenty-seven officers opted for “occasional difficulty”, twenty-one indicated that they have “no difficulty” while seven have “some difficulty”. Nobody opted for “great difficulty”.

The next question aimed to explore the officers’ familiarity with the STANAG Level 3 Speaking Descriptors, an issue which has actually propelled the conception of this study. The respondents could select more than one answer that applied to them. Only sixteen per cent of the officers stated that they are familiar with the Speaking Descriptors because they have actually read through them while another sixteen per cent of them stated that they “have been informed about [the Speaking Descriptors] from other colleagues/candidates”. Most officers (36%) stated that they have no idea what the Speaking Descriptors are about while an equally considerable percentage (31%) of them seem to have an intuitive knowledge of the Speaking Descriptors (“I feel I know what they are about although I haven’t read them”). Finally, seven per cent of the surveyed officers see no reason why they should be familiar with the Speaking Descriptors.

More than half of the officers (53%) took the interview because it was part of the English language examination they were required to take as they had applied for a NATO post. Some others (31%), being more prudent, took the interview in order to obtain a valid score in case they apply for a NATO post in the next three years while sixteen per cent of them chose to take the interview simply because they wanted to check their level of speaking proficiency for personal reasons.

When asked how they prepared for the interview, sixteen officers (29%) answered that they did not prepare at all because they did not have any time, while fifteen officers had lessons with an English teacher. Eleven officers felt really confident and stated that they did not prepare because they did not need to, while from the remaining thirteen, nine officers practiced by conversing in English with relatives/friends/colleagues and four others practiced by conversing with English-speaking relatives/friends/colleagues.

Subsequently, I wanted to investigate the extent to which the officers were familiar with the purpose and content of the interview before they took it so as to know what to expect. Only seven officers claimed that they were clear about the purpose and content of the interview prior to taking it, while the majority of them, that is twenty-nine officers, stated, rather compromisingly, that they were “more or less” familiar with the purpose and content of the interview and the remaining nineteen officers answered that they did not know anything about the purpose and content of the interview before they took it. However, the surveyed officers proved to be aware of how test takers can find out about the purpose and content of the interview. When asked to indicate all the possible ways that test takers can obtain access to this kind of information, thirty-one respondents indicated that test takers can simply telephone the Military Language Testing Center and twenty respondents indicated that test takers can drop by the Military Language Testing Center and learn everything they need to know about the interview. Twenty-five respondents also stated that test takers can refer to test specifications (the official document issued by the Hellenic National Defence General Staff). However, there are still twelve of the surveyed officers who seem not to be aware of any of the above-mentioned ways.

Reflecting on their experience of taking the interview, a large majority of the respondents (75%) characterized it as “positive” while the rest of them answered that they have “mixed feelings” about it (nobody opted for “negative”). Regarding their levels of anxiety before the interview, twenty-four officers answered that they felt slightly anxious, twenty-one did not feel any anxiety while ten officers reported that they felt anxious before the interview. The officers’ answers to the next question indicate that in general their levels of anxiety did not increase during the interview (62% answered “no”). However, thirteen officers opted for “slightly” and eight thought that their levels of anxiety increased during the interview.

When asked to comment on the presence of more than two examiners during the interview, twenty-four officers commented that it affects the interview positively, thirteen stated that it affects the interview negatively, the same number of officers (13) answered that it does not affect the interview in any way, while five officers did not have an opinion on it.

A clear majority (84%) of the surveyed officers found the number of questions that they were asked during the interview “sufficient” and, thus, they were able to “fully demonstrate their oral proficiency”. Only five officers had the opposite opinion while four were ambivalent and opted for “not sure”. Fifty respondents agreed that the questions were clear and only a mere handful of them (five military officers) suggested that “not all of them” were clear. Similarly, a high percentage (80%) indicated that they were given “enough time to respond to the questions and elaborate on [their] thoughts and ideas”. Eight respondents opted for “quite” suggesting that the time they were given was almost enough, while there were three of them who believed that they were not always given enough time to respond to the questions and elaborate on their thoughts and ideas. Nobody thought that the time they were given was not enough.

The next question sought to investigate the officers' perceptions on the language functions elicited by the interview. It employed the main language functions included in the STANAG (Edition 4) Speaking Descriptors at Level 3 (which were adapted accordingly) and asked participants to state whether the tasks that they were engaged in during the interview elicited the indicated eight language functions. To this end, a rating scale (*Yes-No-Probably-Don't Remember*) was used to facilitate the officers' answers. Four functions received very high ratings. More specifically, the respondents indicated that during the interview they had the opportunity to *support their opinion and justify their decisions* (95%), *participate in a discussion on a social and/or professional topic* (86%), *answer objections and clarify points raised by their interlocutors* (78%), and *hypothesize and express their opinion on a situation* (75%). Furthermore, sixty per cent, that is thirty-three officers, answered that during the interview, they had the opportunity to *respond to challenges from their interlocutors*. Twenty-five of the respondents were of the view that during the interview they also had the opportunity to *defend and state policy*. However, fifteen had a different opinion and opted for "no", eight could not decide and opted for "probably" and seven did "not remember". With regard to *eliciting information and opinion from their interlocutors*, a meager majority, twenty-one officers, opted for "yes", while fifteen officers answered "no", eleven chose "probably" and eight did "not remember". Finally, the respondents were torn on whether they were given the opportunity to *deliver a monologue* during the interview; nineteen stated they were, while some other nineteen officers thought they were not, twelve opted for "probably" and five did "not remember".

