SUBJECTIVE EXPERIENCE OF POETRY: LATENT STRUCTURE AND DIFFERENCES BETWEEN EXPERTS AND NON-EXPERTS

Filip Nenadić[a], Dušan Vejnović[b], & Slobodan Marković[c]

note that Filip Nenadić and Dušan Vejnović are joint first authors

a Department of Linguistics, University of Alberta, Canada

b Department of Psychology, Faculty of Media and Communications, Singidunum University, Serbia

c Laboratory for Experimental Psychology, Department of Psychology, Faculty of Philosophy, University of Belgrade, Serbia

This is an Accepted Manuscript of an article published by Elsevier the *Poetics* journal on 24 April, 2019, available online: https://doi.org/10.1016/j.poetic.2018.11.005

Abstract

We present a series of studies investigating subjective experience of poetry in expert and non-expert readers. In the preliminary study, literature and psychology students produced adjectives descriptive of poetry. Separate lists of adjectives with 5-point Likert-type scales, one for experts, another for non-experts, were created based on their responses. In a series of three subsequent poem rating studies, we show (1) that a more complex factor structure of subjective experience of poetry is formed based on expert responses, (2) that the obtained factor structure for non-experts becomes more similar, yet not entirely identical to the expert solution when non-experts use the adjective list generated by literature students, and (3) that the factor solution obtained for experts remains stable even when a different group of experts rates an entirely different set of poems. The extracted factor structures and their content, differences between expert and non-expert readers, and suggestions to how the developed scales of subjective experience of poetry can be used in future studies are discussed.

Keywords: subjective experience, poetry, literary reading, expertise

1. Introduction

Reader's subjective experience induced by reading literary texts has recently received considerable attention within the field of empirical study of literature. Jacobs (2016) advanced his view of literary experience as comprised of bodily, neurocognitive, and behavioral reactions to a certain text and proposed that measuring each of the three provides us with different information about the process of literary reading. This prompted Kuiken (2016) to refer to the construct as "literary response" and to argue in favor of the more subjective, phenomenologically based methods of data collection, which should offer us a more profound understanding of a reader's experience. Additionally, Dixon and Bortolussi (2016) stressed the importance of longer-lasting effects of literary reading and methods that can produce evidence for them. Taken together, these and several other papers (e.g., Dixon & Bortolussi, 2011; Miall, 2011) brought forth an important discussion of the ways in which literary reading and its outcomes should be studied empirically.

In this paper we present a somewhat different approach to address the question of readers' subjective literary experience. More specifically, we aim to investigate one of the principal subjective aspects of literary reception – the reader's opinion on different characteristics of a given text. Operationalizing it as ratings on a number of distinct dimensions of their subjective experience, we advance an approach that could be viewed as an intersection of Jacobs', Kuiken's, and Bortolussi & Dixon's main points: it examines a measurable, behavioral response that offers a more comprehensive view (in comparison to physiological measures such as pupil dilation or heart-rate indices, for example) of the longer-lasting subjective experience. After presenting the procedure through which subjective literary experience was measured, we consider the latent structure that emerges when it is applied to the genre of poetry.

1.1. Subjective experience of poetry and its assessment

Subjective experience of various art forms has been empirically studied rather extensively since the 1950's. The most common technique used for this purpose was the Semantic Differential, which received wider attention due to Osgood's studies of connotative meaning of words (Osgood, Succi, & Tannenbaum, 1957; Osgood, May, & Miron, 1975). A semantic differential normally features two opposing adjectives as its poles. These adjectives are assumed to be antonyms (e.g., *pleasant-unpleasant*) and are separated by several steps marked by numbers (ranging, for example, from -3 to 3). The task of the participants is to rate the different stimuli on such adjective pair scales. Importantly, statistical analyses (most commonly some form of factor analysis) can then be performed to identify groups of adjective pairs that exhibit similar patterns across stimuli and participants. The extracted factors are then typically taken to represent some basic, latent

dimensions of subjective experience of the rated stimuli.

In a classical study of Berlyne and Ogilvie (1974), for example, three factors of subjective experience of paintings were found in this way: Hedonic Tone, Arousal, and Uncertainty. However, whereas Osgood et al. (1957, 1975) originally used adjective pairs' frequency as an indicator of their importance, Berlyne and Ogilvie's (1974) selection criteria were whether the pair was used in earlier studies or certain theoretical assumptions (see also Berlyne, 1974). Three decades later, Marković & Radonjić (2008) extended this line of research creating a new scale of subjective experience of paintings, which yielded four principal factors: Hedonic Tone, Regularity, Arousal, and Relaxation. Similar procedures have subsequently been applied in various other domains, such as dance (Vukadinović & Marković, 2012) and architectural objects (Marković & Alfirević, 2015; Marković, Stevanović, Simonović, & Stevanov, 2016).

Variations of the described approach have also been applied to the study of poetry or literature in general. Several attempts were made in order to measure a single feature – 'poeticity' – while manipulating rhyme, meter, or some other important characteristic of poetry (e.g., Hanauer, 1996; Hoffstaedter, 1987). In other studies, the number of examined items was somewhat larger, but still fairly low. For example, Kruger, Wirtz, Van Boven, and Altermatt (2004) asked their participants to rate poems according to their personal appreciation (ranging from *I hate it* to *I love it*), and its "overall quality" (ranging from *terrible* to *excellent*). Obermeier et al. (2013) had participants rate the poems on Likert-type scales based on liking (*very bad* to *very good*), strength of emotional response (*very weak* to *very strong*), emotion represented (*negative* to *positive*), and emotion experienced (*negative* to *positive*). Hanauer (2016) administered a scale in which participants rated whether a poem was well-written, beautiful, and whether they experienced an emotional response to it, while Kraxenberger and Menninghaus (2017) examined the beauty of the poem and readers' appreciation of it. Comprised of a small number of items, such scales were instrumental in the investigation of specific effects of manipulated conditions, yet they are not very helpful if one is concerned with a comprehensive account of readers' subjective experience of poetry. One way to sidestep the arbitrariness of item selection is exemplified by the application of the repertory grid technique (O'Hare, 1981; Miall, 1985). Indeed, repertory grids are quite akin to semantic differential scales, given that multiple elements are rated on a number of constructs. However, the crucial difference is that instead of responding to semantic differential scales provided by the researcher, each participant is asked to devise their own dimensions of comparison, which unfortunately renders joint data analysis very difficult.

