

Cvetka Sokolov

Faculty of Arts, University of Ljubljana
Slovenia
cvetka.sokolov@ff.uni-lj.si

UDK 811.111'243:37.091.279.7
DOI: 10.4312/vestnik.14.275-292
Izvirni znanstveni članek



THE CHALLENGE OF ASSESSING CONTENT AND COHERENCE IN WRITING

1 INTRODUCTION

In analytic scoring, raters of written work find it more difficult to objectively and reliably assess the quality of content and coherence than that of grammar and vocabulary. One obvious reason for this is that, for example, a grammatical error is easier to spot than a supposedly illogical, underdeveloped, or irrelevant proposition. “Dimensions such as coherence and content do not have overt linguistic markers that are countable” (Bae 2001: 62; cf. Glenn/Goldthwaite 2014: 129). No wonder readers disagree about the quality of the content rather than the linguistic means by which it is expressed. In the words of Peter Elbow (1996: 121):

We have long known that readers bring their own diverse values to what they read – indeed, they help construct the very meanings they find in a text. (...) Thus we shouldn’t be surprised that even the most skilled readers characteristically disagree with one another not only in their valuations of a text but even about its meanings. (Elbow 1996: 121; cf. Bean 2011: 277; Cushing Weigle 2002: 71; Holdstein 1996: 219; White 1996: 16; Wilson/Hanna 1993: 236; Yu 2007: 541)

This means that the score assigned at the end of the day “(reflects), not only of the quality of the performance, but of the qualities as a rater of the person who has judged it” (McNamara 2000: 37; cf. Bean 2011: 279).

Teacher education and training programmes play an important role in keeping this unfortunate, albeit natural, condition in check. The same is true for standardization meetings organized by groups of teachers teaching in the same school (see, e.g., Bean 2011: 287 and Sokolov 2014: 120) and for national external examination bodies, such as the State Examination Centre (Slov. *Državni izpitni center*) in Slovenia, in order to provide

a common basis for the interpretation and implementation of assessment criteria and to minimize their biased use (see, e.g., Bacha 2001: 375; Bean 2011: 269; Eckes 2008, Hughes 2003: 103; Hyland 2003: 229; Skoufaki 2020: 118; Spandel/Stiggins 1990: 68–72). However, Sara Cushing Weigle (2002: 72) points out that “while training can help bring raters to a temporary agreement on a set of common standards, research has consistently shown that raters will never be in complete agreement on writing scores.”

To make matters worse, the use of analytical scoring and standardization procedures is not as helpful as it could be if the problem were not exacerbated by overly vague descriptors (see, e.g., Knoch 2007: 109; Skoufaki 2020: 120; Sokolov 2018: 173), and if there were more agreement in the perceptions of writing experts and scoring scale developers about the demarcation line between content and coherence. Are theoretical considerations concerning coherence reflected in the evaluation criteria? More specifically, should an irrelevant paragraph count as a coherence break, or should it be penalized within the content category? This article looks at the descriptors in the two categories of two grading scales used in Slovenia to examine them for specificity/vagueness and (in) consistency, as well as the distinction between the two graded categories, and the extent to which they take into account some relevant theoretical insights. The second part includes a case study of secondary school teachers’ perceptions of the two categories and their interpretations of the key descriptors used to assess content and coherence.

2 TERMINOLOGY AND THEORETICAL CONSIDERATIONS

2.1 Content

The concept of the content of a text seems easy to grasp, but less so to define. Simply put, it refers to the meaning(s) and message(s) of the text. Jungok Bae (2001: 54) defines the term more comprehensively:

(C)ontent represents the semantic domain of language. (...) It (encompasses) the relevance of a written text to a given task, as well as thoroughness, persuasiveness, and creativity consistent with task expectations. The quality of content is thus viewed as the degree to which the writing impresses the reader in terms of these criteria. (Bae 2001: 54)

The definition includes the basic characteristics, such as “persuasiveness”, “creativity” and “the degree to which the writing impresses the reader”, that make writing and reading texts exciting, but also challenging to evaluate objectively.

On the other hand, “the relevance of a written text” and “thoroughness” (development) are more tangible and therefore easier to evaluate. However, since relevance is also

important to ensure the coherence of written texts, it becomes more difficult to distinguish between the two categories and to respect the rule that the same weakness in a written composition may be punished only once.

