

## AN INVESTIGATION OF AFFECTIVE REACTIONS TO THE FIRST-TIME ADMINISTRATION OF AN ENGLISH ORAL ELICITED IMITATION TEST AND ORAL NARRATIVE TEST

JAKUB BIELAK<sup>1</sup>, ANNA MYSTKOWSKA-WIERTELAK<sup>2</sup>, MIROŚLAW PAWLAK<sup>3</sup>

### ABSTRACT

As some language tests may be more anxiety-provoking than others, anxiety, other affective reactions, and related perceptions evoked by the English oral elicited imitation test (EI), a sentence repetition task measuring the implicit knowledge of grammar in a way not resembling natural communication, were investigated during first-time administration by means of a 10-point rating scale and a thought-listing tool. Because anxiety and other emotions cannot be interpreted in absolute terms, the same reactions induced by a special type of an English oral narrative test (ON) were investigated for comparison. A quantitative and qualitative analysis revealed EI to be more anxiety-provoking than ON as it created considerably higher levels of tension and worry. The possible causes include the perception of EI as very difficult, the uncertainty and confusion generated by the oral nature of its instruction and stimuli, and lack of an openly declared focus. Careful administration of EI is recommended to reduce anxiety and unfavorable perceptions. Other, much less frequent affective reactions to the tests and perceptions included satisfaction, curiosity, excitement, hope, confusion, interest, boredom, uncertainty, and concentration.

Keywords: Anxiety; test anxiety; language anxiety; elicited imitation test; oral narrative test; test difficulty.

### 1. Introduction

Both positive and negative emotions experienced in the course of second/foreign language (L2) learning and use (e.g., Dewaele & MacIntyre 2014; Oxford 2015)

---

<sup>1</sup> Department of English Studies, Adam Mickiewicz University, Poznań/Kalisz; ul. Nowy Świat 28–30, 62–800 Kalisz, Poland; [kubabogu@amu.edu.pl](mailto:kubabogu@amu.edu.pl)

<sup>2</sup> Institute of English Studies, University of Wrocław; ul. Kuźnicza 22, 50–138 Wrocław, Poland; [anna.mystkowska-wiertelak@uwr.edu.pl](mailto:anna.mystkowska-wiertelak@uwr.edu.pl)

<sup>3</sup> Department of English Studies, Adam Mickiewicz University, Poznań/Kalisz; ul. Nowy Świat 28–30, 62–800 Kalisz, Poland; [pawlakmi@amu.edu.pl](mailto:pawlakmi@amu.edu.pl)

have often been studied by researchers. Although much less frequently, positive and negative emotions occurring during language tests have also been researched (e.g., Scott 1986; Kenyon & Malabonga 2001). The focus on emotions with reference to language use, learning, and testing is warranted given the importance of affect for human behavior (MacIntyre 2002). Among the negative emotions researched in relation to L2 acquisition, anxiety has been studied the most (Dewaele & MacIntyre 2014; Gkonou, Daubney & Dewaele 2017). This is not surprising when the stresses of using an L2, which are especially severe in evaluative situations, are taken into account.

Anxiety is a cognitive and emotional reaction to difficult and threatening enterprises which is usually viewed as negative. Test taking and L2 learning and use are two such undertakings (Shohamy 2001; Horwitz 2010). Tests frequently induce high levels of apprehension, which typically confounds test scores by lowering them (Zeidner 1998). Learning and using L2s also tend to provoke anxiety, which, again, has mainly deleterious consequences such as low language achievement and poor performance on language tasks (Horwitz, Horwitz & Cope 1986). Given these negative tendencies, it seems that taking a language test may be especially anxiety-provoking and may abound in negative consequences in the case of at least some individuals.

Both test anxiety (TA) and language anxiety (LA) are constructs especially relevant for the present study. Its purpose was to investigate and compare anxiety, and, additionally, other affective reactions and perceptions related to affect, induced by two English oral language tests not taken by the participants before, namely the oral elicited imitation test (EI) and what will be called the oral narrative test (ON). On the basis of the study conducted by Scott (1986), which found that anxiety was a single major emotional reaction to language tests, and LA literature in general (see the literature review section), we expected to encounter primarily anxiety among the affective reactions to the two tests. As the qualitative results of the study presented in the results section show, this was a largely accurate assumption.

EI, which typically requires the repetition of oral stimuli and is usually used to assess implicit knowledge of grammar (Erlam 2006), and, based on this, global speaking proficiency (Van Moere 2012), is well known in applied linguistics but is somewhat controversial, mainly due to its contrived, “artificial” character not resembling natural communication. In fact, we were interested primarily in the reactions to this test rather than ON, which was included mostly for comparison. Being similar to EI in many important respects such as the oral mode and the tapping of implicit grammatical knowledge, this test was intended as a kind of standard of comparison as it is not possible to say anything interesting about the levels of anxiety and other emotions in absolute terms. However, the findings concerning ON, a focused communication task used to elicit particular

grammatical structures in relatively spontaneous communication, may in and of themselves also be of interest to those using it (or some variation on it; see Section 3.2 for more detail on the test).

EI is currently rarely used in L2 pedagogy. However, the recent resurgence of interest in this test (Van Moere 2012; Bielak & Pawlak 2013; Gaillard 2014; Van Compernelle & Zhan 2014; Sarandi 2015; Yan et al. 2016; Campfield 2017; see Sections 2 and 3.2 for more detail on the test) suggests that its use in the L2 classroom may soon be much more common. Since the use of EI is very infrequent, students are likely to encounter it for the first time when a particular teacher decides to employ it. Our personal experience has shown that in English-as-a-foreign-language contexts the test is received unfavorably, with a mixture of apprehension and surprise, which is likely to make it unreliable (cf. Bielak & Pawlak 2013). A good way to investigate EI-induced anxiety as well as other affective reactions to and perceptions of this test seems to be their comparison with those occurring in response to another oral test which is also new to test takers. For this reason, the study compared the anxiety, other emotions and perceptions related to affect induced by the two tests when administered to learners for the first time. The primary difference between the tests, both of which tap the implicit knowledge of grammar, is their similarity or otherwise to natural communication, with EI greatly divorced from naturalistic conversation and ON much closer to it.

## 2. Literature review

In psychology, anxiety is regarded as one of the most pervasive reactions to stressful situations (Sarason 1986) and is therefore often taken as a near-synonym of stress (Lazarus 1991). It occurs when a threat to ego or self-esteem or one's existence is encountered. Common manifestations of anxiety are excessive worry, tension, irrelevant and chaotic thinking, and such physiological reactions as fast heartbeat or excessive perspiration (Zeidner 1998). Anxiety is often conceptualized in three different ways which give rise to three distinct constructs: trait, situation-specific, and state anxiety. Trait anxiety is a relatively stable tendency to become anxious in a range of different situations (Spielberger 1983). Situation-specific anxieties are similarly trait-like, but they make an individual likely to become anxious in certain kinds of situations only. These more specific anxieties arise as a result of repeated experience of anxiety in situations of one type, for example in language learning or test-taking situations. In fact, one example of a situation-specific anxiety relevant in the present context is TA, which makes one susceptible to experiencing apprehension and fear during test-taking. It is a relatively widespread performance anxiety related to the fear of failure which often results in poor test performance (Zeidner 1998). Test anxiety,

in addition to communication apprehension and fear of negative evaluation, was conceptualized by Horwitz, Horwitz and Cope (1986) as one of the correlates of foreign language anxiety (or LA), which is another situation-specific anxiety of relevance for this study. Language anxiety was defined as “a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning process” (Horwitz, Horwitz & Cope 1986: 128). In contrast to the trait-like anxieties linked with test-taking and learning foreign languages, which predict the occurrence of anxiety in certain circumstances, state anxiety is the actual experience of the cognitive and emotional symptoms of anxiety in a concrete situation.