To our knowledge, Van Peer (1990) was the first to use a larger number of semantic differential scales within the field of literary reading. Seeking to examine the effects of metre on poem reception, Van Peer employed a total of sixteen scales. Nonetheless, his study was still susceptible

to the issue of scale selection criterion, as the scales were chosen “in a largely intuitive manner” (Van Peer, 1990; p. 265). Furthermore, the latent structure based on the participants' responses was not presented in detail, even though Van Peer reports that a factor analysis yielded a six-factor solution. Since inference tests on the extracted factor scores were not significant, each scale was analyzed separately and the latent structure was not considered. Miall (1992) asked participants to rate selected parts of poems using six pairs of adjectives. Although two adjective pairs and one adjective from the third pair were also used in Van Peer’s study, three of the pairs were novel. Forty scales, either taken from various sources or again arbitrarily selected as suitable for assessing “literary quality”, were used by Martindale and Dailey (1995). Interestingly, only four of the adjectives in this study were common to Van Peer’s. Once again, each of the adjectives was analyzed separately, i.e., ignoring the potential latent structure, which might have been quite informative considering that as many as 13 different poems were rated by 32 readers. Finally, Viana, Zyngier, Chesnokova, Jandre, and Nero (2009) created a semantic differential scale questionnaire in Ukrainian, Portuguese, and English as a foreign language. An important difference between this and the previously mentioned studies was that the administered adjective pairs were not arbitrarily selected by the researchers. Instead, the authors initially asked a group of participants to produce relevant adjectives, out of which fifteen were later included in the questionnaire. Even though an effort was made to address the issue of adjective selection, the study featured one major drawback: a single poem, E. A. Poe’s *Annabel Lee*, was used as a prompt for adjective production. The selected adjectives were thus heavily biased (towards elegiac expression and sadness that are characteristic to this particular poem). The follow-up study (Chesnokova, Zyngier, Viana, Jandre, Rumbesht, & Ribeiro, 2017) compared poem ratings using the most frequent of the adjectives in the three languages and in English as a foreign language, but once again it had a similar limitation: a single poem (E. A. Poe’s *The Lake*) was administered to the participants. The need to identify the basic dimensions of the subjective experience was recognized by the same authors in another report (Chesnokova, Zyngier, Viana, Jandre, & Nero, 2009), where they have classified the adjectives based on their meaning into those focused on the text itself (e.g., *difficult/easy* or *long/short*), emotion-driven (*sad/happy*), and appraisal-oriented (*beautiful/ugly*).

As the previous overview suggests, although earlier studies provided the field with some valuable findings, scales for rating poetry that were employed in them were typically flawed in one or more of the following respects: they either (1) consisted of few items, (2) were based on arbitrary selection, (3) varied between participants, or (4) were elicited by presenting participants with a single poem.

1.2. The role of expertise

Despite the challenges in its definition and measurement (see, e.g., Hoffman, 1998), expertise, i.e., domain-specific knowledge or performance, is an indispensable agent in a myriad of cognitive tasks (for detailed reviews, see Ericsson, Charness, Feltovich, & Hoffman, 2006; Sternberg & Grigorenko, 2003). I. A. Richards' (1929) seminal study was the first to examine the individual and group differences in the highly specialized domain such as literary reading. More recently, these differences have been investigated and operationalized in various ways (see, for example, Cupchik, Oatley, & Vorderer, 1998; Dixon, Bortolussi, Twilley, & Leung, 1993; Hoffstaedter, 1987; Miall & Kuiken, 1995; Van Rees, Vermunt, & Verboord, 1999), and the most frequent among the comparisons was the one between readers that could be considered experts vs. less experienced, non-expert readers[1]. A growing interest in this topic is not surprising, since literary expertise was shown to influence performance on a number of tasks, such as thinking aloud during short prose and/or poetry reading (Earthman, 1992; Meutsch, 1989; Peskin, 1998), short story assessment in terms of comprehension, induced affect, and literary value (Dorfman, 1996), 'poeticity' of manipulated poems (Hanauer, 1996), and description of a complex narrative (Graves & Frederiksen, 1991). Expertise effects were even registered between groups with similar backgrounds (Bortolussi & Dixon, 1996; Janssen, Braaksma, Rijlaarsdam, & van den Bergh, 2012), and when comparing children of different age (Peskin, 2010). All these findings indicate that the neglect of expert/non-expert distinction can potentially obscure the characterization of readers' subjective experience of poetry or literature in general.

1.3. Present studies

In this paper, we present three studies (preceded by the preliminary one). All the materials used in them were printed in Serbian, given that this was the native language of all the participants. The studies were conceived to get around the discussed shortcomings of previous research and provide us with the latent structure of readers' subjective experience. In order to do so, we used poetry because it is a genre that is arguably the furthest removed from other "ordinary" language use, and because it is easily distinguished based on the formal characteristics alone (Fabb, 2015; Hanauer, 1996, 2001; Miall & Kuiken, 1998). Furthermore, poems tend to be short enough to make their use in experiments feasible.

The proposed approach expands our understanding of poetry in three major regards. Firstly, a clear latent structure of subjective experience of poetry allows the identification of specific factors that drive poetic reception. With the advent of literary theories that put the reader in the focus of

[1] Admittedly, pinpointing expertise in literary reading is far from an easy task (cf. Graves, 1996). Researchers seem to opt for a single simple indicator of expertise (e.g., they often compare graduate to freshmen students), but what exactly makes a person a literary expert, and whether one should focus on a generic or a more domain-specific type of expertise (cf. Warren, 2011), are questions that remain largely unresolved (but see also Hanauer, 1999).

research, a critical debate ensued about the weight the individual differences have in literary reading. Empirical studies provided the evidence both for general effects that text characteristics produce in all readers, and the tendencies that are shaped by individual or group differences between them (for a review see e.g., Dixon & Bortolussi, 2011). Accordingly, readers belonging to the same population should share a somewhat similar latent notion of what we would refer to as "the space of poetic experience", and we set as our goal to examine to what extent this is true, and what are the common dimensions that define this space. In a manner akin to the one established in the psychology of individual differences, text and reader characteristics are taken to interact within the space of poetic experience, creating different outcomes of reading. The hypothetical construct of the subjective experience of poetry is an attempt to include both the general tendencies and the differences between readers (and poems) in order to describe a variety of these outcomes using shared terms.

Secondly, we hope that a list of carefully selected adjectives, on which the poems are rated, and an established factor structure could provide a basis for the construction of an inventory to be used in further research. Indeed, we sought to create criterion variables that can be employed in both the experimental studies (i.e., those concerned with examination of specific effects of various text, context, or paratext manipulations on readers' subjective experience), and the non-experimental ones (i.e., those that seek to establish relationships between participant characteristics, such as personality traits or reading habits, and their reading experience).

Finally, with the separate treatment of experts and non-experts we choose to take into account the factor of literary expertise. While we accept that readers can form a stable structure of subjective experience, we also hypothesize that certain parts of this structure are likely to develop as the readers' experience with texts increases, and more so in a highly specialized domain as poetry. Expert/non-expert comparison thus seems to be an obvious starting point in the search for the experience-driven differences between readers, because the two groups are clearly distinct in terms of their experience with poetry (yet similar in many other regards). In short, the purpose of the expert/non-expert distinction in our studies was to keep us at bay from potentially erroneous generalizations across importantly divergent groups of readers, and to enable us to specify the aspects of poetic reception that dissociate experts from non-experts.

2. Preliminary study

The goal of the preliminary study was to obtain two lists of descriptors (expert and non-expert one) that can be utilized to characterize poems. We already noted that variations of the Semantic Differential technique were used to investigate subjective experience of various art forms, but it is

worth pointing out that there are several studies in which the method of free production of descriptors had also been employed for similar purposes. For example, lists of descriptors were created in order to describe the appreciation of music (Istók, Brattico, Jacobsen, Krohn, Müller, & Tervaniemi, 2009), various other art forms or objects (Augustin, Wagemans, & Carbon, 2012; Knoop, Wagner, Jacobsen, & Menninghaus, 2016), and even the term "aesthetics" itself (Jacobsen, Buchta, Köhler, & Schröger, 2004).