2.2 Coherence and Cohesion (and Organization)

A coherent text makes sense. The ideas in such a text are logically related to one another and are clearly expressed so that the reader does not have to keep going back to what he or she has already read to think about parts of the text in order to understand them. “When you focus on coherence, you’re providing readers with the all-important context they need to help them understand what they’re reading” (Douglas 2015: 90). In the words of A. Franking Parks et al. (1991: 108), a coherent text allows the reader to “move smoothly from sentence to sentence without becoming confused or losing the writer’s train of thought.” Ronald Carter (1993: 9) defines coherence as a “semantic and propositional organisation” of a text that manifests itself in “the concepts, propositions, or events (being) related to each other and (being) consistent (with the overall subject of the text)” (Cf. Alaro 2020: 41; Briesmaster/Etchegaray 2017: 187; Skoufaki 2020: 104). When the ideas/supporting points in such a text are “related to each other and consistent with the overall subject of the text”, they are relevant. And this is where the categories of content and coherence merge.

The quality of coherence, “the relationships which link the meanings of sentences in a text” (Richards et al. 1992: 61), also affects the way we perceive the content of a text in other ways. A text that is difficult to process because of many breaks in coherence is bound to seem less convincing and impressive than a well-structured, coherent text. Bae (2001: 58) also points out the close relationship between coherence and content: “An incoherent text with disjointed connections cannot communicate content effectively.” Apart from the relevance of ideas, which is obviously an indispensable component of content and coherence, this also makes the separate assessment of each of the two categories a challenge, thus explaining the sometimes confusing situation of some of the comparable descriptors being assigned to one or the other in various assessment criteria.

Cohesion, “the appropriate linguistic links between sentences” (Carter 1993: 8), is not essential in text composition (see, e.g., Brown/Yule 1983: 199; Hoey 1991: 12; Nunan 1993: 61; Widdowson 1978: 29). Nevertheless, grammatical and lexical devices that connect ideas in a clear and logical way are a prerequisite for the production of more complex texts (Cf. Alaro 2020: 41). It is important to point out, though, that cohesion alone does not ensure coherence. According to Bae (2001: 56–57), “(c)ohesive markers alone (...) do not necessarily make the text coherent and comprehensible. A text full of cohesive markers that are locally correct could be incoherent and incomprehensible as a whole” (see also Enkvist 1990). Moreover, the reader of a coherent text must use his or

her knowledge of the world to interpret its meaning(s) (see, e. g., Brown/Yule 1983; Carrel 1982; Douglas 2015: 23–25; Enkvist 1990: 14; Halliday/Hasan 1976).

There are scoring criteria in which the category “coherence” is replaced by the category “organization” (see, e.g., Brown/Bailey 1984 in Brown/Abeywickrama 2019: 252; Jacobs et al. 1981 in Hughes 2003: 104; ReadWriteThink in Bean 2011: 271; Spandel et al. 1990); in other words, some experts on testing writing consider organization to be synonymous with coherence. Bae (2001: 55) puts it somewhat tentatively, “Organization (...) **may** be considered **similar to or part of** coherence (my emphasis).” Interestingly, the Matura Exam Rating Scales (Ilc et al. 2018: 15–16) refer to the category as “organization and cohesion”, suggesting that the constructors of the scales felt that “organization” was not enough. Personally, I associate organization with the basic structure of a text, that is paragraphing, and therefore consider it a sub-category of the umbrella term “coherence”. Since cohesion refers to the tools used to establish coherence, which goes beyond the linguistic means used, I would refrain from using the term “cohesion” to replace the label “coherence” in the criteria.

2.3 Coherence Breaks

Coherence breaks interrupt the reader’s smooth processing of a text. Eleanor Wikborg (1990: 133) defines them as “(w)hat happens when the reader loses the thread of the argument when reading a text attentively”. Her classification of coherence breaks (Wikborg 1990: 134) is divided into two groups, namely Topic-Structuring Problems and Cohesion Problems. While breaks in cohesion clearly belong to the category of coherence, two of three subgroups in the first group of coherence breaks, namely Irrelevance and Unjustified Implicit Coherence¹ (see Table 1), include textual features that could also be attributed to the category of content (see 2.1. and 2.2. above).