Regarding the duration of the interview, the officers widely agree (82%) that it was as long as it should be (six thought that "it was rather long", two thought that "it was rather short", one opted for "it was too long" and another one for "it was too short"). In addition, most officers (30) found the interview to be "quite relevant" to the language requirements and communication needs in the NATO work environment and ten officers thought of it as being "very relevant". However, fourteen officers were of the view that it was "not very relevant" (only one opted for "not relevant at all").

When asked to predict how they would do if they retook the interview, most officers (29) stated that they would do better, eighteen indicated that they would do the same, eight were undecided, while nobody opted for "worse". With regard to their score on the interview, the majority of the respondents (66%) believed that it was what they had expected to get, followed by thirty-one per cent who reported that it was lower than what they had expected to get and four per cent who thought that their score was actually higher than what they had expected to get. In this context, twenty-nine officers indicated that their score on the interview probably reflects their performance on it while sixteen officers stated that their score does not reflect their performance (and opted for "yes"). On the other hand, six officers hold the view that their score probably does not reflect their performance while four officers maintain that it does not (and opted for "no"). Similarly, twenty-seven officers believe that their score on the interview probably reflects their overall speaking proficiency. However, the numbers resulting from the dichotomous yes/no answers are reversed here compared to the ones in the previous question. In other words, the number of officers who believe that their score does not reflect their overall speaking proficiency (and opted for "no") has increased (13) while the number of those who support that it does (and opted for "yes") has decreased (9). The number of officers who chose "probably no" has remained the same (6).

Subsequently, the officers were provided with a list of six language functions and were asked to rate them in terms of their importance for officers working in NATO. These language functions are part of the STANAG (Edition 4) Speaking Descriptors at Level 3. However, this was not stated in the questionnaire. The officers were asked to rate the given language functions using a *Very Important – Quite Important – Not Important – Not Sure – Don't Know* scale. More specifically, the great majority of respondents (75%) indicated that it is “very important” for officers working in NATO to be able *to participate effectively in formal and informal conversations on practical, social, and professional topics*. Regarding officers’ ability *to discuss particular interests and special fields of competence with considerable ease*, the respondents were torn between “very important” and “quite important”; forty-seven per cent opted for the former and forty-nine per cent chose the latter. A very high percentage (78%) believes that it is “very important” for officers working in NATO to be able *to use the language to perform a series of common professional tasks, such as answering objections, clarifying points, justifying decisions, responding to challenges, supporting opinion, stating and defending policy*. Furthermore, sixty-nine per cent rated officers’ ability *to demonstrate language competence when conducting meetings, delivering briefings or other extended and elaborate monologues, hypothesizing, and dealing with unfamiliar subjects and situations* as being “very important”, while twenty-six per cent opted for “quite important”. Fifty-eight per cent think that it is “very important” for officers working in NATO to be able *to reliably elicit information and informed opinion from native speakers or from highly competent English language speakers* (thirty-six per cent opted for “quite important”). Finally, regarding officers’ ability *to convey concepts in discussions of such topics as economics, culture, science, technology, philosophy as well as his/her professional field*, forty-seven per cent think of it as being “quite important”, thirty-eight per cent consider it to be “very important”, and nine per cent of the respondents, that is five respondents, dismiss it as being “not important” (the only time that the “not important” option was chosen by more than one respondent and the only time that the “don’t know” option was ever chosen by any of the respondents in this question – here two respondents opted for it). It is feared that the reference to *culture* and *philosophy* must have somehow discouraged respondents from giving higher ratings to the importance that they actually attach to this particular language function.

The last question asked participants to put forward their suggestions for improving the quality of oral assessment at the Military Language Testing Center. Only twenty-five officers expressed their opinions on the matter. A small number of the answers were deemed to be irrelevant to the purpose of the question as they addressed issues relevant to other sections of the examination or simply offered positive remarks on the existing model of oral assessment. The remaining suggestions essentially addressed issues relating to interview topics and tasks. Regarding the former, it was suggested that additional (but not too specific) focus be given to NATO-oriented topics, such as policy, operations, communications, intelligence, or other military-related topics. In a similar vein, some respondents called for more questions on a wider variety of topics (without mentioning what kind of topics these should be). Regarding the latter, I was pleasantly surprised to find out that a number of suggestions concerned the inclusion of role plays as part of the interview (role play ideas/scenarios were also provided). Other suggestions put forward involved the use of discussion prompts in the form of a short newspaper/magazine article or a paragraph from a book and the creation of a pool/list of discussion topics from which the examinee will be given the opportunity to choose the one s/he will feel most confident to talk about. The need for candidates to become more familiar with