In our preliminary study, third, fourth, and fifth year students of literature (i.e., experts; N = 53) and psychology (i.e., non-experts; N = 42) were tasked with producing as many adjectives that are characteristic of poetry as they could. Admittedly, whether our sample of literature students can indeed be regarded as truly expert is open to debate. However, we opted to treat their responses as the production of presumed future experts because (a) authentic poetry experts are neither abundant nor easily available for the research and their participation was crucial in our main studies (Study 1 and 3); and (b) the task of adjective production is a very simple one, therefore one could hope that the sophistication and very high levels of expertise may not be necessary for the differences between this group and their non-expert peers to emerge. No particular poem was given as an example or as reading material to elicit the production in either of the groups. Instead, participants were instructed to think about poetry in general when generating adjectives.

A total of 446 unique adjectives (M = 15.31) were produced by literature students, while psychology students provided us with 173 (M = 8.52) adjectives, lending empirical support to the assumed distinction between the two groups. Adjectives produced by at least ten percent of the participants within each group were then retained and were subsequently used in the three main studies. Thus, an adjective was retained if it occurred five or more times in the group of literature students, or four or more times in the group of psychology students. Synonyms and antonyms were allowed to enter the final list of descriptors (i.e., the final list could contain both *cheerful* and *sad*[2], although they may be considered antonyms). Additionally, if a certain adjective was not frequent enough, yet its apparent synonyms and/or antonyms were also produced by the participants, a sum frequency of all the produced synonyms and/or antonyms was calculated in order to ensure that no potentially important descriptor is discarded. In cases when this sum frequency reached the threshold, the most frequent form in the group was included in the descriptor list. For example, neither *warm* nor *cold* occurred frequently enough to warrant the inclusion in the descriptor list on their own, but the sum of their frequencies reached the threshold and thus the more frequent adjective *cold* was included in the final list. Six such adjectives were included for the group of psychology students and twelve for the group of literature students.

[2] In order to facilitate the reading, throughout this report we use the English translations of the original Serbian adjectives that the participants produced. Complete lists of the original Serbian adjectives used in the studies and their English translations are provided in Appendices A and B.

The created scales were unipolar (i.e., they were not defined by pairs of antonyms but rather by single descriptors) for three reasons. First, we wanted to avoid any preconceptions about which adjectives form antonym pairs. Second, we wanted to allow for potential or apparent antonyms to be rated separately on the same stimulus. By accepting that a poem can be rated, for example, as both *sad* and *happy*, we aimed to account for grotesque, paradoxical, or absurd experiences. Third, our intent was to identify synonymous or antonymous adjectives empirically, through high positive or negative correlations.

Lists of descriptors were created separately for expert and non-expert readers. The list of 34 descriptors based on literature students' responses was intended for subsequent expert readers' examination and the list 25 of descriptors based on psychology students' responses was intended for the use in non-experts (see Appendices A and B). Each of the descriptors was followed by a 5-point Likert-type scale, with 1 indicating that the descriptor is not a good characterization of the poem at all, and 5 indicating that the descriptor is very highly characteristic of the poem.

3. Study 1

The main goal of the first study was to establish separate structural characterizations of subjective experience of poetry for two groups of participants: expert and non-expert readers. Additionally, we sought to compare the structures obtained in the two groups for similarities and divergences.

3.1. Participants

Fifteen experts (2 female, 11 male, and 2 of unspecified gender; age: M = 53, SD = 13.31) and nineteen non-experts (8 female, 10 male, and 1 unspecified; age: M = 54.18, SD = 7.56)[3] participated in Study 1. Following Graves's (1996) definition of literary experts, the expert group was comprised of poets (with at least two published books), literary critics, or university professors of literature or languages whose work is related to the field of poetry. Non-experts were selected to match the experts as much as possible according to their age and educational level, but their profession was not related to poetry or arts in general.

3.2. Stimuli

The stimuli set consisted of twenty poems selected from a pool of several hundred poems published in eleven poetry anthologies available in Serbian by two independent selectors (students

[3]Age information was missing for three expert and two non-expert participants. Missing data could not be reconstructed retrospectivelly due to the anonymous character of administration.

in their final year of comparative literature studies). In the first phase of the selection process each of the selectors would be tasked with making a list of proposed poems. In the second phase the two individual lists would be discussed by the selectors in order to determine the final set of texts. The selectors were asked to choose the poems that are as diverse as possible so that the variability of poetic expression is encompassed. No specific criteria (e.g., time of writing, authors' characteristics, membership in a movement, theme, topic, formal features, etc.) were explicitly stated as crucial, although the selectors did discuss them during the process. The selectors, being advanced literature students, were previously familiar with many of the poems. Besides the brief initial instruction, the authors did not interfere with the selection process in any way.

Four of the selected poems were originally written in Serbian and the remainder was available in Serbian translation. Selected poems were uniformly formatted and printed on a sheet of paper without any additional information such as author's name or publication date. The list of poems and author names is provided in Appendix C.

3.3. Procedure

The participants were tasked to read a poem and then rate it on the corresponding list of descriptors (i.e., 34 descriptors in the case of experts and 25 descriptors in the case of non-experts). Poem reading order was randomized per participant. Participants were allowed to peruse the poems throughout the rating process and the time for the task was not limited.

3.4. Data analysis and results

Participant ratings were analyzed using minimum rank exploratory factor analysis (Ten Berge & Kiers, 1991) with Promin rotation (Lorenzo-Seva, 2013) performed in *Factor* software v. 10.03.01 (Lorenzo-Seva & Ferrando, 2013). Data were organized using the string-out method (Osgood, May, & Miron, 1975), which enables extraction of factors irrespectively of the stimulus or the participant. Missing values were treated by Hot-Deck Multiple Imputation procedure (Lorenzo-Seva & Van Ginkel, 2016) and only cases with less than five missing values were included in the analyses. Given that Likert-type scales are ordinal and that certain items had skewness and kurtosis values higher than 1, polychoric correlation matrices were used instead of the Pearson's (Muthén & Kaplan, 1985). When the correlation of any item pair was above 0.8 or below -0.8 one of the items was excluded as a synonym/antonym (see Zygmont & Smith, 2014). Optimal implementation of parallel analysis was used as the factor retention criterion (Timmerman & Lorenzo-Seva, 2011), and a factor loading of 0.32 or higher was considered statistically relevant (Yong & Pearce, 2013).

Three items (*wistful*, *emotional*, and *gloomy*) were excluded from the expert group list because they correlated highly with another item, one (*quiet*) due to low communality values, and another

(*cold*) for insignificant loadings on all of the extracted factors. Accordingly, the final number of descriptors retained was 29. Four factors extracted from them had reliability estimates ranging from .92 to .94, and explained 62% of total and 73% of common variance (KMO = .86). Factor names and loadings are shown in Table 1 and their intercorrelations are provided in Table 2.