A. Topic-Structuring Problems

Table 1: Topic-Structuring Problems (Wikborg 1990: 134)

A.1.	IRRELEVANCE
	A missed title
	An irrelevant paragraph
	An illogical, contradictory or untrue proposition
	An unjust change of/drift of topic
	Overcompleteness
A.2.	UNJUSTIFIED USE OF IMPLICIT COHERENCE

¹ The third subcategory, Misleading Paragraph Division, is clearly related to coherence of texts. For the comprehensive classification see Wikborg 1990: 134. For specific authentic examples of coherence breaks see Sokolov (1999 and 2000).

Jacobs et al. (1981 in Hughes 2003: 104), for example, place the descriptor “relevant to assigned topic” in their scales into the category of content (see also ReadWriteThink in Bean 2011: 271), while its equivalent(s) on Wikborg’s (1990: 134; see above) list of coherence breaks, namely “a missed title” and/or “an irrelevant paragraph” and/or “an unjust change of/drift of topic”, are assigned to the category of coherence. On the other hand, it could also be argued that the descriptor “limited support” (Jacobs et al. 1981) in the organization/coherence category actually covers the content aspect defined as “thoroughness”/development (see 2.1. above). After all, the descriptor “limited development of thesis”, which is synonymous with “limited support” (Jacobs et al. 1981), is located in the content category.

On the other hand, “limited development of theses”/“limited support” might be related to the kinds of coherence breaks Wikborg (1990: 134) calls “Unjustified Use of Implicit Coherence”, referring to inappropriate propositional gaps in a text that the reader is expected to fill in without receiving enough clues/development of ideas (a content category, mind you) from the author. In other words, this is another textual feature that is closely related to both content and coherence. Thirdly, Wikborg (1990: 134; see above) and Jacobs et al. (1981) consider illogical sentences as offenders that confuse the reader by committing coherence breaks, while Brown and Bailey (1984 in Brown/Abeywickrama 2019: 252), for example, specify the category of content by calling it the “logical development of ideas”. So, which of the two categories do logical fallacies belong in?

How do the most commonly used analytic scales for assessing (longer) written compositions in grammar schools and other secondary schools in Slovenia, namely the Matura Exam Rating Scales (Ilc et al. 2018: 15-16), and the analytic scales used in the Language in Use courses at the Department of English, Faculty of Arts, University of Ljubljana (Sokolov et al. 2014), reflect this confusing situation?

3 CONTENT AND COHERENCE IN MATURA SCALES AND DEPARTMENT OF ENGLISH, FACULTY OF ARTS, UNIVERSITY OF LJUBLJANA, SCALES

3.1 Content

Grammar/secondary school students receive the highest score for the content of their papers when the content is “completely relevant to the title” and “provides convincing, in-depth and balanced supporting points/complex reasoning”, according to the Matura Exam Rating Scales² (Ilc et al. 2018: 15-16), which are comparable to the first two descriptors in the rating scales used at the Department of English, Faculty of Arts, University of

2 MERS

Ljubljana³ (Sokolov et al. 2014): “content totally relevant to the assigned topic and fully developed (convincing and well-balanced argumentation)”. Similarly to Jacobs et al.’s scales, the two scales used in Slovenia treat relevance as an indispensable part of content, which is consistent with the definition that content is about “the relevance of a written text to a given task” (Bae 2001: 54; see 2.1. above) and seemingly contradicts Wikborg’s (1990: 134) classification of a missed title and an irrelevant paragraph. Considering that irrelevance in a text prevents fluent reading, which is the very characteristic of a coherent text, this conclusion is not valid. On the other hand, Bae (2001: 54) is also right – irrelevant content is useless. In other words, relevance is important in both categories.

The descriptors “convincing, in-depth and balanced supporting points/complex reasoning” in the MERS and “fully developed (convincing and well-balanced argumentation)” in the DoES cover “thoroughness” and (partly) “persuasiveness” in Bae’s (2001: 54) definition, while creativity and impressiveness (ibid.) are not taken into account in the MERS.

This is not the case with the DoES, which contain a number of other, more specific descriptors:

Table 2: Department of English, Faculty of Arts, University of Ljubljana, Scales (Sokolov et al. 2014): Content

Grade	Content
10	<ul style="list-style-type: none"> • Content totally relevant to the assigned topic and • fully developed (convincing and well-balanced argumentation); • effective introductory and concluding paragraphs (interest-catching & closing techniques); • lively/convincing/illustrative/relevant details/supporting points; • original insights into the question (fresh, surprising, daring ideas); • no logical fallacies; • demonstration of critical thinking skills (in the case of opinion and argumentative essays); • the tone and voice give flavour to the writer’s message; • holds the reader’s attention.