Both TA and LA, being situation-specific traits, are usually measured by means of dedicated surveys such as Reactions to Tests (Sarason 1984) and the Foreign Language Classroom Anxiety Scale (Horwitz, Horwitz & Cope 1986), which provide “a good starting point to describe the initial momentum or set of appraisals that a learner brings to a new L2 situation” (Gregersen, MacIntyre & Meza 2014: 576). Sometimes, however, when researchers are interested in anxiety reactions occurring in the moment during specific activities, or over a relatively short period (Piniel & Csizér 2015), state anxiety is measured, with the assumption that the state reflects a situation-specific trait anxiety. For example, Huang and Hung (2013) measured state anxiety in a test taking situation and assumed it reflected (or operationalized) TA. Gregersen, MacIntyre and Meza (2014: 576), in turn, who measured state anxiety using the so-called idiodynamic method, suggested “that a state anxiety reaction in a language-related context be considered *state language anxiety* [emphasis original]”. Another important assumption that seems to be made when state anxiety is viewed as an incarnation of one of the situation-specific anxieties is that it is in a sense constitutive of them because only its repeated occurrences in a given situation type may lead to the origin of the specific anxieties. It seems then that Gregersen, MacIntyre and Meza (2014) were justified in dubbing the state anxiety they measured in a language use situation *language anxiety* inasmuch as they viewed the state measured as potentially giving rise to the trait of LA. As will be seen, for the present study we have in fact assumed a similar view, but we have been more inclusive in the sense of recognizing the potential occurrence of both TA and LA in the form of a state evoked by language use in an evaluative situation.

Generally, both TA and LA are thought to be negative influences in education. Numerous studies (e.g., Hembree 1988; Cassady & Johnson 2002) have shown that TA is inversely, albeit not strongly, related to test performance and course grades. In one study concerning high-stakes English tests, Cheng et al. (2014) revealed a complex set of relationships between TA, motivation, personal and social factors such as age, gender, test importance and test purpose, and test performance, with TA and motivation interacting with the personal and social

factors in predicting performance. Huang and Hung (2013), whose study involved low-stakes, unofficial English speaking tasks, found a negative relationship between TA measured by a state anxiety instrument and oral performance no matter whether the participants were given some background information to use in speaking or not. Similarly to TA, LA has been shown to correlate with moderate strength with measures of language achievement and performance (e.g., Horwitz, Horwitz & Cope 1986; MacIntyre & Gardner 1989, 1991; Woodrow 2006; Hewitt & Stephenson 2012; MacIntyre & Gregersen 2012; Teimouri, Goetze & Plonsky 2019). Other examples of the deleterious consequences of LA are language class procrastination (Gregersen & Horwitz 2002), perfectionism (Gregersen 2003), overstudying (Horwitz, Horwitz & Cope 1986), avoidance (Piechurska-Kuciel 2008) or dropping out of language study (Dewaele & Thirtle 2009). The debilitating effects of test and language anxiety are undesirable because they prevent students from demonstrating their full potential and are a confounding variable in language testing for research purposes.

The explanation for the existence of the negative outcomes of anxiety offered in the literature is that anxiety interferes with the cognitive efforts demanded by test-taking as well as language study and performance. In particular, especially the cognitive component of anxiety, that is, worry, self-rumination and other similar cognitions, may interfere with taking in the input, language processing, and output production by tying up language learners' and test takers' limited attentional and cognitive resources (MacIntyre & Gardner 1994). This is especially likely to happen during the performance of linguistic tasks which are perceived as difficult (Piechurska-Kuciel 2008).

Unfortunately, there are few studies investigating anxiety, other affective reactions and perceptions related to them induced by specific language tests or comparing such reactions in relation to various types of tests. Zeidner and Bensoussan (1988), for instance, compared students' attitudes towards written and oral English tests and found that written tests were favored as they were seen as more pleasant, valuable, fair, and less anxiety-inducing than oral tests, which, however, aroused more interest. Shohamy (1982) compared perceptions of the Hebrew cloze test and the Hebrew oral interview, finding that the attitudes towards the latter were more favorable, and that learners prefer tests resembling real-life performance and, predictably, such which create low anxiety. Madsen (1981) found the English oral interview to be the least anxiety-provoking out of a battery of tests, with the reading test occupying the opposite end of the spectrum.

The small number of studies and their often contradictory results are not conducive to formulating generalizations, which have nonetheless been made. Generally, such factors as the format of the test, its length, time limit, environment, novelty, validity perceptions, and test taker anxiety are thought to impact affective reactions to tests (Scott 1986). More specifically, it has been

asserted that new test formats trigger more anxiety than familiar ones (Young 1991) and that speaking triggers more LA than other skills (MacIntyre & Gardner 1991). It may be added that tests perceived as difficult provoke more anxiety than tests viewed as easy, which is an assertion based on research concerning tests in general rather than just language tests (e.g., Hong 1999; Bonaccio & Reeve 2010). Importantly, in addition to the estimates of the incidence of TA and LA (Horwitz 2016, for instance, claims that 30% to 40% of learners suffer from LA), both anecdotal evidence (Brown & Abeywickrama 2010: 1) and personal narratives (Shohamy 2007) by language assessment experts speak to the ubiquity of anxiety in relation to language tests, which warrants rigorous investigation of this negative emotional reaction.

EI has been used to investigate first and second language acquisition for several decades (see a more detailed description in Section 3.2. below). In a comprehensive review, Vinther (2002) suggests that there is now agreement that the test is useful if carefully administered. Most recently, it has been endorsed as a valid and reliable measure of implicit linguistic knowledge allowing a neat, “surgical” elicitation of specific grammatical features (Ellis 2005; Erlam 2006). In addition, very recent research shows that EI may be suitable for classroom applications (van Moere 2012; Yan et al. 2016; Campfield 2017) including dynamic assessment (Van Compernelle & Zhan 2014) and classroom placement (Gaillard 2014), which is why the use of EI may soon be on the increase. However, the usefulness (Bachman & Palmer 1996) of EI has often been questioned, for example, on account of the possibility that it taps short term memory rather than linguistic knowledge (validity problem); the difficulty in deciding whether comprehension or production is tested (validity problem); and the highly artificial way in which it elicits output (authenticity problem). There are also other problems, including the question of whether a learner’s ability to imitate a structure matches their ability to use it in spontaneous speech (e.g., Slobin & Welsh 1968; Naiman 1974) or the issue of whether ungrammatical sentences should be used as stimuli along grammatical ones. No research to date has investigated affective reactions to EI, let alone those displayed by Polish learners of English.

In view of this evident gap in the existing research, the purpose of this study was to investigate the levels and nature of state anxiety, other affective reactions and perceptions related to affect evoked by English EI in Polish learners and compare them to those induced by English ON. The research questions formulated for this study were:

1. To what extent are the levels of state anxiety induced by English EI and ON different?
2. What is the frequency and nature of anxiety reports, reports of other kinds of affect, and perceptions related to affect in relation to the two tests, and are there differences between EI and ON in these respects?

It should be stressed that anxiety was the affective reaction to the two English tests which we expected to occur the most often (cf. Scott 1986), but other affective responses were also expected, even if with lesser frequency and intensity.

### 3. Method

#### 3.1. Participants

The participants were 34 year-two English majors at a Polish university (25 females and 9 males; all were L1 speakers of Polish) who were members of two intact classes. Their mean age was 21.38 years, ranging from 20 to 30. On average, they had experienced 11.5 years of instruction in English, ranging from 5 to 16, which made their proficiency varied, with the majority at the B2 level of the Common European Framework of Reference (CEFR; Council of Europe 2001). This proficiency level supports the claim that the tests designed for the study matched the participants' L2 ability, despite being quite challenging (see Section 4.2. for details). Considering their experience with language testing, all had taken a variety of tests, mostly relatively short, written ones, customarily employed by teachers. In addition, they had taken at least two longer achievement tests administered in Poland at the end of junior and senior high school and one English-as-a-foreign-language test at the end of year one at university. The first one is a written test including listening comprehension, reading comprehension, and writing components as well as a variety of gap-filling and matching tasks. In addition to these, the other two tests also include an oral component involving a short presentation, acting out roles, discussion with examiners, and the like.

#### 3.2. The tests

EI used in the study included 10 stimulus sentences (see Appendix A) adapted from Erlam (2006), which focused on such English structures as articles, basic tenses, and prepositions, and were therefore within the range of the participants' ability. They had been read at natural speed and recorded by means of a voice recorder by one of the present researchers, who is not a native speaker but speaks with a relatively authentic North American accent. The average length of the stimuli was 19.70 syllables, ranging from 18 to 23. Stimuli length is one of the controversies surrounding EI. For example, Chaudron, Nguyen and Prior (2005) advise the use of stimuli of up to 10–12 syllables while Vinther (2002) reports successful use of sentences including 16–30 syllables. The length of approximately 20 syllables was deemed suitable for the present study because adult learners were involved and the sentences included relatively simple vocabulary and grammar. Half of the stimuli were ungrammatical, with the



expectation that, if truly processed rather than memorized for “parroting”, they might be corrected during reproduction. The sentences were played to the participants by means of a PC and a pair of small-sized speakers. The test instruction was given orally in Polish, and, following Erlam (2006), asked the participants to repeat the stimuli in correct English. To ensure that the test was reconstructive, that is, that the sentences were processed before repetition rather than “parroted”, the participants’ attention was diverted from the aural form of the sentences and the act of repetition for about 3 seconds. Specifically, they were informed that, in addition to being a language (English) test, it was also a survey of their views on a range of issues; therefore, after hearing every sentence, they were supposed to indicate in a special form whether they agreed with the statement or not, or were uncertain. Only after approximately 3 seconds were they allowed to repeat the stimulus at the researcher’s signal.