Table 1. *Factor loadings for the four factors extracted in the expert group of Study 1*

Evaluation		Affectivity		Joy/Sorrow		Opacity	
boring	-.92	emotive	.84	cheerful	.91	incomprehensible	.92
interesting	.80	affective	.83	optimistic	.88	hermetic	.80
profound	.70	romantic	.79	lively	.83	absurd	.67
inspirational	.69	honest	.57	dark	-.75	unusual	.64
beautiful	.68	sensory	.55	sad	-.75	mysterious	.62
exciting	.64	sad	.53	melancholy	-.66	mystical	.48
succinct	.59	melancholy	.52	difficult	-.59	open	-.40
mystical	.47	melodic	.43	light	.52	boring	.33
incomprehensible	-.40	harmonic	.39				
harmonic	.37						
melodic	.35						

Table 2. *Factor correlations for the expert group of Study 1*

	Evaluation	Affectivity	Joy/Sorrow	Opacity
Evaluation	-	.29	-.10	.21
Affectivity	.29	-	-.00	-.22
Joy/Sorrow	-.10	-.00	-	-.22
Opacity	.21	-.22	-.22	-

In the non-expert group, a total of six descriptors (*emotive*, *happy*, *motivating*, *boring*, *relaxing*, and *joyous*) were excluded due to high correlations with another descriptor, and one more (*enigmatic*) due to its low communality value, so the final number of descriptors was 18. Parallel analysis suggested retention of two factors presented in Table 3, which had reliability estimates of

.93 and .94. Together, they accounted for 56% of total and 68% of common variance (KMO = .86), and had low intercorrelation (r = .22).

Table 3. *Factor loadings for the two factors extracted in the non-expert group of Study 1*

Pleasantness		Sorrow	
cheerful	.82	sad	.84
beautiful	.68	wistful	.83
interesting	.67	touching	.80
inspirational	.65	melancholy	.74
dark	-.62	dark	.66
calming	.61	affective	.65
ugly	-.60	nostalgic	.54
uninteresting	-.59		
incomprehensible	-.59		
leisurely	.55		
sad	-.51		
didactic	.46		
affective	.41		
exciting	.39		

3.5. Discussion

The first study showed that a salient structure of subjective experience of poetry is obtainable, and that this structure is unequal for the two groups of readers. Expert readers produced a more complex structure of subjective experience than their non-expert peers: four basic factors were extracted in order to account for their poem ratings, whereas only two factors were encountered in the non-expert group. Before we look into these factors, note that the names given to them are only approximate terms meant to foster easier communication and perhaps comparison between different structures. Participants most probably impose very little control over which of their adjective ratings will correlate, so we should not assume that they are aware of any such latent dimensions nor that they purposely form an opinion of a text consciously attending to these particular terms. Acknowledging that the naming process itself is inherently reductive and to a certain extent

arbitrary, as it replaces a number of adjectives with a single word, for every extracted dimension we provide examples and also recommend that the reader looks into the full list of its adjectives.

The first factor obtained in experts we refer to as Evaluation, and this factor was loaded by descriptors which closely correspond to aesthetic experience (e.g., *inspirational*, *beautiful*; see Marković, 2012), along with several others that indicate the interest of the reader for the poem (e.g., *boring*, *exciting*). The remaining three dimensions pointed to three prominent characteristics of poetic expression. We named those factors Affectivity, Joy/Sorrow, and Opacity. Affectivity delineates the perceived extent of emotional (and/or romantic) content in a poem, with adjectives such as *emotive*, *romantic*, and *honest*. These adjectives primarily relate to the reader's assessment of whether a human or some other anthropomorphized entity has strong and well-presented feelings – usually of love and affection, but often of sadness or grief, too. High scores on this factor are typical of romantic or love poetry. Joy/Sorrow also has to do with emotional valence, but not necessarily with feelings of a subject in the poem. Rather, this factor reflects the poem's overall tone, which can be formed even if no personal experience is described (e.g., a poem can still be *dark* or *sad*). Additionally, adjectives such as *cheerful* and *lively*, besides the theme itself, can also point to the rhythm of the text or some other formal features. Finally, Opacity marks the degree of difficulty the reader encountered in the effort to understand it, or simply the judgment of whether the poem was even intended to be easily understood or have a single, easily accessible, straightforward meaning. Opacity is loaded with adjectives such as *incomprehensible* or *hermetic*, but also *unusual*, which indicate that the reader encountered a text that is unlike other texts they normally read (including even poetry).

Two dimensions extracted from the non-expert ratings we refer to as Pleasantness and Sorrow. In contrast to the expert group, the dimension Sorrow obtained in non-experts was unipolar, i.e., it did not include negative loadings of joyful descriptors. Instead, it indicated the amount of sad or negative emotion detected by the reader (e.g., *sad*), but could also potentially relate to the emotions experienced by some subject in the poem (e.g., *affective*). Non-experts' Pleasantness seemed to envelop a combination of factors that were discerned as separate in experts. Its highest loading descriptor was *cheerful*, the top descriptor of experts' factor Joy/Sorrow, whereas its other important descriptors (e.g., *beautiful*, *interesting*, *inspirational*) in the expert group converged onto the factor Evaluation. Non-experts' Pleasantness was negatively loaded by descriptors that hint negative emotion (e.g., *sad*, *dark*), but also by those that point to difficulties in understanding (or what was termed as Opacity in the expert structure). Simply put, this factor encompasses the notions of value and quality, positive emotions, ease of understanding and/or a lack of negative emotions. This would suggest that non-experts, besides detecting to what extent the overall tone of the poem is cheerful or sad, basically tend to assess the poems with respect to the amount of pleasure they

experience during reading. Pleasing poems would be those that contain or induce positive emotions or are devoid of negative, disturbing ones, that are not too difficult to understand, and are held as beautiful or inspiring. Importantly, a separate factor, akin to Evaluation, was not isolated in non-experts. This does not mean that non-experts do not make evaluative judgments about the poems they read, but rather that their assessments of quality seem to be indistinguishably entangled with other poem characteristics such as its complexity, emotionality, or affective tone.

When considering the previous results, it is important to keep in mind that they were obtained on the basis of two distinct lists of descriptors. Although the overlap between the lists was substantial, some notable differences were present as well (the most prominent one concerned the total number of descriptors they were comprised of). Therefore, before discussing potential causes of the differences in the structure of subjective experience of poetry between experts and non-experts, in Study 2 we wanted to address the issue of different descriptor lists and their potential influence on the results.

4. Study 2

The goal of Study 2 was to examine whether the previously established differences between experts and non-experts were indeed driven by different expertise levels of the raters and not due to different sets of descriptors used. To answer this question, we sampled a new group of eighteen non-experts (13 female and 5 male; age: M = 43.89, SD = 12.06). This new group of non-experts was given the same set of poems as the participants of Study 1. This time, however, the task was to rate the poems on the list of descriptors produced by literature students (i.e., the one that was used by experts in Study 1). Procedure and data analysis were the same as in Study 1.

4.1. Results

Four items (*lively*, *wistful*, *emotional,* and *gloomy*) were excluded from the analysis due to high correlations with another item, and one (*quiet*) due to low communality values. The remaining 29 items yielded a three-factor solution which accounted for 61% of total and 68% of common variance (KMO = .89). Reliability estimates were .96 for the factors we refer to as Pleasantness and Poignancy, and .90 for the factor Opacity. Item loadings on these factors are provided in Table 4 and factor correlations in Table 5.