Creativity and the requirement that the content should “impress the reader” (Bae 2001: 54) are considered with the descriptors “original insights into the question (fresh, surprising, daring ideas)”, “demonstration of critical thinking skills”, “the tone and voice give flavour to the writer’s message” and “holds the reader’s attention”. In addition, some of the added descriptors define “convincing argumentation” more explicitly: “effective introductory and concluding paragraphs (interest-catching & closing techniques)”, “fresh, surprising, daring details”, “no logical fallacies”, “demonstration of critical thinking skills”.

On the other hand, as mentioned earlier (see 2.3.), logical fallacies “make the reader (lose) the thread of the argument when reading a text attentively” (Wikborg 1990: 133), so they could also be moved into the category of coherence.

3.2 Coherence

Grammar/secondary school students receive the highest score for “organization and cohesion” of their essays if their texts are “well-structured and coherent at the paragraph level (divided into introduction, body, and conclusion)”, “ideas are clearly related to each other”, and “cohesive ties work at the sentence, paragraph, and essay level” according to the MERS (Ilc et al. 2018: 15-16), which overlaps with the following descriptor in the DoES (Sokolov et al. 2014): “coherently organized at sentence, paragraph and essay levels: ideas clearly stated, linked and supported”. All the descriptors listed correspond to the various definitions of coherence in the relevant literature (see 2.2. above), the most important point being that a coherent text reads smoothly; in other words, it does not contain coherence breaks.

Regardless of the criterion of clarity of thought, the category of coherence lacks the descriptor covering irrelevant supporting points, which is part of the evaluation of content in the scoring scales under examination. Apart from this, the DoES are also more comprehensive in the category of coherence:

Table 3: Department of English, Faculty of Arts, University of Ljubljana, Scales (Sokolov et al. 2014): Coherence

Grade	Coherence
10	<ul style="list-style-type: none"> • Coherently organized at sentence, paragraph and essay levels: ideas clearly stated, linked and supported; no coherence breaks; • clear thesis statement; • logical sequencing; • makes full and appropriate use of a variety of organisational patterns and a wide range of connectors and other cohesive devices; • well-balanced paragraphing; • reads fluently.

Obviously, university-level examiners explicitly require students to produce thesis statements, something that matura examiners might also consider worth taking onto account. “Logical sequencing” and “reads fluently” are descriptors that refer to the clear outline of ideas, while the descriptor “a variety of organizational patterns and a wide range of connectors and other cohesive devices” reflects the higher expectations at the university level. “Well-balanced paragraphing” could refer to a roughly equal length of

paragraphs reflecting a balanced argument, which again leads us back to content considerations. If the descriptor were more specific, this ambiguity could be avoided. Ute Knoch (2007: 106) may be right that “one reason for (...) vague descriptions of coherence might lie in the rather vague nature of coherence”, but this does not mean that they cannot be improved – specific examples from students’ authentic writing to complement the scoring scales are a possible solution (Sokolov 2014: 190).

4 A CASE STUDY

4.1 Basic Data and Methodology

The aim of the case study, conducted in October 2021, was to gain insight into secondary school teachers’ perceptions of a selected set of seven descriptors from the DoES (Sokolov et al. 2014) in light of the somewhat blurred dividing line between the categories of content and coherence, with a focus on the following questions: In which of the two categories would teachers place the listed descriptors? Does their decision match the category in which they are actually used? Do their paraphrases of three of the seven descriptors confirm their classifications? Do respondents find the descriptors specific enough to be helpful and to make the relevant category easily identifiable?

The DoES, which are composed of different rating scales (Baš et al. 1996; Brown/Bailey 1984 in Brown 2004: 244–245; IB workshops in English B 1992; Jacobs et al. 1981; Spanders et al. 1990), were chosen because they are more comprehensive than the MERS and can be used to revise the latter, providing more specific rating criteria and thus contributing to a more objective and standardized assessment. On the other hand, they partially overlap with the MERS, so they were easier for the survey participants to relate to.

The participant sample comprised 46 secondary school teachers who are not external examiners but use the criteria as part of their regular teaching duties to mark their students’ written compositions, 69% in their original form and 31% in a slightly adapted form. The study was conducted by the researcher using **a questionnaire**.