ON was a limited production response test which was at the same time an oral narrative test. The test was designed and successfully used by Pawlak (2006) to elicit the English passive during a communicative activity. Such tests belong to the category of focused communication tasks (Ellis 2003), which require the use of a specific structure to achieve communication goals. This kind of use is believed to be based on implicit knowledge, making ON similar to EI. The instruction was in Polish and required the creation of a short narrative about an American university and people associated with it on the basis of two complete sentences and 13 prompts (e.g., *found / the government / the university / in the 19th century*) with attention being paid to the use of a proper voice and tense (see Appendix B). The students were given 3 minutes to produce a narrative including sentences based on the prompts, some of which created obligatory contexts for passive voice while others for active voice distractors.

Even though the study did not focus on the test scores obtained by the participants, it should be stressed that the percentage scores (% of correct responses) on the two tests were comparable (EI:  $M = 47.9\%$ ,  $SD = 14.9\%$ ; ON:  $M = 42.2\%$ ,  $SD = 12.1\%$ ), which reflects the care taken to design them as equally challenging. Importantly, there was no statistically significant difference between the two sets of scores, and the distributions of the scores was also similar with no major skewing in either set.

### 3.3. Data collection and analysis

The data were collected by two instruments. The first one was a 10-point single-item rating scale (see Appendix C) measuring the participants’ level of state anxiety before and immediately after the administration of the test with the anchors 1 (*very relaxed*) and 10 (*very anxious*) (both given in Polish). We decided not to label state anxiety measured by this instrument as state TA or state LA as



it was impossible in our view to determine with full confidence the predominant cause/nature of this anxiety. Recognizing the fact that we measured state anxiety during a test on the one hand and in the course of using English as a foreign language on the other hand, we believe it is not possible to relate the state anxiety measured to only one of these situation-specific anxieties; both are very likely to be strongly implicated. It is obviously not possible to exclude other influences on our participants' state anxiety such as some personal or family issues experienced at the time of the study or general trait anxiety. They are however much less likely to have contributed to the state anxiety we measured in any significant way, especially if one takes into account the fact that in our analyses we used mean anxiety scores for the whole sample rather than individual scores, some of which may have reflected anxieties other than TA and LA to a certain degree. In sum, then, because the state anxiety we investigated was tapped immediately before and after oral English tests, it is very likely that both TA and LA were significantly implicated in bringing about the state which was measured. Also, even if the state was experienced by individuals characterized by low levels of TA and LA, it certainly had the potential of leading to the origin of TA and LA as it occurred in relation to language tests. For quantitative data collected in these ways, descriptive statistics and paired samples *t*-tests were calculated to determine the significance and magnitude of the differences in the levels of state anxiety induced by the two tests and by the same test at different points in time. The basic assumptions for paired samples *t*-tests were met as the anxiety scores were to a large extent normally distributed and there were no outliers.

The second tool was a think-aloud instrument often used in TA research called the *thought-listing technique*. The instructions, adapted from Blankstein, Toner and Flett (1989: 273) and translated into Polish, asked participants to list all the thoughts and feelings they were experiencing before, during and after the tests (see Appendix D). Since the instruction was in Polish, all participants responded in this language, which made the process smoother compared to answering in English, given their level of advancement in this L2 (all the subsequent quotations of participants' responses are translations from Polish by one of the present authors). The responses were given in writing, each time within a time limit of 2.5 minutes. The advantage of this kind of instrument is that it "does not artificially restrict a subject's range of responses" (Blankstein, Toner & Flett 1989: 284) and may thus reveal the aspects of affective reactions and related perceptions which would go unnoticed if data were gathered by means of more structured measures such as questionnaires. Qualitative data were content-analyzed by two of the present authors to identify the components of anxiety and other possible affective reactions, as well as perceptions related to them. The categories of anxiety, that is, tension, worry, test-irrelevant thinking, and bodily symptoms, were established following Sarason (1984). Tension refers to general

expressions of anxiety, stress, panic, fear, uneasiness, and so on. Worry is a cognitive component of anxiety and typically takes the form of thoughts of failure and unfavorable comparisons of one's abilities and those of others. Test-irrelevant thinking is related to the inability to focus on the task at hand when the mind wanders off to other matters. Bodily reactions involve such physiological reactions as sweating and trembling hands. All the differences between the results obtained by the two authors were discussed and in every case ultimate agreement as to the best interpretation of the data was reached. Additionally, the frequency with which various reactions occurred was calculated.

### 3.4. Procedure

A few weeks before data collection, informed that the test results would have no bearing on their course grades and that they would be given additional points contributing to their university-course credit, participants gave their consent to take part in the study. The data were collected during regularly scheduled classes on one day in two intact groups by two of the researchers, who were the participants' instructors. The two tests were taken individually one after another but were administered simultaneously in groups of 8 participants, each sitting at a separate desk. First, a short training session demonstrated the EI format to the participants. Subsequently, just before EI, they marked their level of anxiety on the rating scale and listed their current thoughts and feelings as elicited by the thought-listing tool in approximately 2.5 minutes, a time limit that was also used for subsequent occurrences of this activity. Next, they took the test, which lasted approximately 4 minutes, after which they again marked the level of their current anxiety and listed their current thoughts and feelings. This was followed by retrospective thought-listing in relation to the cognitions and emotions experienced during the test. The rating scale was not used in a retrospective fashion (like thought-listing) because it would have been difficult for the participants to recall the exact level of anxiety experienced during the test a few minutes after the actual experience. At the beginning of ON, administered right after EI, the participants read its instruction and examples in approximately 2 minutes and were encouraged to ask questions. Subsequently, they were given exactly 1 minute to scan the prompts to make sure that they were familiar with the vocabulary and ask questions about it. Then they took ON, with the marking of anxiety levels and thought-listing taking place in the same fashion and at the same times as in the case of EI.

## 4. Results

### 4.1. Rating scale

Table 1 displays the descriptive statistics for state anxiety measured with the rating scale before and after the tests. The levels of anxiety before the two tests differed by 0.42. Anxiety remained at a similar level after EI only, with a rise by 0.32. By contrast, after ON, the level was much lower than before, with a fall of 1.18. These trends are also showed graphically in Figure 1. The results of *t*-tests exploring the contrasts between anxiety induced by both tests and by the same test at different points in time are included in Table 2. While the modest difference between anxiety before the two tests was not significant, the one between anxiety after the tests was ( $t = 2.58, p = .014$ ), which suggests that the fall in anxiety after ON was substantial. This is confirmed by the absence of a significant contrast between anxiety before and after EI and the presence of a significant contrast between anxiety before and after ON ( $t = 2.31, p = .027$ ). For both significant differences, the effect sizes were close to but below the medium benchmark of 0.5 (Cohen's *d*).

Table 1. Descriptive statistics for state anxiety before and after EI and ON ( $n = 34$ )

Time	EI		ON	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Before test	4.93	2.10	5.35	2.22
After test	5.26	2.59	4.17	2.12

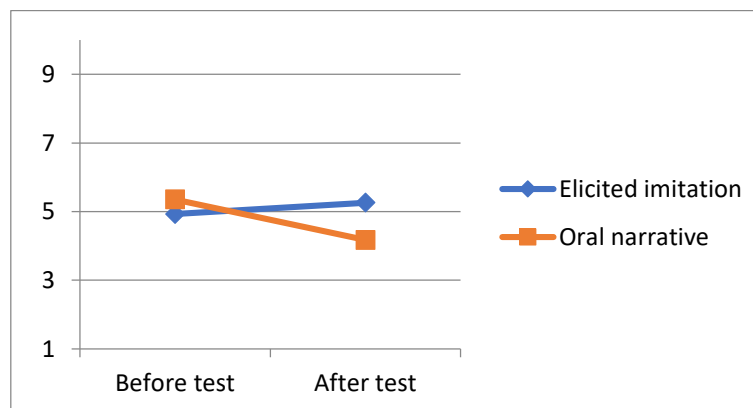


Figure 1. State anxiety before and after EI and ON

Table 2. Dependent samples *t*-tests for the contrasts between state anxiety induced by EI and ON and contrasts between state anxiety before and after the tests.