Table 4. *Factor loadings for the three factors extracted in the non-expert group in Study 2*

Pleasantness	Opacity	Poignancy

cheerful	.90	incomprehensible	.82	emotive	.86
optimistic	.90	unusual	.74	affective	.86
sad	-.84	mysterious	.73	profound	.80
dark	-.78	mystical	.66	honest	.74
difficult	-.67	absurd	.66	sensory	.72
inspirational	.66	hermetic	.63	sad	.67
melancholy	-.58	boring	.49	succinct	.67
light	.56	cold	.34	melancholy	.64
beautiful	.53			interesting	.63
harmonic	.50			exciting	.58
melodic	.43			harmonic	.57
exciting	.42			beautiful	.55
interesting	.41			difficult	.54
cold	-.36			open	.53
				melodic	.52
				romantic	.45
				dark	.42
				mystical	.41
				mysterious	.36
				boring	-.36
				inspirational	.34

Table 5. *Factor correlations for the non-expert group in Study 2*

	Pleasantness	Opacity	Poignancy
Pleasantness	-	-.34	.24
Opacity	-.34	-	-.01
Poignancy	.24	-.01	-

4.2. Discussion

The results of Study 2 support the notion that non-experts are capable of forming a more developed structure of subjective experience given a larger or a more appropriate set of descriptors. This is evidenced by two notable discrepancies between the structures obtained in two groups of non-experts (Study 1 vs Study 2). Firstly, a unique factor Opacity, separate from Pleasantness, was extracted from the non-expert responses once they were offered the experts' descriptor list. However, this factor was not as distinct as its match previously obtained in experts. For example, the descriptor *boring* still had a negative loading on the factor Pleasantness, indicating the reader did not enjoy reading poems that were difficult to understand or unusual. Additionally, factor Sorrow, extracted in Study 1, reoccurred as a broader one in Study 2, where it included descriptors that point to value and emotional intensity (i.e., descriptors that converged to experts' factors Evaluation and Affectivity) and thus in this latter case we refer to it as Poignancy.

Yet, even with these dissimilarities, a notable common aspect for the two non-expert groups is apparent: neither of them was capable of making a clear distinction between their personal experience induced by poem reading from its intrinsic, 'objective' characteristics, since no factor that resembles experts' Evaluation had been extracted in either of the non-expert groups. Instead, 'unusual' poems and those poems whose understanding posed a challenge for a reader were judged as not pleasant by non-experts.

5. Study 3

Study 3 was conducted in order to test the stability of the factor structure obtained in experts of Study 1. For this purpose, a novel group of sixteen experts (5 female and 8 male; age: M = 44.61, SD = 7.21; age and gender information missing for three participants) was tested on an entirely new set of twenty poems (Appendix D). This new set of poems was selected from the same poem pool and by the same raters as in Study 1. The procedure was the same as in the previous two studies and the same list of expert selected descriptors from Study 1 was employed.

5.1. Results

After the elimination of two items (*wistful* and *emotional*) due to their high correlations with another item and five more items (*cold*, *melodic*, *quiet*, *open*, and *succinct*) due to their low communalities, a four factor structure was extracted. The four factors had reliability estimates ranging from .91 to .95, and explained 69% of total and 79% of common variance (KMO = .90). The factors were very similar to those produced by the expert group in Study 1. Three of them are once again referred to as Evaluation, Affectivity, and Opacity, while the fourth, named Sorrow/Joy

was the polar opposite of the previously extracted factor Joy/Sorrow. Factor names and item loadings are presented in Table 6, while the factor intercorrelations are presented in Table 7.

Table 6. *Factor loadings for the four factors extracted in the expert group in Study 3*

Evaluation		Affectivity		Opacity		Sorrow/Joy	
interesting	.88	romantic	.85	incomprehensible	.99	cheerful	-.90
boring	-.85	emotive	.85	hermetic	.95	optimistic	-.88
beautiful	.82	affective	.84	absurd	.84	lively	-.86
inspirational	.80	sensory	.68	unusual	.70	sad	.84
exciting	.73	honest	.40	difficult	.69	dark	.76
profound	.68	melancholy	.38	mysterious	.67	succinct	.72
harmonic	.39			light	-.48	melancholy	.70
unusual	.37			mystical	.48		
honest	.36			harmonic	-.38		
				honest	-.36		

Table 7. *Factor correlations for the expert group in Study 3*

	Evaluation	Affectivity	Opacity	Sorrow/Joy
Evaluation	-	.22	.32	.32
Affectivity	.22	-	-.26	.03
Opacity	.32	-.26	-	.46
Sorrow/Joy	.32	.03	.46	-

5.2. Discussion

In this study both a new set of expert participants and a new set of stimuli were employed. Notwithstanding, the extracted factor structure remained practically identical to the one obtained in Study 1. Factors Evaluation, Affectivity, and Opacity were replicated, with Joy/Sorrow merely changing its polarity to Sorrow/Joy. Therefore, it can be concluded that the experts' notion of basic dimensions for the description of poetry seems to remain rather stable across different groups of experts and different sets of poems. There was but one noteworthy difference between expert factor structures in Studies 1 and 3: whereas factor intercorrelations for the structure in Study 1 were all

low, in Study 3 we registered a moderate intercorrelation between Opacity and Sorrow/Joy. This serves to confirm that, unsurprisingly, stimulus and participant selection can still have some influence on the factor structure obtained. Clearly, a normative study of subjective experience of poetry should take this into account and entail a careful selection of diverse stimuli.

6. General discussion

Reception of art objects has rightfully received considerable attention within the field of empirical study of literature, just as it has in other domains of artistic expression. However, unlike other forms of art (especially visual), the empirical study of literature is yet to yield a refined structure of principal dimensions of recipients' subjective experience. We hope the studies presented in this paper provide a step in that direction. The four-factor structure extracted from expert ratings collected in Study 1, which included factors Evaluation, Affectivity, Opacity, and Joy/Sorrow, was replicated in Study 3, where another group of experts was given an entirely different set of poems. This structure of subjective experience of poetry is far more complex and elaborate in comparison to the ordinarily used small number of separate scales. Considering its stability across participants, it seems that certain implicit agreement on the important characteristics of poetry exists among expert readers. This in turn makes techniques that allow for fully individualized responses but impede joint analysis (e.g., the repertory grid technique) less appealing.

Unfortunately, in the studies in which a larger number of the same adjectives were administered to all participants (Van Peer, 1990; Martindale & Dailey, 1995) latent structures of responses have not been reported, so it is difficult to assess to what extent they overlap. Van Peer (1990) does mention a six-factor structure obtained from 16 adjectives and 34 participants who were first-year students of language and literature. Although the six-factor structure seems questionable from a statistical standpoint (as it means that at least some of the factors are loaded by two adjectives only), it would certainly be interesting to look into its content and compare it to those obtained in our studies. Moreover, it is also worth noting that our studies were conducted on a sample of Serbian speakers and that materials used (poems and descriptor lists) were in Serbian. An application of a similar procedure in other languages would be instrumental both for the sake of cross-linguistic and cross-cultural comparisons but also in order to examine the stability of the structures.