the speaking descriptors, which will in turn foster a better understanding of the requirements of the interview, was also mentioned while it was pointed out that the official document issued by the Hellenic National General Staff, which is the test specification document in effect, should be more analytical and contain more details on the objectives and grading criteria of the interview. There were also occasional comments on the presence of too many examiners in the room during the interview (characterized as “stressful”, “distracting”, even “inconvenient” as the examinees have to “turn their head from one direction to another all the time”). In fact, one of the respondents went so far as to suggest that the furniture in the examination room be rearranged in such a way so that the examiners are seated closer to each other “in a more compact scheme”. Finally, another interesting point was put forward by an officer, who, commenting on the dynamics of the interview and the status of the participants, underlined that “the examiners should not be biased by their military rank but should judge the examinees solely on their language skills and not their personal opinions on the topics discussed during the interview”.

INTERVIEW RESULTS

All five officers who took part in the follow-up interviews have been awarded a Level 3 in Speaking after taking the interview as part of the English language examination at the Hellenic Foreign Language Testing Center during the academic year 2015-2016. Furthermore, three of them hold a C1-level certificate and the other two hold a C2-level certificate and they are all male.

The interviews were conducted in English. The interview guide consisted of ten questions. All ten questions were asked during each interview but not in a fixed order. The purpose of the interview was briefly stated before each interview began. Below the ten questions are presented in the order in which they appear in the interview guide.

1. What is your overall opinion on oral language assessment in the Hellenic Armed Forces?

As an introduction to the main body of the interview questions, the very first question asked for the interviewees’ opinion on oral language assessment in the Hellenic Armed Forces. According to two of them, the fact that there is a language assessment mechanism in place which can effectively assess officers’ oral language skills prior to any NATO posting using the same standards and criteria set by the STANAG is really important because in this way it is ensured that the officers posted abroad meet the oral English language requirements for these posts. It was also indicated by a third officer that oral language assessment in the Hellenic Armed Forces has improved significantly over the last few years and has become “competitive” and “challenging” (judging from the context in which the adjectives were used, it is felt that the adjectives were meant to convey the interviewee’s positive feelings on the interview and emphasize the high level of oral English language competency that test takers need to exhibit in order to be certified as Level 3 speakers) while another one also mentioned that it is “challenging” and “reflects the difficulties that one may encounter during his/her deployment”. However, one of the officers that I interviewed took a rather critical stance on the issue and stated that oral assessment is “at its early stages” and “there are some things that need to be adjusted in order to make sure that the officers who are selected to be posted abroad have a very good level of English”.

2. What is your overall opinion on the interview?

The interviewees' overall opinion on the interview can be characterized as positive. Subsequently, they were asked to mention some of the positive and negative aspects of the way the interview is conducted. Not all officers offered comments on both aspects as asked. On the positive side, it was mentioned that the interview is conducted "in a friendly atmosphere which makes candidates feel comfortable", "gives candidates the opportunity to show what they know", the questions were "accurate" and "well-aimed" and "the examiners are well-prepared and well-equipped to meet the challenges of the various English language levels they may come across". On the minus side, an interviewee pointed out that the presence of so many examiners, who are, in their majority, officers, can "increase candidates' levels of anxiety" and can be "a bit awkward because other officers are present". After being probed to reflect on the issue, he went on to say that the officers-examiners can in fact be viewed as "an audience" simulating a typical NATO work environment where officers need to be able to express themselves, support their opinions, etc. in the presence of others. In this context, he added that the experience from the interview can actually help officers "overcome stage fright" and give them an idea of what it feels like having to "answer demanding questions in English asked by different people". Finally, the officer who expressed more moderate views on the previous issue suggested that the interview become "more standardized" as this "would make the examination fairer and bring out the most realistic language knowledge of each candidate". When asked to explain how in his view the interview can become "more standardized", he commented that "all candidates should be given topics of the same level". I briefly pointed out to him that during the interview candidates are asked questions of different difficulty in order to establish and confirm their level so if a candidate cannot handle, for example, Level 3 topics and questions, then s/he is taken back to Level 2. Subsequently, I asked him to tell me how fair he thought the interview as a means of assessing officers' speaking proficiency was and he replied "quite fair but there is always room for improvement".

3. In your opinion, is the interview relevant to the language requirements in NATO?

In general, the interview was deemed to be relevant to the language requirements in NATO as it basically "assesses the necessary language functions that officers should be able to perform in NATO", with the exception of one officer, the one who had previously pointed out that the interview needs to be further standardized, who stated that he is "not entirely convinced that [the interview] is very relevant" to the language needs in NATO. When asked to elaborate on the matter, he rushed to respond that "This is just how I feel about it!", adding, after a short pause, that "more research into the language needs of officers who are not native speakers of English needs to be done".

4. Do you think that the interview should be oriented towards more military-related issues?

All five officers unanimously agreed that "there is no need to include more military topics, given the interview time limits" as these are already "sufficiently covered". In fact, one of the interviewees drew attention to the fact that we must be very careful what kind of military issues we choose to talk about in the context of an oral language examination because "quite

often military issues have political implications and, without realizing it, we may end up having a political conversation instead”.