The main **limitation** of the study is the small sample of teachers who participated in the survey. Since the research findings are not statistically representative, it is not possible to draw conclusions relevant to a broader Slovene context of the situation, let alone even wider contexts. On the other hand, even case studies involving a limited situational context and a modest number of participants illuminate features of other similar cases in the field they study (Richards 2011: 209; cf. Vogrinc 2008: 76–77; Weir/Roberts 1994: 62), which means that they provide food for thought beyond an individual teacher’s experience and need for self-reflection.

4.2 Results and Discussion

4.2.1 Content or Coherence?

First, the participants were asked to rank a selection of seven descriptors in either the content or coherence category. The results are shown in Table 4. The actual categories in the DoES are indicated in the last column. The majority of the descriptors were taken from the content category, which is the more subjective of the two categories and therefore even more difficult to score.

1. Under which category do the following criteria descriptors used to assess longer written compositions fall: content or coherence?

Table 4: Questionnaire: Content or Coherence?

Descriptor	Content	Coherence	Actual category
no logical fallacies	41% (19)	59% (27)	Content
largely relevant	98% (45)	2% (1)	Content
original insights into the question	98% (45)	2% (1)	Content
unclear or non-existent thesis statement	37% (17)	63% (29)	Coherence
demonstration of critical thinking skills	83% (38)	17% (8)	Content
effective introductory and concluding paragraphs	33% (15)	67% (31)	Content
well-balanced paragraphing	2% (1)	98% (45)	Coherence

Logical fallacies are included in the content category in the actual DoES, where they are classified by less than half the respondents, and it has already been established that the majority (59%) are not wrong to assign them to the coherence category. Although the absence of logical fallacies is a prerequisite for the content of an essay to be convincing and impressive (see 2.1.), it is also true that errors in logical reasoning make readers think about the “logic” before rejecting it, which corresponds to the definition of a coherence break (see 2.3. and 3.2.). In other words, both sets of respondents are right, which is, ironically, bad news. Fortunately, there is a solution – a consensus on where the descriptor should be located. Another option would be to combine the two categories into one.

The descriptors “original insights into the question” and “demonstration of critical thinking skills” are predominantly interpreted as content descriptors, by 98% and 83% of respondents, respectively, which is consistent with the original classification in the scoring scales. Similarly, “well-balanced paragraphing” is understood as a feature of a text related to organization/coherence by 45 of the 46 survey respondents. Despite the high level of agreement among respondents in relation to the relevant categories, it is still

a mystery to me why eight respondents (17%) believe that critical thinking skills could play a role in creating coherence in a text.

The results regarding “effective introductory and concluding paragraphs” are less surprising: two-thirds of the respondents to the survey assign the descriptor to the coherence category, while the actual users of the scales perceive it as an aspect to be evaluated in the content category. The additional explanation in parentheses (“interest-catching & closing techniques”) in the actual scales makes it clear that the pre-modifier “effective” refers to the content category, mainly because of the focus on two specific paragraphs rather than the general paragraph structure in an essay. On the other hand, the constructors of the ReadWriteThink scales (in Bean 2011: 271) place the descriptor “Writing includes a strong beginning, middle and end” in the organization category. Obviously, the structural elements of a text can also be “effective” and “strong”; in order for raters to perceive them as belonging to the content category, the additional indication in parentheses is essential. Another explanation for the results is the absence of this or a similar descriptor in the MERS – and thus adding it would definitely be worth considering.

Finally, 63% of the respondents are aware that a clear thesis statement plays an important role in establishing coherence, while 37% assign this descriptor to the category of content. The relatively high percentage is also due to the fact that there is no mention of thesis statements in the MERS, which should be changed when they get revised. Announcing the topic of the essay and the aspect on which the writer is going to focus not only conforms to the conventions of Anglo-Saxon writing, but also helps young writers to stay focused and write coherently. On the other hand, the topic of a written paper is the very essence of its content, so it would be unacceptable to claim that it is wrong to categorize the descriptor under content.

In summary, the respondents agree on the classification of descriptors that they are either more or less familiar with from the MERS (“largely relevant” and “well-balanced paragraphing”) or that are easily assigned to a particular category (“original insights into the question” and “demonstration of critical thinking skills”), while they tend to disagree on the classification of descriptors that are either ambiguous/vague (“effective paragraphing”) or lie on the blurred dividing line between the two categories (“logical fallacies”). In the case of the descriptor “unclear or non-existent thesis statement”, 37% of the respondents do not even seem to be familiar with the concept of a thesis statement, which explains why they assign it to the wrong category.