Nevertheless, some support for the notion of the generality of the dimensions extracted in our studies is already present, and it comes from research of Viana et al. (2009) in which participants were asked to produce pertinent adjectives after reading E. A. Poe's *Annabel Lee*. The most frequent adjectives in this study substantially overlap with those in the preliminary study of ours. Among them we find examples of adjectives that load each of the factors we extracted – *sad*

(Joy/Sorrow), *romantic* (Affectivity), *beautiful* (Evaluation) were the top three most frequent adjectives, while *mysterious* (Opacity) was within the top ten. In another paper (Chesnokova et al., 2009) the same authors even discuss groups of similar adjectives; for example, appraisal-oriented adjectives resemble our Evaluation dimension, while emotion-driven ones are akin to our factors Joy/Sorrow and Affectivity. In the study conducted by Knoop et al. (2016) hundreds of participants produced adjectives that they would use to label various literary genres. In the case of poetry, among the most frequent ones once again we encounter adjectives present in our Studies 1 and 3: *beautiful* (Evaluation), *sad* (Joy/Sorrow), *romantic* (Affectivity). We should also add that the term *beautiful* is probably the most frequent and most common descriptor in any rating of art objects, and that it was even isolated as the most crucial in defining the term "aesthetics" (Augustin et al., 2012; Jacobsen et al., 2004). However, none of the most frequent adjectives from Knoop et al. study corresponds to our experts' fourth factor – Opacity. This observation, in addition to the fact that the adjectives that characterize opaqueness were also less prominent in the study of Viana et al. (2009), suggests that further research is required in order to clarify its position and importance in the structure of subjective experience of poetry.

Another consequence of factor structure stability across the two expert samples relates to the second goal of our study, which was to construct a measure of subjective experience of poetry that could be used to test the effects of various factors in reading experiments, or correlated with text or reader characteristics. We propose that the factor Evaluation, extracted from experts' ratings, could be much more informative for this purpose than the commonly used ratings of liking or other similar measures. The remaining three factors, in turn, could be used to investigate whether certain changes in text or reading context affect not only the overall assessment of poem quality but also the assessment of its particular characteristics; or they could, again, be correlated with objective characteristics of the text.

The use of factor ratings for the investigation of resemblances or differences between particular poems could be another application of the proposed approach. Looking into the extracted factors may be instrumental in the succinct characterization of different types of poems. For example, a poem assessed as high on Affectivity, very low on Joy/Sorrow, and very low on Opacity may be a straightforward lament over a lost love, while a poem judged as moderate on Affectivity, very high on Joy/Sorrow, and very high on Opacity might be a jovial children's poem with a lively rhythm and many nonsense words. An empirical investigation of ratings for a large set of poems could show whether their position in such multidimensional space of extracted factors can be conductive for the purpose of establishing distinct groups or classes of texts. Such a classification could then be contrasted against already existing historical or thematic categories, or even measures obtained from corpus stylistics studies.

Discussion of the findings related to our first two goals underlines the differences between experts and non-experts. To a certain extent, these differences are already observable in our preliminary study, where final year literature students were capable of producing more adjectives that can be used to describe a poem than their psychology peers. In the three main studies, where actual experts are contrasted to non-experts, these differences are evidenced by importantly different structures of subjective poetry extracted from their responses. Study 2 showed that this can partially be explained by different adjective lists used: when non-expert readers were asked to rate the poems on a larger adjective list, the same one that was used by experts of Study 1, the extracted structure more closely resembled the one found in experts. However, even in this case, the structures remained unequal for the two groups, with the conspicuous absence of factor Evaluation in non-experts being the most prominent difference between them. Instead of forming a distinct factor, non-experts' assessment of poem value seems to be strongly related to other poem characteristics, and especially to readers' impression of how pleasing it is.

One potential explanation of the previous discrepancy is that non-experts and experts may have access to genuinely distinct latent spaces of subjective experience of poetry. Poetry reading, and consequently the assessment of experience it leads to, is necessarily founded on the readers' domain or genre knowledge (see Hanauer, 2001; Peskin, 2010). Whereas such knowledge permits experts to establish fine-grained differentiations among poems on several important dimensions, non-experts' much more limited experience with poetry seems insufficient to provide the basis for this. Given the fact that the poems used in our studies were sampled by (future) expert selectors, one could perhaps wonder whether such an expert bias could be the reason for the more elaborate structures obtained in expert groups, and whether non-experts would be capable of producing similarly complex results given a set of poems more suitable for them. As the Van Rees et al.'s study (1999) demonstrates, at least in the case of prose reading, there are indeed classes of readers that enjoy different types of texts ('lowbrow' and 'highbrow' readers). A proper answer to this doubt should be an empirical one. However, keeping in mind the particularities of poetry reading, i.e., the fact that poetry, in difference to prose, is presently seldom read outside the educational or professional context (Bradshaw & Nichols, 2004; Gallik, 1999), we do not believe that such a possibility is likely. Instead, we are inclined to argue that it is the domain knowledge the experts ('highbrow' readers) acquire that enables them to position the poems in the presumed 'poetic space' more appropriately, just as it allows them a more pertinent production and selection of the adequate adjectives. In other words, we believe that the poetry expertise acquisition results in a more developed structure of subjective experience, which enables experts to make more refined judgments, whereas the lack of such domain knowledge prevents non-experts from noting much more than the overall affectivity of the poem and whether they find reading it a generally pleasant experience.

In line with this, the lower percentage of variance explained in the non-expert groups and extraction of a lower number of more general factors also points to lower intergroup agreement. Experts, conversely, seem to be more homogeneous in this respect, presumably due to their more similar backgrounds, extensive experience with poetry, and shared opinions (which in turn facilitate the isolation of a higher number of separate dimensions of subjective experience). Moreover, non-experts may also be more inclined to treat the descriptors differently from one poem to the other. Thus, it can be argued that smaller between-participant agreement and less consistency in ratings within individual non-experts (both indicating a lack of a firm and elaborate notion of what truly important characteristics of a poem are) offer at least a partial explanation for their less refined factor structure in comparison to experts.

Finally, we believe that, besides the application of the scales of subjective experience of poetry for further investigation of the subject (especially in other languages and cultures), these scales could also be instrumental in the applied, educational research concerned with literary reading or expertise – particularly in cross-sectional or longitudinal studies aimed at depicting the development of domain/genre knowledge, or as a tool to assess students' achievement in a classroom context.

7. Conclusions

The presented series of studies enabled us to identify a stable four-factor structure of experts' subjective experience of poetry, independent of stimulus and participant sampling, and a considerably simpler, two-factor structure in non-expert readers. However, when it was built upon a more numerous and more appropriate set of descriptors, the non-expert factor structure became more refined, although still not as elaborate as experts'. Both of these findings corroborate and emphasize the notion of expertise as an important factor in literary reading. It is our hope that the created scales could be employed in future studies as dependent variables or correlated with particular text or reader characteristics. We also look forward to the prospect of similar studies being conducted in other languages, which would enable cross-linguistic and cross-cultural comparisons and clarify the issue of generalizability of the dimensions of poetic subjective experience.

References

Augustin, M. D., Wagemans, J., & Carbon, C. C. (2012). All is beautiful? Generality vs. specificity of word usage in visual aesthetics. *Acta Psychologica, 139*(1), 187-201.