5. Does the interview effectively assess speaking proficiency at Level 3?

Four officers responded positively to this question. In an attempt to summarize their views on the issue, it can be deduced that these officers are of the opinion that the interview effectively assess speaking proficiency at Level 3 because “it assesses the necessary language functions that officers should be able to perform in NATO ... these language functions are in accordance with STANAG”. The fifth officer stated that “probably it does” adding that “further standardization will help towards this direction” and referred back to his answers to questions 2 and 3.

6. Do officers need to be familiar with the STANAG Level 3 Speaking Descriptors?

There is a general consensus among the interviewees that officers should be familiar with the STANAG Level 3 Speaking Descriptor prior to taking the interview, though their answers exhibited varying degrees of enthusiasm (“it would be helpful”, “they can help them understand the requirements of the examination”, “they definitely need to know about them before they take the examination”, “absolutely ... in order to be adequately prepared ...”). Furthermore, it was pointed out that the STANAG is easily accessed, so “if someone is really motivated, it is really easy to find out what the speaking descriptors are about”. When asked if they had read through the speaking descriptors before they took the interview, three of them said they had.

7. In your opinion, what kind of language functions should a speaker of English at Level 3 be able to perform?

The following list of language functions was compiled from the officers’ answers:

- To be able to have conversations about social and professional issues
- To be able to give presentations
- To be able to talk about issues on an academic level
- To be able to comprehend nuances and important notions
- To be able to communicate and carry out a conversation with native and non-native speakers of English
- To be able to have a conversation about abstract topics and elaborate on his/her ideas
- To be able to talk about general and military issues fluently and confidently and with grammatical and syntactic accuracy
- To be able to talk about professional issues in detail and in depth
- To be able to talk about political, economic and social matters fluently and without mistakes
- To be able to argue in a civilized manner and convince the audience with logical and well-developed arguments
- To be able to use military terminology correctly

- To be able to express his/her opinion on his/her country's policy and interests
- To be able to express himself/herself accurately, especially under stressful conditions

8. In your opinion, do the STANAG Level 3 Speaking Descriptors reflect the language requirements in a NATO work environment?

All five officers unequivocally agreed that the STANAG Level 3 Speaking Descriptors do reflect the language requirements in a NATO work environment. In general, it is felt that “the descriptors are quite indicative of the language functions that officers should be able to perform at Level 3”. However, one of the officers reaffirmed his view that “more research into the language needs of officers who are not native speakers of English needs to be done”, a point made earlier (see Question 3), in order to have more realistic expectations of them.

9. Do you think that you are a Level 3 speaker?

Based on their scores on the interview and the way they perceive themselves as speakers of English, all five officers indicated their firm belief that they are Level 3 speakers. In fact, three of them made reference to their previous NATO-related work experience as further evidence for rightly considering themselves to be Level 3 speakers since they had “no difficulty using and understanding English” during their deployments.

10. Do you think that you meet the language requirements for NATO posts requiring Level 3 in Speaking?

All five officers promptly stated that they firmly believe that they meet the language requirements for NATO posts requiring Level 3 in Speaking and declared their readiness to take up relevant posts.

PART FIVE: DISCUSSION

SETTING THE CONTEXT OF THE DISCUSSION

At first glance, two kinds of biases can be easily detected in the officer sample. More specifically, the officers who participated in the online survey come from the three military branches of the Hellenic Armed Forces but their number is not even. The exceptionally low number of Army officers has already been accounted for in the previous chapter. However, the difference in the number of the Navy and Air Force officers who completed the questionnaire (there are more than twice as many Navy officers as there are Air Force officers in the sample although the initial email was sent out to practically the same number of officers from the respective military branches) cannot be explained in any straightforward manner; it can only be assumed that the higher number of Navy officers can be attributed to the fact that this study is conducted by a colleague of theirs in the Navy, so they must have felt a more direct connection with it and a stronger obligation to contribute to it. It should also be pointed out that a significant part of this study was conducted during the summer months which made it even more difficult to track potential participants down either because they were on summer leave or transferred to a new post or department. Finally, regarding the gender of the respondents, the

sample is obviously biased. In a (not so) male-dominated field such as the military, this kind of bias can be perceived as “understandable” by many. The truth, however, lies somewhere in the middle; the majority of female officers have not yet reached the military ranks which would make them eligible for a NATO post (the situation is expected to change in the next few years) while for those (mainly Supply/Logistics officers) who already hold the required military ranks, the number of respective posts is already significantly limited (irrespective of gender). As for the reasons why female officers might not decide to apply for a post abroad as easily as their male colleagues usually do, these go beyond the scope of this study.