Contrast	<i>t</i>	<i>df</i>	<i>p</i>	Cohen's <i>d</i>
Before EI/before ON	-0.99	33	.329	0.16
After EI/after ON	2.58	33	.014*	0.44
Before EI/after EI	-0.62	33	.535	0.10
Before ON/after ON	2.31	33	.027*	0.39

\**p* < .05

#### 4.2. Thought-listing

When interpreting the thought-listing results, it should be remembered that the participants recorded their thoughts only before and after the tests, and not in the course of their performance. Therefore, whenever we write of the participants reporting something during the test, this means that they simply recalled a particular thought as occurring at that time. As expected, presumably because language tests are closely associated with both TA and LA, references to anxiety were by far the most frequent of the affective reactions to the two tests revealed by thought-listing. Table 3 presents the numbers of self-reports of the components of anxiety in relation to the two English tests. The reports of tension before, during and after the test were altogether mentioned 48 and 35 times with reference to EI and ON, respectively. The figures reflecting the overall frequency of reported worry stand at 64 (EI) and 33 (ON). Tension and worry were mentioned with comparable frequency, but tension tended to subside more than worry as the test progressed, especially in the case of EI. The tension and worry figures were much higher than those for test-irrelevant thinking, with 6 (EI) and 7 (ON) occurrences; and bodily symptoms, with 4 (EI) and 0 (ON) reports. The lack-of- or diminishing-tension reports were fairly numerous, occurring 20 (EI) and 22 (ON) times. The results of thought-listing are discussed in more detail in the subsections that follow.

Table 3. Frequency of reports of the components of anxiety induced by EI and ON

Anxiety reactions	EI			ON		
	Before	During	After	Before	During	After
Tension	24	12	12	15	14	6
Worry	22	18	24	18	6	9
Test-irrelevant thinking	5	1		1	4	2
Bodily symptoms	4					
Diminishing/lack of tension, relaxation, etc.	7	3	10	5	4	13

#### 4.2.1. Tension

As Table 3 above shows, 24 participants declared tension before EI, and only 12 during and after EI. Some of the tense participants (7 before, 1 during EI) qualified their tension reports by *a little*. Most of those who reported tension did so by using the Polish equivalents of such English words as *stress* and *anxiety*. Additionally, some of them experienced fear (3 before, 4 during), panic (1 during), tiredness/sleepiness (3 before), feeling like laughing (1 during), and the desire to leave the classroom (1 before), which were all interpreted as symptoms of tension following Sarason (1984). Few participants gave reasons for experiencing tension caused by EI, but some mentioned the oral nature of EI (1 before), being recorded and the possibility of the recorder breaking (1 before), silence in the classroom (1 before), the presence of other testees (1 before, 1 during), taking a new kind of test (1 before), and an event unrelated to EI (1 before).

The temporal spread of the tension reports for ON was rather different. 15 participants reported tension before the test and 14 during the test; a considerable drop to 6 occurred only after ON. Similarly to EI, some ON-tension experiencers acknowledged the relatively low intensity of the reaction (6 before, 5 during, 3 after). The vast majority of tension admitters used the Polish lexis mentioned earlier, but additionally fear was listed (3 reports before, 2 after). The following reactions were also mentioned, each one just once: light panic (after), tiredness (after), a funny feeling (before), and a blank mind (during). The identified causes of ON tension included the time limit (1 before, 1 during), being recorded (1 before, 1 during), and the desire to be grammatically correct (1 during). Interestingly, one participant said that her stress was not greater than in the case of EI because she could read the prompts.

Some participants reported no or lessening tension, especially after the tests. A variety of relaxation triggers were mentioned, such as acquaintance with the test task, the presence of other participants, or the test not being part of academic assessment.

#### 4.2.2. Worry

Twenty-two participants reported worry before EI, and this level was generally maintained later on, with a small drop to 18 during EI and an insubstantial increase to 24 afterwards. Before EI, participants admitted to worrying mostly about their performance and test results (13 participants), including, more specifically, not being able to comprehend, remember, or repeat the stimulus sentences; the possibility of making mistakes; and making a good impression. In addition, there were concerns about lack of knowledge (3), not understanding the task (2), making a fool of oneself in the presence of others (2), and others'

opinions about one's pronunciation (1). Furthermore, one participant felt defenseless and another worried partly because of an earlier situation unrelated to EI.

The worry cognitions experienced *after* EI will be discussed before those occurring *during* the test, because certain trends having to do with self-criticism and negativity occurred at this point with particular force. Similarly to before the test, worry after EI usually concerned test performance and results, but the difference was that very often participants reported shame, disappointment, and regret (17 participants). Compared to the earlier ones, these admissions usually involved stronger, more negative language, sometimes expressing lack of self-confidence and self-deprecation (4). At times, they were very specific and concerned such things as memory; knowledge of English; and ability to concentrate, handle the voice recorder, comprehend, remember or repeat the stimulus sentences, or understand the instructions.

The following thoughts illustrate the strong terms in which self-criticism was expressed (and the numerous perceptions of EI as particularly difficult; see Section 4.2.4):

*I don't know why I came [to take the test] if after a moment I forgot what a given "conversation" was about. My memory is poor, I can't concentrate for a little longer because after a moment everything escapes me.*

*I feel stupid, I did not understand the majority of the sentences or I could not focus attention on listening to the sentences, I wish the ground would swallow me up, again I feel defenseless, helpless, stupid, lost opportunities, missed chance*

The worry reports during EI were quite similar to those after it. Very frequent was concern with one's performance in terms of the (in)ability to hear/understand, remember, or repeat the stimuli (15 reports). Other manifestations of worry included disappointment (2), low self-esteem and discouragement (4), resignation replacing positive thinking because of awareness of mistakes (1), comparison of one's performance with others' (1), and worry about the opinions of test administrators (1) and about the ability to manage the recording in the allotted time (1). The language voicing helplessness and self-criticism was milder than that expressing post-EI worry, but it was still quite strong and negative.

The number of participants and comments signaling worry both during (6) and after (8) ON was much smaller than before it (18). Similarly to EI, before ON, participants were troubled (one just a little and one only initially) mostly by their performance and test results (10 participants), but also by not (fully) reading (1) or comprehending the instruction (5), their abilities (1) and the time limit (1). Only one lacked self-confidence "*because of tension*". Similarly to EI, one

participant did not worry, despite not doing very well, because ON was conducted for research purposes only.

While in one case the object of pre-ON worry was very general: “*I am bound to do something wrong*”, the worry was sometimes very specific as it concerned concrete language subsystems and skills as well as evaluation criteria:

*Grammar is not my strong side. I worry that I might make some obvious mistakes and not manage to complete the task within the time limit.*

*I am wondering ... whether I can put everything in the right order, use the right tenses and avoid making obvious mistakes such as –s in the 3rd person singular.*

*Anxiety caused by the fear of failure. Lack of evaluation criteria, preparation for the speech, etc.*

Among the objects of worry during and after ON, participants listed their performance and results (6 comments during, 6 after), quite often mentioning correctness, and in single cases pointing to, more specifically, choice of tenses (during), frequency of using the passive (after), not being the worst in the group (during), “reception” of the test (during), probably by the teacher/researcher, coherence (during), fluency (during), not using all the prompts (after), and not finishing the task (after). While one participant “*got completely off the track*” (during), some others admitted to worrying moderately (2 during, 1 after) or not at all (during).

#### 4.2.3. Test-irrelevant thinking and bodily symptoms

As is shown above in Table 3, few participants indulged in test-irrelevant thinking at any point. Irrelevant thoughts concerned such issues as an event on the following day, forgetting to make tea in the morning or hunger. Only four occurrences of bodily symptoms were reported in relation to EI (before): fast/audible heartbeat (3 occurrences) and shaking hands, with no such reports related to ON.