Berlyne, D. E. (1974). Verbal and exploratory responses to visual patterns varying in

uncertainty and in redundancy. In D. E. Berlyne (Ed.), *Studies in the new experimental aesthetics* (pp. 121-158). Washington, DC, New York: Hemisphere Publishing Corporation.

Berlyne, D. E. & Ogilvie, J. C. (1974). Dimensions of perception of paintings. In D. E. Berlyne (Ed.), *Studies in the new experimental aesthetics* (pp. 181-226). Washington, DC, New York: Hemisphere Publishing Corporation.

Bortolussi, M., & Dixon, P. (1996). The effects of formal training on literary reception. *Poetics, 23*(6), 471-487.

Bradshaw, T., & Nichols, B. (2004). *Reading at risk: A survey of literary reading in America* (Research Report No. 46). Washington, DC: National Endowment for the Arts.

Chesnokova, A., Zyngier, S., Viana, V., Jandre, J., & Nero, S. (2009). Universal Poe(try)? Reacting to "Annabel Lee" in English, Portuguese and Ukrainian. In: S. Zyngier and J. Jandre (Eds.), *Linguagem, criatividade & ensino: Abordagens empíricas e interdisciplinares*, (pp. 191-209). Rio de Janeiro: Publit Solucoes Editoriais.

Chesnokova, A., Zyngier, S., Viana, V., Jandre, J., Rumbesht, A., & Ribeiro, F. (2017). Cross-cultural reader response to original and translated poetry: An empirical study in four languages. *Comparative Literature Studies, 54*(4), 824-849.

Cupchik, G. C., Oatley, K., & Vorderer, P. (1998). Emotional effects of reading excerpts from short stories by James Joyce. *Poetics, 25*(6), 363-377.

Dixon, P., & Bortolussi, M. (2011). The scientific study of literature: What can, has, and should be done. *Scientific Study of Literature, 1*(1), 59-71.

Dixon, P., & Bortolussi, M. (2016). Measuring literary experience. *Scientific Study of Literature, 5*(2), 178-182.

Dixon, P., Bortolussi, M., Twilley, L. C., & Leung, A. (1993). Literary processing and interpretation: Towards empirical foundations. *Poetics, 22*(1), 5-33.

Dorfman, M. H. (1996). Evaluating the interpretive community: Evidence from expert and novice readers. *Poetics, 23*(6), 453-470.

Earthman, E. A. (1992). Creating the virtual work: Readers' processes in understanding literary texts. *Research in the Teaching of English, 26*(4), 351-384.

Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (Eds.). (2006). *The Cambridge handbook of expertise and expert performance*. New York, NY, US: Cambridge University Press.

Fabb, N. (2015). *What is poetry?: Language and memory in the poems of the world*. Cambridge: Cambridge University Press.

Gallik, J. D. (1999). Do they read for pleasure? Recreational reading habits of college students. *Journal of Adolescent & Adult Literacy, 42*(6), 480-488.

Graves, B. (1996). The study of literary expertise as a research strategy. *Poetics, 23*(6), 385-

403.

Graves, B., & Frederiksen, C. H. (1991). Literary expertise in the description of a fictional narrative. *Poetics, 20*(1), 1-26.

Hanauer, D. (1996). Integration of phonetic and graphic features in poetic text categorization judgments. *Poetics, 23*(5), 363-380.

Hanauer, D. (1999). Attention and literary education: A model of literary knowledge development. *Language Awareness, 8*(1), 15-29.

Hanauer, D. (2001). What we know about poetry reading. In: D. Schram & G. Steen (Eds.), *The psychology and sociology of literature: In honor of Elrud Ibsch*, (pp. 107-128). Amsterdam: John Benjamins.

Hanauer, D. I. (2016). Beauty judgements of non-professional poetry. *Scientific Study of Literature, 5*(2), 183-199.

Hoffman, R. R. (1998). How can expertise be defined? Implications of research from cognitive psychology. In R. Williams, W. Faulkner, & J. Fleck (Eds.), *Exploring expertise: Issues and perspectives* (pp. 81-100). London: Palgrave Macmillan.

Hoffstaedter, P. (1987). Poetic text processing and its empirical investigation. *Poetics, 16*(1), 75-91.

Istók, E., Brattico, E., Jacobsen, T., Krohn, K., Müller, M., & Tervaniemi, M. (2009). Aesthetic responses to music: A questionnaire study. *Musicae Scientiae, 13*(2), 183-206.

Jacobs, A. M. (2016). The scientific study of literary experience. *Scientific Study of Literature, 5*(2), 139-170.

Jacobsen, T., Buchta, K., Köhler, M., & Schröger, E. (2004). The primacy of beauty in judging the aesthetics of objects. *Psychological Reports, 94*(3), 1253-1260.

Janssen, T., Braaksma, M., Rijlaarsdam, G., & van den Bergh, H. (2012). Flexibility in reading literature: Differences between good and poor adolescent readers. *Scientific Study of Literature, 2*(1), 83-107.

Knoop, C. A., Wagner, V., Jacobsen, T., & Menninghaus, W. (2016). Mapping the aesthetic space of literature “from below”. *Poetics, 56*, 35-49.

Kraxenberger, M. & Menninghaus, W. (2017). Affinity for poetry and aesthetic appreciation of joyful and sad poems. *Frontiers in Psychology, 7*, 2051.

Kruger, J., Wirtz, D., Van Boven, L., & Altermatt, T. W. (2004). The effort heuristic. *Journal of Experimental Social Psychology, 40*(1), 91-98.

Kuiken, D. (2016). The implicit erasure of “literary experience” in empirical studies of literature. *Scientific Study of Literature, 5*(2), 171-177.

Lorenzo-Seva, U. (2013). *Why rotate my data using Promin?*. Technical Report. Department of

Psychology, Universitat Rovirai Virgili, Tarragona.

Lorenzo-Seva, U., & Ferrando, P.J. (2013). FACTOR 9.2. A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and IRT Models. *Applied Psychological Measurement, 37*(6), 497-498.

Lorenzo-Seva, U., & Van Ginkel, J. R. (2016). Multiple Imputation of missing values in exploratory factor analysis of multidimensional scales: estimating latent trait scores. *Anales de Psicología/Annals of Psychology, 32*(2), 596-608.

Marković, S. (2012). Components of aesthetic experience: aesthetic fascination, aesthetic appraisal, and aesthetic emotion. *i-Perception, 3*(1), 1-17.

Marković, S., & Radonjić, A. (2008). Implicit and explicit features of paintings. *Spatial Vision*, *21*(3), 229-259.

Marković, S., Stevanović, V., Simonović, S., & Stevanov, J. (2016). Subjective experience of architectural objects: A cross-cultural study. *Psihologija, 49*(2), 149-167.

Marković, S., & Alfirević, Đ. (2015). Basic dimensions of experience of architectural objects' expressiveness: Effect of expertise. *Psihologija*, *48*(1), 61–78.

Martindale, C., & Dailey, A. (1995). IA Richards revisited: Do people agree in their interpretations of literature?. *Poetics, 23*(4), 299-314.