Undoubtedly, the officers’ opinions on the interview have been shaped by their education background (including their English language proficiency), their previous test taking experience (of English and Greek examinations) as well as their work and life experience. A great number of the participants hold graduate degrees or have received graduate level education and training either in Greece or abroad in an English-speaking environment. Similarly, the majority of them are highly competent in English as they hold a C2-level certificate. However, the fact that even a handful of them stated that they hold a B1-level certificate or do not hold an English language certificate at all is problematic in itself and raises a number of questions since the basic language requirement for all candidates, which is stated explicitly in the official document issued by the Hellenic National Defence General Staff, is that they should hold a B2-level certification in order to be eligible to take the English language examination at the Military Language Testing Center. A final observation concerning the participants’ background is that many of them have previously worked in NATO or in NATO-related environments, so it is assumed that they know firsthand the true language and communication needs in a NATO work environment.

In this context, the officers have been called upon to give their opinions and make their judgments about a number of interview-related issues despite the fact that they lack any subject-matter expertise. No matter how enlightening and informative these judgments can prove to be, they must be treated cautiously before any final conclusions are drawn as they are likely to be subjective and carry emotional overtones.

Given all the above, I now turn to the discussion of the findings in relation to the three research questions.

RESEARCH FINDINGS VIS-À-VIS RESEARCH QUESTIONS

1. What do test takers think of the content of the interview and the way it is conducted?

The respondents’ views on the content of the interview and the way it was conducted were indicated mainly through their answers to the following questionnaire items:

- Were you familiar with the content and purpose of the interview before you took it? (Question 12)
- How would you characterize your overall experience of taking the interview? (Question 13)
- Did you feel anxious before the interview? (Question 14)
- Did your anxiety increase during the interview? (Question 15)

- What do you think of the presence of more than two examiners in the room during the interview? (Question 16)
- Were you asked a sufficient number of questions which enabled you to fully demonstrate your oral proficiency? (Question 17)
- Were the instructions clear? (Question 18)
- Were you given enough time to respond to the questions and elaborate on your thoughts and ideas? (Question 19)

The majority of the respondents were superficially familiar with the purpose and content of the interview prior to taking it though it is generally accepted that information on the purpose and content of the interview can be easily accessed in more than one way (Question 11). It is surprising that quite a few officers not only were they completely unaware of the purpose and content of the interview before they took it but they also did not have a clue as to how this kind of information could be accessed. Brown (1993) urges that “[p]otential test takers need to be educated as to the purpose of the test if they are to accept it as a suitable assessment instrument” (p. 295) and suggests that “adequate information on the content and format of the test [be] provided before the test administration” (p. 296).

Overall, the interview is described as a test-taker friendly experience which does not seem to induce or add to test takers’ anxiety. This view is reinforced by the interviewees’ favorable comments on the way the interview is conducted (see Interview Question 2). In addition, the respondents hold the view that the interview allows test takers to fully demonstrate their oral proficiency and express and elaborate on their thoughts and ideas through a sufficient number of questions posed by the examiners in a clear and comprehensible way. It can be said then that officers consider the interview to have face validity. Building on Alderson, Clapham and Wall (1995), who acknowledge the importance of face validity in testing, it can be said that face validity is particularly important in military testing as it can enhance the credibility of the testing instrument in the eyes of military officers, who may not trust or have confidence in the test, and reduce potential face threat – after all, the concept of “face” (Brown & Levinson, 1987; Goffman, 1967; Thomas, 1995) is crucial in the military. As Alderson et al. (1995, p. 173) point out, “tests that do not appear to be valid to users may not be taken seriously for their given purpose”. On the other hand, “if test takers consider a test to be face valid, ... they are more likely to perform to the best of their ability on [it] ...”.

One of the issues that I have been particularly interested in investigating in the context of this study concerns the presence of more than two examiners during the interview and how it can potentially affect test takers’ performance. In most well-established oral examinations there are two examiners (e.g. an interlocutor and an assessor) present in the examination room or even one (in both capacities). At the Military Language Testing Center the test taker is confronted with four or five examiners who are usually military officers. Not all examiners are obliged to participate in the interview but they all have to give a mark in the form of a score for each candidate at the end of each interview session. According to the questionnaire findings, the majority of the respondents (though not a high one - 44%) indicated that the presence of more than two examiners affects the interview positively. Taking a closer look at the rest of the responses, it can be assumed that a number of officers chose to take a neutral stance on the issue either by indicating that it does not affect the interview in any way or by not offering an opinion on it. On the other hand, many of those who were of the view that it affects the interview negatively voiced their concerns quite intensely in their comments to the final open-

ended question. Undoubtedly, the officers-examiners bring to the interview their previous work experience and knowledge on military issues which can add authenticity to the tasks (Green & Wall, 2005) and give candidates a more realistic idea of possible language challenges in a NATO work environment (see officer's relevant comments to Interview Question 2 in the previous chapter). However, it can also be speculated that the respondents, who are officers and were asked to comment on the presence of other officers, did not want to appear disrespectful or challenge the status of the examination committee in the context of military chain of command.

In the context of the above discussion, it can be deduced that the officers who participated in this study have positive feelings both about the content of the interview and the way it is conducted.

2. Do test takers think that the interview can effectively assess their speaking proficiency in English for interoperability purposes at Level 3?