#### 4.2.4. Test difficulty

A separate thought category concerned the difficulty of the tests. These thoughts were especially numerous with respect to EI (12 during, 12 after). The vast majority pointed to EI being perceived as very or too demanding. Some stated explicitly that a certain element of EI was difficult, and others admitted to an inability to perform. In particular, the inability to hear and/or comprehend the stimuli or problems with doing so were reported (10 reports during, 1 after), with



some participants mentioning the difficulty in understanding *words* in the sentences (3 during, 1 after). Similar problems occurred with respect to remembering (11 during, 1 after) and repeating the stimuli (8 during, 4 after). Concerning their production, for one participant it was hard to “*put the sounds into meaningful units*”. A different participant was unable to use the recorder. The following exemplify the EI difficulty reports:

*During the test I focused on remembering as many sentences as possible, which was difficult because I did not understand all the words, so I could not remember them.*

*I did not understand certain words and I tried to focus on them, which made it impossible for me to remember the whole [sentence].*

*a sense of not understanding the task instruction, the inability to repeat the sentence despite grasping its meaning ...*

Some participants tried to identify the sources of EI difficulty. Four mentioned problems involved in simultaneously attending to two different aspects of the task. In particular, two wrote that providing answers to the distractor task and simultaneously repeating the stimuli was difficult, another “*did not know whether to focus on the sentences [words was intended] in the sentences or whether to remember the content*”, and the last one had problems simultaneously comprehending words and remembering sentences. Similarly, another participant reported difficulty remembering the sentences caused by not knowing all the words. Three mentioned excessive length and/or complexity of (some) of the stimuli as the cause of difficulty. Two ascribed their problems to the characteristics of the recordings: One complained about its “*unEnglish*” and poor quality, and another wished a British native speaker had read the stimuli. Finally, one participant thought the test instruction was difficult, another was distracted by the presence of other testees, and yet another thought EI was easy.

There were fewer ON-difficulty remarks (6 during, 5 after), and four of the after-ones were qualified by expressions such as *a little*. Most comments were rather indirect in that the participants admitted their inability to perform very well rather than explicitly mentioning difficulty. Four referred to the time limit, with one participant mentioning the inability to create logical sentences, one not using all the prompts, and one making mistakes. Other difficulties were creating a coherent story, maintaining fluency, and using the passive more often. Furthermore, one participant complained about having to perform in the presence of and simultaneously with others, and another perceived ON as difficult in general. Additionally, 5 participants claimed that ON was easy, including one stating that it was easier than EI.

#### 4.2.5. Other reactions

Several favorable opinions about the tests were expressed as well. Two participants had a positive attitude to EI. Another wrote that taking it resulted in the desire to record herself at home to practice pronunciation, an outcome she also ascribed to ON. A positive result of taking ON reported by 6 participants was a sense of satisfaction. Its causes, expressed in positive and sometimes even exuberant language, were managing to smoothly deliver a story and doing better on ON than on EI. Additionally, one participant appreciated ON by wishing he did such “*exercises*” more often because “*they significantly improve the ability to form sentences quickly*”.

There were also other thoughts unrelated to anxiety and test difficulty. Before EI, curiosity (12 participants) and excitement/enthusiasm were expressed (5). Five participants were confused during EI, with 4 suggesting that the instruction was not entirely clear and one wondering whether to focus on the meaning or wording of the stimuli, and 2 were confused after EI. Additionally, general (unspecified) uncertainty (3 reports before EI, 1 during) and uncertainty about the test format or contents (7 before) and purpose (2 before) were expressed quite often. The participants also tried to focus on the task, especially as it was in progress (12 participants), but also before (2) and after (3). Additionally, they reported being happy to take part in the research (2 participants before), the desire to test themselves (1 before) and to do their best (1 before), as well as impatience (2 before).

Curiosity was expressed by 6 participants before ON, with one specifying its object: how well she would do, and by 4 after ON, probably in anticipation of the next test.<sup>4</sup> One participant before and one during ON were excited. Several expressed interest in ON (3 before, 1 during, 1 after), while another was bored (1 after). Additionally, one participant reported a lack of knowledge and two stated confusion during the test. General unspecified uncertainty was also experienced (1 before, 3 after). There were also numerous reports of concentration/focus (3 before, 13 during, 1 after), including focus on the construction of the sentences, syntax, selecting tenses, ordering words, grammatical correctness, and making sense rather than just being correct. Additionally, two participants hoped that they would do well (before), and single individuals reported a sense of creativity (during), energy for further action (after), feeling like laughing at one’s speaking (after), dislike of the test format (after), and happiness that the next test was written (see Note 1).

---

<sup>4</sup> After the two tests, another one, this time written, was administered, because the study reported here was part of a larger research project to be reported elsewhere.

## 5. Discussion

On the basis of the quantitative results it may be claimed, with respect to the first research question, that the level of state anxiety induced by EI was greater than that evoked by ON. This finding, however, needs to be treated with caution because it could be related to the ordering of the tests, a point that will be addressed below. A possible explanation for this result seems to be the perception of EI as more demanding. Thus, the present study confirms the findings of earlier research (e.g., Hong 1999; Bonaccio & Reeve 2010) that tests viewed as challenging provoke more anxiety than those viewed as easy. Following Shohamy (1982), who found that tests resembling real-life language use are received more favorably, an additional possible explanation is that test takers view EI, which involves the repetition of oral stimuli seldom performed extensively in natural communication, as less natural than ON, which elicits a short monologue, a much more common type of oral production.

The most conspicuous components of anxiety in relation to the two tests were tension and worry, rather than test-irrelevant thinking or bodily symptoms, so it may be inferred that these are the most common anxiety reactions to EI and ON. In addition, some participants might have simply refrained from recording thoughts greatly divorced from the testing situation due to the belief that they would be irrelevant to researchers. The very small number of reports of bodily symptoms, also reported by Blankstein, Toner and Flett (1989), may have resulted in part from participants' inability to detect them, especially the more subtle ones such as slight perspiration, or perhaps their reluctance to even admit them. The finding that tension decreased more than worry along with the progression of the tests reflects earlier research findings (Zeidner 1998).

The number of tension reports before EI (24) was much greater than before ON (15). This may have been caused by the different reception of the two tests, with the perceived greater difficulty of EI and its unusual character possibly playing an important role. However, it may have also been caused, at least in part, by the simple fact that EI was administered first. The general level of anxiety and apprehension may have been high before it due to the novelty of the research testing situation. This possible cause of the difference in the quantity of the before-the-test tension reports constitutes an important weakness of the study (it is further addressed in the conclusion section). When it comes to the temporal spread of tension reports, considerable decreases were in evidence in both cases, which, however, were not identical. The number of tension reports dropped faster in the case of EI, with relatively few participants reporting tension both during (12) and after (12) the test in comparison with the time before it (24). In the case of ON, the drop from before the test (15) to during it (14) was barely noticeable, and its magnitude increased only after the test (6). The difference between the responses to the two tests could have

been caused by the perception of the EI format as less natural and familiar and the fear that the participants would be unable to deal with it. This difference may have also resulted from the possibly “artificially” inflated number of tense participants before EI due to its administration right at the beginning of data collection. However, the discrepancy between the relatively substantial drop in the number of tension reports from during to after ON and the absence of such a drop in the case of EI may reflect the significant difference between state anxiety after the two tests. Concerning the quality of the tension reports accompanying the two tests, there were no major differences.

The reports of relaxation in relation to the two tests were quite common and comparable in number. Their high frequency is probably related to the tests’ having no influence on grades and to the fact that, according to estimates, the majority of the population do not suffer from excessive TA or LA (Zeidner 1998; Horwitz 2016).

One difference between the tests is that while the number of worry reports (and participants expressing them) during (6 reports/participants) and after (8) ON was much smaller than before it (18), for EI all the worry numbers (22 before, 18 during, 24 after) were very similar. It thus seems that participants worried much less during and after ON than during and after EI. This is in line with the quantitative results discussed earlier and may also reflect the difference in perceptions of test difficulty and naturalness.

The ON worry reports seem to have been more concrete than the EI ones in the sense that they sometimes referred to particular linguistic elements and skills as well as test characteristics. This may have resulted from the fact that ON and its requirements were more transparent to the students in that they could see the linguistic material to be manipulated and were instructed to use appropriate tenses and the right voice (active or passive), neither of which was the case in EI, where they had to rely solely on what they heard and were asked to use “correct English”. This difference suggests the possibility that at least some participants understood the ON instruction better than the EI one and felt more “comfortable” with it as it gave them something specific to focus on. Although this speculation might be undermined by the fact that more participants worried about not understanding what the task was about before ON (5) than before EI (2), it is supported by the numbers of reports of confusion in relation to the two tests at different points in time (7 for EI and just 2 for ON).