4.2.2 Specific enough?

Secondly, the participants were asked to indicate whether or not they found the seven descriptors specific enough. Six out of the 46 participants skipped this question. The results from 40 respondents are thus shown in Table 5.

2. Are the descriptors specific enough?

Table 5: *Questionnaire: Specific Enough?*

Descriptor	Specific enough	Not specific enough
no logical fallacies	55% (22)	45% (18)
largely relevant	48% (19)	53% (21)
original insights into the question	68% (28)	33% (12)
unclear or non-existent thesis statement	70% (28)	30% (12)
demonstration of critical thinking skills	78% (31)	23% (9)
effective introductory and concluding paragraphs	78% (31)	23% (9)
well-balanced paragraphing	83% (33)	18% (7)

What is immediately striking is the fact that there is not a single descriptor that all respondents find specific enough. The lowest percentage of responses in the not-specific-enough column is 18% (i.e. seven respondents) and is attributed to the descriptor “well-balanced paragraphing”, which is still surprising as it produced a much more consistent result in 4.2.1 (98% agree that it belongs to the coherence category). Other descriptors with high agreement in 4.2.1, namely “largely relevant”, “original insights into the question” and “demonstration of critical thinking skills”, are also criticized as not being specific enough by 21 (53%), 12 (33%) and nine (23%) of the survey respondents, respectively.

On the other hand, “demonstration of critical thinking skills” is considered specific enough by 31 respondents (i.e. 78%). Another high score was given to the descriptor “effective introductory and concluding paragraphs”, although in reality this item is rather ambiguous (see 4.2.1).

In two cases, the respondents’ assessment of whether the selected descriptors are specific enough might reflect their ranking of the same descriptors from earlier (see 4.2.1). First of all, almost half the respondents, 45%, think that the descriptor “no logical fallacies” is not specific enough, which reflects their uncertainty as to which category it should be assigned to. The same is true for the descriptor “unclear or non-existent thesis statement”: 37% of respondents think it belongs in the content category of content, and 30% say it is not specific enough. Of course, it is impossible to determine how many of the former are included in the 30%, but the result indicates some uncertainty on the part of survey respondents, which, as noted above, is due to the fact that some are unfamiliar with the existence and role of the thesis statement.

The most surprising result seems to be that more than half of the participants (53%) consider the descriptor “largely relevant” not specific enough, as 45 out of 46 (98%) agree (see 4.2.1 above) that it belongs to the content category, which implies a more solid common

ground, especially since the descriptor overlaps with the one in the MERS and has been regularly used by teachers for years. Overall, the results show considerable uncertainty among teachers, suggesting that the descriptors would need to be revised to be more helpful.

4.2.3 Interpretations of Selected Descriptors

Finally, the participants were asked to interpret three out of seven descriptors to shed light on their understanding and categorization (see 4.2.1). Not all the responses are shown in Table 6, as some were given more than once or repeated the same ideas in insignificantly different ways. The responses have thus been grouped according to the underlying philosophy they share. I also distinguish between unproblematic interpretations and those that are too vague or could be argued about (indicated by question marks in brackets). The one incorrect answer is marked by a cross in brackets.

3. How do you interpret the descriptors listed below?

Table 6: *Questionnaire: Interpretation of Descriptors*

Descriptor	Interpretation
<u>Example</u> : fairly one-sided	<i>presents arguments only for or against a controversial topic</i>
original insights into the question	<ul style="list-style-type: none"> • Unexpected, fresh, unique, creative, in-depth, intelligent, individual ideas; new, out-of-the-box solutions/arguments; • interesting (?); thought-provoking (?); • his or her own point of view/attitude (?); critical thinking (?).
demonstration of critical thinking skills	<ul style="list-style-type: none"> • Different perspectives, evaluating validity and strength of arguments; critical analysis; doubts the obvious; intelligent & insightful/sees the big picture (?); • goes beyond generally expected interpretation (?); sharp & innovative ideas (?); persuasive arguments (?); deep understanding of the topic (?); • logical connections and development of ideas (x).
effective introductory and concluding paragraphs	<ul style="list-style-type: none"> • Catching attention & thesis/re-stating the thesis & opinion or solution/food for thought (no new ideas); relevant to the title; • impressive, out of the ordinary; • concise & to the point; • well-balanced.