The majority of the respondents found the interview to be quite relevant to the language requirements and communication needs in the NATO work environment, while most of the interviewees hold the view that it effectively assesses speaking proficiency at Level 3. However, it would be interesting to find out why a number of the participants think that the interview is not very relevant to the language requirements and communication needs in the NATO work environment since none of them commented on it in any way in the space provided at the end of the questionnaire.

The language functions elicited by the interview (Question 20) have been shown to correspond to a large extent with some of the language functions which are part of the STANAG (Edition 4) Speaking Descriptors at Level 3 (Question 27) and which the respondents rated mainly as being "very important" and "quite important" for officers working in NATO. One of the proposed language functions included in Question 20, that of *delivering a monologue*, was shown not to be successfully elicited by the interview as indicated by the participants' answers. It can only be assumed that the word *monologue*, though of Greek origin, must have confused the respondents and consequently must have been misinterpreted as *delivering a monologue* is in essence giving an extended answer in which you *support your opinion* in the context of a *discussion on a social or professional issue* given the time constraints of the examination, two language functions which are claimed to be successfully elicited by the interview as the questionnaire findings indicate a high agreement among respondents. With regard to *stating and defending policy*, a central concept in the definition of interoperability, it is felt that it is probably not as adequately elicited by the interview as it should be because the respondents appear to be somehow divided (despite the fact that there is a small majority who responded positively on the issue). In fact, one of the comments to the open-ended question addressed this issue by calling for additional focus on NATO-oriented topics, such as policy, operations, communications, intelligence, or other military-related topics. However, the interviewees unanimously agreed that military-related topics are sufficiently covered in the interview which typically should give candidates the opportunity to state and defend policy. Finally, the fact that the respondents appear to be largely undecided as to whether the interview did or did not give them the opportunity to *elicit information and opinion from their interlocutor(s)* can be attributed to the absence of role plays, which would have allowed them more easily to have this kind of exchange. In addition, a number of

respondents have suggested that role plays be included in the interview. It becomes apparent then that role plays can significantly help improve the assessment of military officers' speaking proficiency in English for interoperability purposes at Level 3 by making the interview more standardized and the assessment possibly more accurate and fairer.

One final point should be made regarding the familiarity of test takers with the STANAG Level 3 Speaking Descriptors and how this can affect their perceptions of how effectively the interview can assess their speaking proficiency in English for interoperability purposes. As it was pointed out in the previous chapter, the surveyed officers seem to have a rather intuitive knowledge of what the Speaking Descriptors are about while some of them stated that they know nothing about them. In fact, very few of the surveyed officers appear to be truly familiar with the Speaking Descriptors although there seems to be a shared belief among officers that being more knowledgeable about the Speaking Descriptors prior to taking the interview can help candidates gain a better understanding of its requirements which can in turn benefit their performance on it. In addition, knowledge of the Speaking Descriptors can help officers develop a better understanding of the language requirements in the NATO work environment as it has been pointed out by the interviewees. Regardless of their extent of familiarity with the Speaking Descriptors at Level 3, the officers have in general acknowledged the importance of the language functions included in them (Question 27). In fact, the interviewees came up with a number of language functions which reflect the language functions included in the Speaking Descriptors quite accurately. It can be assumed that the officers' work experience, professional knowledge and language background have helped them develop an implicit understanding of what the Speaking Descriptors are actually about.

In the context of the above discussion, it can be deduced that the officers who participated in the study think that the interview can satisfactorily assess their speaking proficiency in English for interoperability purposes at Level 3.

3. Do test takers think that their scores on the interview reflect their linguistic readiness for NATO posts requiring Level 3 in Speaking?

Commenting on their scores on the interview, the majority of the surveyed officers stated that they actually got what they expected to get. Since officers are required to be Level 3 speakers in order to be eligible for any international posting, it can be deduced that the majority of the surveyed officers managed to achieve Level 3 in Speaking. This is not very surprising taking into consideration the high level of language proficiency of the majority of the surveyed officers (Question 6). In fact, the number of the officers who hold C2- and C1-level certificates is roughly equivalent to the number of the officers who stated that their score on the interview was what they expected to get (Question 24). In a similar vein, the majority of the surveyed officers were of the view that their scores on the interview probably reflected their performance on it (Question 25), followed by those who were positive about it. In other words, most of the respondents believe that they probably deserved the score they received on the interview. These findings seem to reflect to a certain extent the officers' answers to the previous question. In addition, the majority of the surveyed officers indicated that their score on the interview probably reflects their overall speaking proficiency (Question 26), followed by those who had a negative opinion on it. A few things are worth noting here. To begin with, the fact that the majority of the respondents (49%) opted for "probably yes" in Question 26 somehow contradicts the overwhelming majority of the respondents (84%) who unequivocally stated in

Question 17 that during the interview they were asked a sufficient number of questions which fully enabled them to demonstrate their oral proficiency. Regarding Question 25, the total number of officers who opted for “probably yes” and “yes” is equivalent to the number of officers who responded positively (and opted for “yes”) to Question 17. However, the total number of officers who opted for “probably yes” and “yes” in Question 26 is much lower than the respective number in Question 25 (and consequently in Question 17), not to mention the fact that the second most popular answer to Question 26 was “no” as opposed to the second most popular answer to Question 25 which was “yes”. It can only be assumed that Questions 25 and 26 being quite similar in their content (and probably in the way in which they were expressed) must have somehow confused some of the respondents, who might have also grown fatigued. According to Dörnyei and Taguchi (2010, p. 9), fatigue, known as the *fatigue effect*, is “more likely to influence responses toward the end of the questionnaire”.