An obvious difference between the worry reports during/after the two tests is that the frequency, strength, and negativity of the self-criticisms in relation to ON did not even remotely match those relating to EI. Perhaps the strongest self-directed remark in relation to ON was about getting completely off the track, while the strong ones related to EI expressed such feelings as shame, defenselessness, and a sense of being stupid. What also points to the presence of less worry during/after ON is the

qualification of three ON worry reports as mild, with EI ones never being qualified. Another difference was in the objects of the self-directed worry in relation to the two tests. While the EI critical self-references were sometimes more “global” in that they concerned general cognitive capacities such as memory, ability to concentrate, or the general knowledge of English, the ON ones were only “local” in that they referred to specific abilities or aspects of English such as speaking or tenses. This might mean that EI, contrary to ON, is likely to strongly undermine test takers’ self-confidence and self-worth by exposing their real or imagined cognitive deficiencies.

The comparison of the difficulty, easiness, and satisfaction thoughts reveals that EI was perceived as definitely more demanding, and perhaps too demanding for the participants. First, the number of the difficulty comments for EI (24) was much higher than for ON (11). Second, the negative reactions to EI were much stronger and more definitive. Third, there was only one remark that EI was easy, in contrast to the several ones for ON. Finally, six participants were satisfied with their performance on ON, with nobody making a comparable claim for EI.

The different nature of the difficulty remarks concerning the two English tests reveals why they were not perceived as equally challenging. Numerous comments on EI concerned its fundamental characteristics, that is, the comprehension, retention, and delivery of the stimuli. In addition, several EI remarks mentioned the necessity to simultaneously attend to two cognitively demanding activities such as answering the distractor questions and remembering the stimuli, or remembering both the individual words and whole sentences. In the former case, for example, numerous participants focused on the comprehension of individual words before and, consequently, because of their current processing capacity, perhaps in preference of, conceptualizing the meaning of whole sentences. This points to the participants’ tendency to try to memorize the sentences word by word rather than reconstructing them. These findings suggest that EI may have been too demanding *cognitively*, with its difficulty going beyond the purely linguistic level (grammar, vocabulary, etc.). Incidentally, this may also indicate that the test may not always be the most valid and reliable measure of implicit knowledge. This certainly does not apply in the same measure to ON, as in this case comments do not seem to concern general cognition to the same extent, but, rather, such more specifically linguistic factors as fluency and coherence. What may at least in part explain the cognitive challenge posed by EI is the possibly excessive length of the stimuli and the medium quality of the recordings. However, the strength of the negative reactions to EI makes their neutralization unlikely even if these two aspects were manipulated. The greater perceived difficulty of EI linked to its inherent cognitive challenge is related to the automaticity it requires in linguistic processing and the fact that it tests not only spoken production but also listening comprehension, with both features contributing to its use as a measure of global proficiency (Van Moere 2012).

The two tests elicited comparable, in kind and number, reports of curiosity, interest, and excitement, which probably reflects their novelty. However, EI and ON differed in the number, and, especially, the nature of the uncertainty reports. First, EI was accompanied by 13 reports of uncertainty (excluding the ones concerning performance), this number being much higher than the four occurrences sparked by ON, including three after the test, which might have been caused by the test to follow (see Note 1). Second, there were seven reports of uncertainty about EI format/contents and none concerning ON. While the differences may in part be explained by EI being administered before ON, they also testify to the lesser “palpability” of EI. This elusiveness was probably fueled by the audio nature of its instruction and stimuli, and lack of a declared narrow focus (versus the written stimuli and narrow focus of ON, the instruction of which asked for attention to certain grammatical features of English). The numbers of the concentration reports for the two tests look comparable, but their nature was different, with EI often accompanied by the focus on the processing and reproduction of the sentences, and ON by the focus on the grammar, syntax, tenses, and the like. The difference probably reflects the “palpability” of the tests and the “wide” scope of EI (general correctness) versus the “narrow” one of ON (tenses, voices). The remaining thoughts the tests evoked were quite idiosyncratic and do not warrant any generalizations.

## 6. Conclusion

The present research found that EI, at least when taken for the first time, induces relatively high levels of anxiety and evokes especially high levels of tension and worry, because it is perceived as being very difficult. The challenge of the test has to do not only with purely linguistic features it taps but also, or perhaps especially, with the high cognitive demands on the test takers. This may in fact confirm the suitability of EI as a measure of implicit knowledge and general proficiency in a language such as English, granted that special steps are taken to prevent its administration from inducing too much anxiety. As convincingly argued by Van Moere (2012: 326), measures of implicit representation should focus not only on the comprehension and construction of utterances in communicative tasks such as interviews, which often enable test takers to avoid performance relying mostly on automatic processing, but also on linguistic automaticity, which “includes psycholinguistic processing speed and the interaction between observed fluency, complexity, and accuracy of the speech act”. Our research supports the conclusion that to do its job of tapping implicit linguistic knowledge properly, EI should be administered very carefully with special attention paid to such features as the length of the stimuli, the quality of their recording or the wording of the instruction. For example, the instruction

should perhaps discourage rote memorization of the stimuli word by word in a more direct fashion as the distractor task we used did not do this job very well. Steps of this kind should impact on the perceptions of the test as excessively challenging and therefore reduce anxiety, which, it should be recalled, is likely to confound scores.

More research is needed to learn how we might best employ EI and capitalize on the curiosity it inspires in some learners. In particular, investigations of the best ways of mitigating anxiety evoked by EI, which might involve preparing learners for its high linguistic and cognitive demands, would be welcome. Needless to say, follow-up studies involving counterbalancing the order of tests and higher-stakes administrations would shed more light on the issues investigated in this study.

It should be admitted that the present findings must be treated with circumspection due to some obvious weaknesses such as the possible test order effects and the tests not being part of academic assessment but also the nature of anxiety experienced before and after the tests. When it comes to the first of these, the ordering was conditioned by the logistical constraints in the context in which the study was conducted. With respect to the second, due to the nature of the program, the students were concerned about their use of the target language in the performance of all the activities done in the classroom. It can thus be assumed that anxiety levels (which were relatively high, at least with respect to EI) and other reactions to the tests in our study would be at least as severe during “real” English tests, given their impact on grades. In other words, high-stakes administrations are likely to elicit even stronger negative reactions of the type that occurred in the study, and, perhaps, some additional ones. Therefore, the study might underestimate the severity and quality of the reactions, but it does generalize to higher-stakes testing situations in that the reactions our tests elicited are not likely to disappear during “real” tests. Finally, while it could be argued that the anxiety experienced before the administration of the tests was different in nature from that reported afterwards, there are grounds to assume that both of them were connected with the specific tasks to be performed, thus being state-related. It would also be interesting to look into the relationship between the affective reactions reported by the students and the outcomes of the two tasks, however they were to be measured, a goal that was not addressed in the present study. It is for future research to take account of such issues, thus shedding more light on how the performance of different language learning tasks is linked to the appearance of different affective reactions. This is clearly important since, whether some measures are used for the sake of testing or research, we want to be sure that they tap into the mastery of the language that learners are striving to learn.