Meutsch, D. (1989). How to do thoughts with words II: Degrees of explicitness in think-aloud during the comprehension of literary and expository texts with different types of readers. *Poetics, 18*(1), 45-71.

Miall, D. S. (1985). The structure of response: A repertory grid study of a poem. *Research in the Teaching of English, 19*(3), 254-268.

Miall, D. S. (1992). Response to poetry: Studies of language and structure. In E. F. Nardocchio (Ed.), *Reader response: The empirical dimension* (pp. 153-170). The Hague: Mouton.

Miall, D. S. (2011). Science in the perspective of literariness. *Scientific Study of Literature, 1*(1), 7-14.

Miall, D. S., & Kuiken, D. (1995). Aspects of literary response: A new questionnaire. *Research in the Teaching of English, 29*, 37-58.

Miall, D. S., & Kuiken, D. (1998). The form of reading: Empirical studies of literariness. *Poetics, 25*(6), 327-341.

Muthén, B., & Kaplan D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171-189.

Obermeier C., Menninghaus W., von Koppenfels M., Raettig T., Schmidt-Kassow M., Otterbein S., & Kotz, S. A. (2013). Aesthetic and emotional effects of meter and rhyme in poetry. *Frontiers in*

Psychology, 4, 10.

O'Hare, D. (1981). Cognition, categorization and aesthetic response. In H. Bonarius, R. Holland, & S. Rosenberg (Eds.), *Personal construct psychology: Recent advances in theory and practice* (pp. 147-155). New York: St. Martin's Press.

Osgood, C.E., May, W., Miron, M. (1975). *Cross-cultural universals of affective meaning*. Urbana-Chicago, IL: University of Illinois Press.

Osgood, C.E., Succi, G.J., Tannenbaum, P.H. (1957). *The measurement of meaning*. Urbana-Chicago, IL: University of Illinois Press.

Peskin, J. (1998). Constructing meaning when reading poetry: An expert-novice study. *Cognition and Instruction, 16*(3), 235-263.

Peskin, J. (2010). The development of poetic literacy during the school years. *Discourse Processes, 47*(2), 77-103.

Richards, I. A. (1929). *Practical Criticism*. London: Routledge & Kegan Paul.

Sternberg, R. J., & Grigorenko, E. L. (Eds.) (2003). *The psychology of abilities, competencies and expertise*. Cambridge, UK: Cambridge University Press.

Ten Berge, J.M.F., & Kiers, H.A.L. (1991). A numerical approach to the exact and the approximate minimum rank of a covariance matrix. *Psychometrika, 56*, 309-315.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16*(2), 209.

Van Peer, W. (1990). The measurement of metre: Its cognitive and affective functions. *Poetics, 19*(3), 259-275.

Van Rees, K., Vermunt, J., & Verboord, M. (1999). Cultural classification under discussion: Latent class analysis of highbrow and lowbrow reading. *Poetics, 26*(5), 349-365.

Viana, V., Zyngier, S., Chesnokova, A., Jandre, J., & Nero, S. (2009). Quantitative research in practice: applying a differential scale questionnaire to literature. *Філологічні трактати, 1*(3-4), 67-77.

Vukadinović, M., & Marković, S. (2012). Aesthetic experience of dance performances. *Psihologija, 45*(1), 23-41.

Warren, J. E. (2011). "Generic" and "specific" expertise in English: An expert/expert study in poetry interpretation and academic argument. *Cognition and Instruction, 29*(4), 349-374.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9*(2), 79-94.

Zygmont, C., & Smith, M. R. (2014). Robust factor analysis in the presence of normality, violations, missing data, and outliers: Empirical questions and possible solutions. *Quantitative Methods in Psychology, 10*(1), 40-55.

APPENDICES

Appendix A. *Adjective list obtained in the preliminary study based on responses from literature students (expert adjective list)*

Original (Serbian)	English
vesela	cheerful
hladna	cold
inspirativna	inspirational
duboka	profound
lepa	beautiful
polenta	lively
ljubavna	romantic
apsurdna	absurd
melodična	melodic
mračna	dark
tajanstvena	mysterious
uzbudljiva	exciting
laka	light
optimistična	optimistic
tiha	quiet
osećajna	affective
iskrena	honest
setna	wistful
melanholična	melancholic
emocionalna	emotional
dosadna	boring
otvorena	open
harmonična	harmonic
tužna	sad
teška	difficult
tamna	gloomy
jezgrovita	succinct
nerazumljiva	incomprehensible
hermetična	hermetic
čulna	sensory
emotivna	emotive
neobična	unusual
zanimljiva	interesting
mistična	mystical

Appendix B. *Adjective list obtained in the preliminary study based on responses from psychology students (non-expert adjective list)*

Original (Serbian)	English
uzbuđujuća	exciting
emotivna	emotive

nostalgična	nostalgic
poučna	didactic
srećna	happy
zagonetna	enigmatic
lepa	beautiful
osećajna	affective
vesela	cheerful
zanimljiva	interesting
tužna	sad
nerazumljiva	incomprehensible
dirljiva	touching
inspirativna	inspirational
motivišuća	motivating
lagana	leisurly
mračna	dark
dosadna	boring
nezanimljiva	uninteresting
melanholična	melancholic
setna	wistful
opuštajuća	relaxing
umirujuća	calming
ružna	ugly
radosna	joyous

Appendix C. *List of poems used in Study 1*

Author	Title
Rafael Alberti	Buster Keaton Searches the Woods for his Sweetheart a Genuine Cow
Anna Akhmatova	Requiem V [For seventeen months I have been screaming]
Hertha Kräftner	*No title*
Lucretius	On the Nature of Things (excerpt)
William Blake	I Saw a Chapel
Friedrich Hölderlin	Diotima
Marie Luise Kaschnitz	Genazzano
Erich Fried	Constructive Self Criticism
Arthur Rimbaud	Marine
Dragan Aleksić	Village Calm
Carl Michael Bellman	Nota Bene
Jovan Subotić	Clouds
Nikolaus Lenau	The Oppressive Evening
Sappho	Love Poem
Gerhard Fritsch	Afternoon
Charles Baudelaire	Correspondences
Vasko Popa	The Pig
Francesco Petrarca	Sonnet 134
Antonio Machado	The Plaza Held a Tower
Vojislav Ilić	Abandoned Spring

Appendix D. *List of poems used in Study 3*

Author	Title
Marin Sorescu	Segment
Günter Eich	Inventory
Jorge Luis Borhes	Labyrinth
Wystan Hugh Auden	There is No Change of Place
Nichita Stănescu	Natural Unusualness
Octavio Paz	The Tanghi-Garu Pass
Gunnar Harding	Puberty
Leopold Senghor	What are you doing?
Ingeborg Bachmann	The Large Freight
Stanisław Grochowiak	Clean Men
Nikolay Gumilyov	The Giraffe
Victor Hugo	Sowing Season. Evening
Théophile Gautier	The Hippopotamus
Vasa Živković	The Flower
Georg Heym	Shadow of Life (Umbra Vitae)
Monny de Boully	The Chameleon
Ilarie Voronca	The Piano
Srečko Kosovel	The Laugh of King Dada
Anacreon	The Grasshopper
Walther von der Vogelweide	Under the Lime Tree