In the context of the above discussion, it can be deduced that the majority of the officers who participated in this study think that their scores on the interview reflect their linguistic readiness for NATO posts requiring Level 3 in Speaking.

CONCLUSIONS

I believe that this study has shown that when test takers are given the opportunity to voice their opinions and concerns about language assessment practices, they are in fact able to make some really insightful and perceptive comments despite the fact that they lack subject matter expertise or at times may come across as rather emotional or too confident about their true abilities.

In an attempt to demystify the construct of the oral proficiency interview which is taken by all officers of the Hellenic Armed Forces prior to any staff appointment within NATO, a qualitatively-oriented study was conducted which focused exclusively on test takers’ perceptions of the issue under examination. The qualitative orientation of the study was dictated by the fact that this is the first exploratory research project in the field of language testing in the Hellenic Armed Forces which also put the test takers of this particular professional group at the heart of it. In this context, a qualitative approach to data analysis was adopted because it was deemed to lend itself to the purposes of this study as it is inherently language-based and helps the researcher to establish patterns (Dörnyei, 2007) by “tap[ping] mental processes that may be suggestive of some particular characteristics not easily detected in a purely quantitative study” (Winke, 2012). In addition, the qualitative orientation was further dictated by the fact that the collected data were largely biased and for this reason no correlations could have been carried out; almost all the respondents were male, the number of officers of the three military branches was disproportionate, and the number of C2-level certificate holders was disproportionately higher than the rest (C1, B2).

The discussion of the research findings indicated that the research questions, and by extension the validation questions which map onto the respective inferences in Shaw et al.’s (2012) framework, were answered quite satisfactorily. However, this should not lead those involved in the administration of the interview to be complacent since “there is always room for improvement”, as one of the military officers so rightly pointed out.

It was established that test takers need to grasp a better understanding of the Speaking Descriptors which will reinforce their existing intuitive knowledge of them. By becoming more familiar not only with Level 3 but also with Level 2 and Level 4 Speaking Descriptors, test takers will be able to compare and contrast the development of speaking proficiency per level and form a clearer and more realistic picture of the interview language requirements as well as of their own speaking abilities. Crossey (2005) calls it an understanding of what represents a “professional” level of foreign language knowledge. To this end, online tutorials can prove useful, effective and test taker friendly means of instruction.

Regarding the interview structure, roles plays should be adopted as level checks and probes. In this way, the interview will become more standardized and will be in full accordance with the Oral Proficiency Interview (OPI) structure endorsed by the BILC. Moreover, it is hoped that a wider range of language functions as described in the STANAG will be elicited through role plays. For this reason, it is important that a list of possible role play topics and scenarios be compiled so as to be subsequently trialed before they are incorporated in the interviews. It is also important that all members of the examination committee are familiar with conducting role plays with a view to eliciting ratable speech samples and have received appropriate training.

Sireci (2007, p. 477) points out that “[e]valuating test validity is not a static, one-time event; it is a continuous process”. It is therefore imperative that this work be followed up and its scope be expanded to include both qualitative and quantitative studies from different perspectives employing different data collection methods. Future follow-up studies can address the “Scoring” and “Decision-making” inferences in Shaw et al.’s framework (2012) which were not addressed in the context of the present study or look into interrater reliability, an issue which calls for careful and special consideration given the number of examiners who conduct and score the interviews.

In addition, this study has attempted to raise military officers’ awareness of the importance of language assessment in the NATO context and call their attention to the importance of linguistic interoperability as within the NATO there has been “a perceived dominance of English language native speakers in decision-making in its command structures” (Crossey, 2009, p. 151). As a result, “non-native speakers may effectively find themselves cut out of key processes” which can undermine their sense of self-worth and reduce the influence of national delegations (Crossey, 2005). This has obvious political implications as the NATO aims at fostering cooperation between nations, not creating, even unwittingly, an impression of “cultural discrimination”, Crossey suggests, who draws attention to the fairness and transparency of national language testing practices and systems.

For all the above, directly or indirectly related to this project, I hope that this study, besides paving the way for future research, will help to enhance the status of language testing in the Hellenic Armed Forces and lay the basis for more active involvement and collaboration of all interested parties in the implementation of military language testing policy.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Jayanti Banerjee for her guidance and advice. I am also particularly grateful to the officers of the Hellenic Armed Forces who participated in my research and helped me bring this project to fruition.