## REFERENCES

- Bachman, Lyle F. & Adrian S. Palmer. 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Bielak, Jakub & Mirosław Pawlak. 2013. *Applying Cognitive Grammar in the foreign language classroom: Teaching English tense and aspect*. Springer.
- Blankstein, Kirk R., Brenda B. Toner & Gordon L. Flett. 1989. Test anxiety and the contents of consciousness: Thought-listing and endorsement measures. *Journal of Research in Personality* 23(3). 269–286. DOI: [10.1016/0092-6566\(89\)90001-9](https://doi.org/10.1016/0092-6566(89)90001-9)
- Bonaccio, Silvia & Charlie L. Reeve. 2010. The nature and relative importance of students' perceptions of the sources of test anxiety. *Learning and Individual Differences* 20(6). 617–625. DOI: [10.1016/j.lindif.2010.09.007](https://doi.org/10.1016/j.lindif.2010.09.007)
- Brown, H. Douglas & Priyanvada Abeywickrama. 2010. *Language assessment: Principles and classroom practices* (2nd edn.). Pearson Longman.
- Campfield, Dorota E. 2017. Lexical difficulty – using elicited imitation to study child L2. *Language Testing* 34(2). 197–221. DOI: [10.1177%2F0265532215623580](https://doi.org/10.1177/2F0265532215623580)
- Cassady, Jerrel C. & Ronald E. Johnson. 2002. Cognitive test anxiety and academic performance. *Contemporary Educational Psychology* 27(2). 270–295. DOI: [10.1006/ceps.2001.1094](https://doi.org/10.1006/ceps.2001.1094)
- Cheng, Liying, Don Klinger, Janna Fox, Christine Doe, Yan Jin & Jessica Wu. 2014. Motivation and test anxiety in test performance across three testing contexts: The CAEL, CET, and GEPT. *TESOL Quarterly* 48(2). 300–330. DOI: [10.1002/tesq.105](https://doi.org/10.1002/tesq.105)
- Chaudron, Craig, Hanh Nguyen & Matthew Prior. 2005. *Manual for the Vietnamese Elicited Imitation Test* (Test manual). National Foreign Language Resource Center, University of Hawai'i at Manoa. <http://scholarspace.manoa.hawaii.edu/bitstream/handle/10125/10601/RN41m.pdf?sequence=116>
- Council of Europe. 2001. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Dewaele, Jean-Marc & Peter D. MacIntyre. 2014. The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching* 4(2). 237–274. DOI: [10.14746/ssllt.2014.4.2.5](https://doi.org/10.14746/ssllt.2014.4.2.5)
- Dewaele, Jean-Marc & Helen Thirtle. 2009. Why do some young learners drop foreign languages? A focus on learner-internal variables. *International Journal of Bilingual Education and Bilingualism* 12(6). 635–649. DOI: [10.1080/13670050802549656](https://doi.org/10.1080/13670050802549656)
- Ellis, Rod. 2003. *Task-based language learning and teaching*. Oxford University Press.
- Ellis, Rod. 2005. Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition* 27(2). 141–172. DOI: [10.1017/S0272263105050096](https://doi.org/10.1017/S0272263105050096)
- Erlam, Rosemary. 2006. Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics* 27(3). 464–491. DOI: [10.1093/applin/aml001](https://doi.org/10.1093/applin/aml001)
- Gaillard, Stéphanie. 2014. *The elicited imitation task as a method for French proficiency assessment in institutional and research settings*. University of Illinois PhD dissertation. <http://hdl.handle.net/2142/50562>
- Gkonou, Christina, Mark Daubney & Jean-Marc Dewaele (eds.). 2017. *New insights into language anxiety: Theory, research and educational implications*. Multilingual Matters.
- Gregersen, Tammy S. 2003. To err is human: A reminder to teachers of language-anxious students. *Foreign Language Annals* 36(1). 25–32. DOI: [j.1944-9720.2003.tb01929.x](https://doi.org/j.1944-9720.2003.tb01929.x)

- Gregersen, Tammy S. & Elaine K. Horwitz. 2002. Language learning and perfectionism: Anxious and non-anxious language learners' reactions to their own oral performance. *The Modern Language Journal* 86(4). 562–570. DOI: [10.1111/1540-4781.00161](https://doi.org/10.1111/1540-4781.00161)
- Gregersen, Tammy S., Peter D. MacIntyre & Mario D. Meza. 2014. The motion of emotion: Idiodynamic case studies of learners' foreign language anxiety. *Modern Language Journal* 98(2). 574–588. DOI: [10.1111/modl.12084](https://doi.org/10.1111/modl.12084)
- Hembree, Ray. 1988. Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research* 58(1). 47–77. DOI: [10.3102/00346543058001047](https://doi.org/10.3102/00346543058001047)
- Hewitt, Elaine & Jean Stephenson. 2012. Foreign language anxiety and oral exam performance: A replication of Phillips's *MLJ* study. *The Modern Language Journal* 96(2). 170–189. DOI: [10.1111/j.1540-4781.2011.01174.x](https://doi.org/10.1111/j.1540-4781.2011.01174.x)
- Hong, Eunsook. 1999. Test anxiety, perceived test difficulty, and test performance: Temporal patterns of their effects. *Learning and Individual Differences* 11(4). 431–447. DOI: [10.1016/S1041-6080\(99\)80012-0](https://doi.org/10.1016/S1041-6080(99)80012-0)
- Horwitz, Elaine K. 2010. Foreign and second language anxiety. *Language Teaching* 43(2). 154–167. DOI: [10.1017/S026144480999036X](https://doi.org/10.1017/S026144480999036X)
- Horwitz, Elaine K. 2016. Reflections on Horwitz (1986), “Preliminary evidence for the validity and reliability of a Foreign Language Anxiety Scale”. *TESOL Quarterly* 50(4). 932–935. DOI: [10.1002/tesq.295](https://doi.org/10.1002/tesq.295)
- Horwitz, Elaine K., Michael B. Horwitz & Joann Cope. 1986. Foreign language classroom anxiety. *Modern Language Journal* 70(2). 125–132. DOI: [10.1111/j.1540-4781.1986.tb05256.x](https://doi.org/10.1111/j.1540-4781.1986.tb05256.x)
- Huang, Heng-Tsung Danny & Shao-Ting Alan Hung. 2013. Comparing the effects of test anxiety on independent and integrated speaking test performance. *TESOL Quarterly* 47(2). 244–269. DOI: [10.1002/tesq.69](https://doi.org/10.1002/tesq.69)
- Kenyon, Dorry M. & Valerie Malabonga. 2001. Comparing examinee attitudes toward computer-assisted and other oral proficiency assessments. *Language Learning & Technology* 5(2). 60–83. DOI: [10.125/25128](https://doi.org/10.125/25128)
- Lazarus, Richard S. 1991. *Emotion and adaptation*. Oxford University Press.
- MacIntyre, Peter D. 2002. Motivation, anxiety and emotion in second language acquisition. In Peter Robinson (ed.), *Individual differences and instructed language learning*. John Benjamins. 45–68. DOI: [10.1075/llt.2.05mac](https://doi.org/10.1075/llt.2.05mac)
- MacIntyre, Peter D. & Robert C. Gardner. 1989. Anxiety and second-language learning: Toward a theoretical clarification. *Language Learning* 39(2). 251–275. DOI: [j.1467-1770.1989.tb00423.x](https://doi.org/10.1111/j.1467-1770.1989.tb00423.x)
- MacIntyre, Peter D. & Robert C. Gardner. 1991. Investigating language class anxiety using the focused essay technique. *The Modern Language Journal* 75(3). 296–304. DOI: [10.1111/j.1540-4781.1991.tb05358.x](https://doi.org/10.1111/j.1540-4781.1991.tb05358.x)
- MacIntyre, Peter D. & Robert C. Gardner. 1994. The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning* 44(2). 283–305. DOI: [10.1111/j.1467-1770.1994.tb01103.x](https://doi.org/10.1111/j.1467-1770.1994.tb01103.x)
- MacIntyre, Peter D. & Tammy S. Gregersen. 2012. Affect: The role of language anxiety and other emotions in language learning. In Sarah Mercer, Stephen Ryan & Marion Williams (eds.), *Psychology for language learning: Insights from research, theory and practice*. Palgrave Macmillan. 103–118.
- Madsen, Harold S. 1981. Measuring the impact of ESL-test anxiety on test performance. *Deseret Language and Linguistic Society Symposium* 7(1). 6.