REFERENCES

- Adubato, M., & Efthymiopoulos, M. P. (2014). Capacity language building in NATO: language skill as a crucial tool for interoperability at all levels. *Transforming joint air power: The journal of the JAPCC*, 19, 26-30. Retrieved December 10, 2015, from https://www.japcc.org/wp-content/uploads/JAPCC_Journal_Ed-19_web.pdf
- Alderson, J. C., & Bachman, L. F. (2004). Series editors' preface. In Luoma, S., *Assessing speaking*. Cambridge: Cambridge University Press.
- Alderson, J. C., & Clapham, C. (1992). Applied linguistics and language testing: A case study for the ELTS test. *Applied Linguistics*, 13(2), 149-167. doi: 10.1093/applin/13.2.149
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Bachman, L. F. (1988). Problems in examining the validity of the ACTFL Oral Proficiency Interview. *Studies in Second Language Acquisition*, 10(2), 149-164. doi: 10.1017/S0272263100007282
- Banerjee, J. (2004). Qualitative analysis methods. In *Reference supplement to the preliminary pilot version of the manual for relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment* (pp. 1-53). Strasbourg: Language Policy Division, Council of Europe.
- Berger, A. (2015). *Validating analytic rating scales: a multi-method approach to scaling descriptors for assessing academic speaking*. Frankfurt am Main: Peter Lang.
- Brown, A. (1993). The role of test-taker feedback in the test development process: test-takers' reactions to a tape-mediated test of proficiency in spoken Japanese. *Language Testing*, 10(3), 277-301. doi: 10.1177/026553229301000305
- Brown, P., & Levinson, S. C. (1987). *Politeness: some universals in language use*. Cambridge: Cambridge University Press.
- Chapelle, C. A. (2008). The TOEFL validity argument. In C. A. Chapelle, M. K. Enright, & J. Jamieson (Eds.), *Building a validity argument for the test of English as a foreign language* (pp. 319-352). New York; London: Routledge.
- Crossey, M. (2005). Improving linguistic interoperability. *NATO Review*. Retrieved December 15, 2015, from <http://www.nato.int/docu/review/2005/issue2/english/art4.html>
- Crossey, M. (2009). The role of micropolitics in multinational, high-stakes language assessment systems. In J. C. Alderson (Ed.), *The politics of language education: individuals and institutions* (pp. 147-165). Bristol, UK; Buffalo NY: Multilingual Matters.

- Cumming, A. (2004). Broadening, deepening, and consolidating. *Language Assessment Quarterly*, 1(1), 5-18. doi: 10.1207/s15434311laq0101_2
- Dörnyei, Z. (2007). *Research methods in applied linguistics: quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
- Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: construction, administration, and processing* (2nd ed.). Abingdon: Routledge.
- Dubeau, J. J. (2006). *Are we all on the same page? An exploratory study of OPI ratings across NATO countries using the NATO STANAG 6001 scale*. Unpublished MA thesis, Carleton University, Ottawa, Ontario, Canada. Retrieved October, 25, 2015, from <http://www.natobilc.org/documents/Projects/OPIRaterSTUDY.pdf>
- Gillham, B. (2008). *Developing a questionnaire* (2nd ed.). London: Continuum.
- Ginther, A. (2012). Assessment of speaking. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Blackwell Publishing Ltd. doi: 10.1002/9781405198431.wbeal0052
- Goffman, E. (1967). *Interactional ritual: Essays on face-to-face behavior*. Chicago: Aldine Publishing Company.
- Green, R., & Wall, D. (2005). Language testing in the military: problems, politics and progress. *Language Testing*, 22(3), 379-398. doi: 10.1191/0265532205lt314oa
- Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 17-64). Westport, Ct: Praeger Publishers.
- Kunnan, A. J. (2012). High-stakes language testing. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Blackwell Publishing Ltd. doi: 10.1002/9781405198431.wbeal0054
- Rea-Dickins, P. (1997). So, why do we need relationships with stakeholders in language testing? A view from the UK. *Language Testing*, 14(3), 304-314. doi: 10.1177/026553229701400307
- Shaw, S., Crisp, V., & Johnson, N. (2012). A framework for evidencing assessment validity in large-scale, high-stakes international examinations. *Assessment in Education: Principles, Policy & Practice*, 19(2), 159-176. doi: 10.1080/0969594X.2011.563356
- Shohamy, E. (2001a). Democratic assessment as an alternative. *Language Testing* 18(4), 373-391. doi: 10.1177/026553220101800404
- Shohamy, E. (2001b). *The power of tests: A critical perspective on the uses of language tests*. London: Pearson.
- Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher*, 36(8), 477-481. doi: 10.3102/0013189X07311609

- Standardization Agreement 6001 (Edition 4), (2010). *Language proficiency levels*. NATO Standardization Agency, Brussels, Belgium. Retrieved on February 12, 2015, from <http://www.natobilc.org/documents/LanguageTesting/NUST%206001%20NTG%20ED4.pdf>
- Thomas, J. (1995). *Meaning in interaction: an introduction to pragmatics*. London; New York: Longman.
- Wall, D., Clapham, C., & Alderson, J. C. (1994). Evaluating a placement test. *Language Testing*, 11(3), 321-344. doi: 10.1177/026553229401100305
- Winke, P. (2012). Rating oral language. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Blackwell Publishing Ltd. doi: 10.1002/9781405198431.wbeal0993