- Naiman, Neil. 1974. The use of elicited imitation in second language acquisition research. *Working Papers on Bilingualism* 2. 1–37.
- Oxford, Rebecca L. 2015. Emotion as the amplifier and the primary motive: Some theories of emotion with relevance to language learning. *Studies in Second Language Learning and Teaching* 5(3). 371–393. DOI: [10.14746/ssllt.2015.5.3.2](https://doi.org/10.14746/ssllt.2015.5.3.2)
- Pawlak, Mirosław. 2006. *The place of form-focused instruction in the foreign language classroom*. Wydział Pedagogiczno-Artystyczny w Kaliszu, UAM w Poznaniu.
- Piechurska-Kuciel, Ewa. 2008. *Language anxiety in secondary grammar school students*. Uniwersytet Opolski.
- Piniel, Katalin & Kata Csizér. 2015. Changes in motivation, anxiety and self-efficacy during the course of an academic writing seminar. In Zoltán Dörnyei, Peter D. MacIntyre & Alastair Henry (eds.), *Motivational dynamics in language learning*. Multilingual Matters. 164–194. DOI: [10.21832/9781783092574-015](https://doi.org/10.21832/9781783092574-015)
- Sarandi, Hedayat. 2015. Reexamining elicited imitation as a measure of implicit grammatical knowledge and beyond...? *Language Testing* 32(4). 485–501. DOI: [10.1177/0265532214564504](https://doi.org/10.1177/0265532214564504)
- Sarason, Irwin G. 1984. Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Cognitive Psychology* 46(4). 929–938.
- Sarason, Irwin G. 1986. Test anxiety, worry, and cognitive interference. In Ralf Schwarzer (ed.), *Self-related cognitions in anxiety and motivation*, 19–35. Erlbaum.
- Scott, Mary L. 1986. Student affective reactions to oral language tests. *Language Testing* 3(1). 99–118. DOI: [10.1177/026553228600300105](https://doi.org/10.1177/026553228600300105)
- Shohamy, Elana. 1982. Affective considerations in language testing. *Modern Language Journal* 66(1). 13–17. DOI: [10.1111/j.1540-4781.1982.tb01015.x](https://doi.org/10.1111/j.1540-4781.1982.tb01015.x)
- Shohamy, Elana. 2001. *The power of tests: A critical perspective on the uses of language tests*. Longman.
- Shohamy, Elana. 2007. Tests as power tools: Looking back, looking forward. In Janna Fox, Mari Wesche, Doreen Bayliss, Liying Cheng, Carolyn E. Turner & Christine Doe (eds.), *Language testing reconsidered*. University of Ottawa Press. 141–152.
- Slobin, Dan Isaak & Charles A. Welsh. 1968. Elicited imitation as a research tool in developmental psycholinguistics. *Working Paper* 10. Reprinted in Charles Albert Ferguson & Dan Isaak Slobin. 1973. *Studies of child language development*, Holt, Rinehart and Winston, 485–497.
- Spielberger, Charles D. 1983. *Manual for the state-trait anxiety inventory* (Form Y). Consulting Psychologists.
- Teimouri, Yasser, Julia Goetze & Luke Plonsky. 2019. Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition* 41(2). 363–387. DOI: [10.1017/S0272263118000311](https://doi.org/10.1017/S0272263118000311)
- Van Compernelle, Rémi A. & Haomin (Stanley) Zhang. 2014. Dynamic assessment of elicited imitation: A case analysis of an advanced L2 English speaker. *Language Testing* 31(4). 395–412. DOI: [10.1177/0265532213520303](https://doi.org/10.1177/0265532213520303)
- Van Moere, Alistair. 2012. A psycholinguistic approach to oral language assessment. *Language Testing* 29(3). 325–344. DOI: [10.1177/0265532211424478](https://doi.org/10.1177/0265532211424478)
- Vinther, Thora. 2002. Elicited imitation: A brief overview. *International Journal of Applied Linguistics* 12(1). 54–73. DOI: [10.1111/1473-4192.00024](https://doi.org/10.1111/1473-4192.00024)
- Woodrow, Lindy. 2006. Anxiety and speaking English as a second language. *RELC Journal* 37(3). 308–328. DOI: [10.1177/0033688206071315](https://doi.org/10.1177/0033688206071315)

- Young, Dolly J. 1991. Creating a low-anxiety classroom environment: What does language anxiety research suggest? *The Modern Language Journal* 75(4). 426–437. DOI: [10.1111/j.1540-4781.1991.tb05378.x](https://doi.org/10.1111/j.1540-4781.1991.tb05378.x)
- Yan, Xun, Yukiko Maeda, Jing Lv & April Ginther. 2016. Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing* 33(4). 497–528. DOI: [10.1177/0265532215594643](https://doi.org/10.1177/0265532215594643)
- Zeidner, Moshe. 1998. *Test anxiety: The state of the art*. Kluwer.
- Zeidner, Moshe & Marsha Bensoussan. 1988. College students' attitudes towards written versus oral tests of English as a Foreign Language. *Language Testing* 5(1). 100–114. DOI: [10.1177/026553228800500107](https://doi.org/10.1177/026553228800500107)

## Appendix A

### Stimulus sentences used in EI

1. Many people have been using computers for many years, which has damaged their eyes.
2. The software Bill Gates invented it changed the world in many different ways.
3. People in love usually want getting married as soon as possible.
4. People should always report the police stolen money as quickly as possible.
5. Middle-aged people often worry about their parent's health and their children's future.
6. When man invented the motor car life change for everyone in the first world.
7. Everyone wants to know what President Komorowski is like in his free time.
8. The number of Europeans with AIDS increased last year and the year before.
9. Every child needs good father to grow into a responsible adult.
10. Not everyone in Poland can learn a second language very quickly.

## Appendix B

### The instruction (translated from Polish) and prompts used in ON

Imagine you have recently visited an American university. Using the prompts given below, talk about the university and some people associated with it. Each numbered prompt should correspond to exactly one sentence in your speech, which must not include more sentences (the number of your sentences should be the same as the number of prompts). The order of sentences should follow the order of the prompts. The underlined sentences must appear in your talk in an unaltered form. Your speech should include all the relevant information given in the prompts, but some nouns may be replaced with pronouns (*he, they, it*, etc., as in example 3 below) or left out (as in example 2 below). In some cases the order of words in your speech should be different from that in the prompts. In a given sentence, try to use the right tense (e.g. *Jerry is writing this letter* vs. *Jerry wrote this letter*) and the right voice, either active or passive (e.g. *Jerry wrote this letter* vs. *This letter was written by Jerry*). Time limit: 3 minutes.

## EXAMPLES:

Prompts	Responses
1. <u>The hotel I want to describe is very nice.</u>	1. <i>The hotel I want to describe is very nice.</i>
2. build/someone/the hotel/one year ago	2. <i>This hotel was built one year ago.</i>
3. visit/many celebrities/the hotel/every year	3. <i>It is visited by many celebrities every year.</i>
4. <u>One of these celebrities is Brad Pitt.</u>	4. <i>One of these celebrities is Brad Pitt.</i>
5. write/Brad Pitt/his first book/last year	5. <i>Interestingly, Brad Pitt wrote his first book last year.</i>

1. The university I will talk about is quite old, but really good.
2. locate/somebody/the university/about twenty kilometers from Washington
3. see/easily/people/the university/from the airport, which is quite close
4. found/the government/the university/in the 19th century
5. so far/attend/two American presidents/the university
6. destroy/an earthquake/the university/in 1919
7. in fact/cause/the earthquake/great panic/in the whole area
8. rebuild/someone/the university/a year later
9. fund/a rich rock star/one department/at that time
10. name/somebody/the department/the School of Popular Music/a few years ago
11. introduce/somebody/a lot of new courses/since then
12. think/people/the university/to be a school for the rich
13. One of such rich students is Bill Gat's son, Mike Gat.
14. own/Mike Gat/two chains of restaurants
15. also/run/Mike Gat/his own company/while studying at university

**Appendix C**

**The English translation of the rating scale**

Mark the level of your anxiety experienced at this moment.

**Example**

Very relaxed \_\_\_\_\_ X \_\_\_\_\_ Very anxious  
Before the test

Very relaxed \_\_\_\_\_ \_\_\_\_\_ Very anxious  
Before the test

Mark the level of your anxiety experienced at this moment.

Very relaxed \_\_\_\_\_ \_\_\_\_\_ Very anxious  
After the test

Mark the level of your anxiety experienced during the test.

Very relaxed \_\_\_\_\_ \_\_\_\_\_ Very anxious  
During the test



## Appendix D

### **The English translation of instructions for the thought-listing instrument**

Please list in 2.5 min as many thoughts and feelings as you are experiencing **RIGHT NOW, BEFORE THE TEST**. Every thought and feeling you are experiencing is important (i.e., thoughts and feelings about yourself, the test situation, or unrelated to the test). Please be spontaneous; it is important that you list all thoughts and feelings that occur.

Please list as many thoughts and feelings as you are experiencing **RIGHT NOW, RIGHT AFTER THE TEST**. Every thought and feeling ... (*the instruction continues as above*)

Please list as many thoughts and feelings you experienced **DURING THE TEST**. Every thought and feeling you **EXPERIENCED DURING THE TEST** is important (i.e., thoughts and feelings about yourself, the test situation, or unrelated to the test). Please be spontaneous; it is important that you list all thoughts and feelings that occurred.