Many respondents showed a clear understanding of what the descriptor “original insights into the question” implies. However, some interpretations were rather vague (possibly because 33% of respondents find the original descriptor vague – see 4.1.2 above): are “interesting” and “thought-provoking”, “personal” and “critical” ideas necessarily synonymous with “original insights”? In general, some respondents do not distinguish

between the descriptor from the second box in Table 6, i.e. “demonstration of critical thinking skills”, which is reflected in paraphrases such as “goes beyond generally expected interpretation” and “sharp & innovative ideas”, and “critical thinking” in the first box of Table 6. Of course, critical thinking is likely to be original, and original thinking can also be critical – but not necessarily so. As both descriptors fall into the same category, the fine distinction is not of immense importance, although separate descriptors are helpful as they specify what vaguer descriptors in the MERS, such as “convincing, in-depth and balanced supporting points” (Ilc et al. 2018: 15), merely imply.

“His or her own point of view” may or may not be original, while “persuasive arguments” and “deep understanding of the topic” are not necessarily critical. In other words, some interpretations of the descriptor “demonstration of critical thinking skills” are vaguer than the descriptor themselves (while 23% of the respondents complain that the descriptor is not specific enough – see 4.1.2 above), which does not add any clarity or allow for a simpler interpretation. The interpretation of “logical connections and development of ideas” as equivalent to “demonstration of critical thinking skills” is incorrect – logical connections fall under the category of coherence.

The interpretations of the last descriptor in Table 6 are surprising – all of them are acceptable to begin with. Most of them refer to the content, although 31 (67%) of the respondents assign it to the coherence category and nine (23%) think it is not specific enough. The last interpretation, namely “well-balanced”, is the only one that (probably) refers to coherence – it corresponds to the last item on the list in 4.1.1 and 4.1.2, and merges two descriptors into one. If we agree that “well-balanced paragraphing” also includes (“effective”) introductory and concluding paragraphs, this explanation works, too – and brings us back to the beginning: the dividing line between the two categories is often blurred and needs to be regularly negotiated and defined.

In the future, further research with secondary school teachers who are external examiners would be welcome to investigate the extent to which their experience as raters in a large-scale, high-stakes examination involving standardization of marking and wider use of the MERS influences their interpretation of the analytic descriptors, and whether they are perceived by this group of individuals as less vague than by raters without such experience.

5 CONCLUSION

Theoretical considerations of content and coherence are reflected in the confusing situation in the field of analytic scoring scales. In addition, not only do we read texts differently, but we also interpret descriptors differently. Ute Knoch (2007: 109) quotes Watson Todd et al. (2004), who “argue that while analytic criteria are intended to increase the reliability of rating, descriptors (...) inevitably require subjective interpretations of the raters and might lead to confusion” (cf. Shaw/Weir 2007: 145), a conclusion that is also

confirmed by the case study presented in the present article. It is thus essential that raters working in the same educational context have a shared and standardized understanding and use of the related assessment criteria.

As standardization proceeds, there will be opportunities to further develop the scoring scales by creating grading guides supplemented with specific authentic examples of (portions of) student work, by revising and/or supplementing ambiguous descriptors, by adding new ones, and/or by possibly merging the categories of content and coherence into a single, more authentic, category.

Admittedly, there are some writing instructors who raise “philosophical” objections to analytic assessment “on the grounds that writing cannot be analysed into component parts. They argue that ideas cannot be separated from organization, or clarity of expression from clarity of thought” (Bean 2011: 270). Although such considerations have merit, foregoing analytic assessment would result in students receiving less detailed feedback on their work and thus learning less. Since writing instruction is an essential component of assessment, analytic scoring should be discussed, standardized, and developed rather than replaced with holistic assessment. On balance, however, a combination of the two, as suggested by John C. Bean (2011: 280-282), seems to be the best choice: holistic assessment to do justice to the text as a whole, followed by analytic assessment as a test of the former’s reliability (prompting the teacher to revise the assessment if the gap between the two is large) and as detailed and helpful feedback for students. This process may seem time-consuming, but the approach definitely leads to more reliable and rewarding assessment of student writing.

BIBLIOGRAPHY

- ALARO, *Abeyayehu Anjulo* (2020) An Assessment of Cohesion and Coherence in Students’ Descriptive and Narrative Essays. *Journal of Literature, Languages and Linguistics* 64, 41–46.
- BACHA, Nahla (2001) Writing evaluation: what can analytic versus holistic essay scoring tell us? *System* 29, 371–383. DOI:10.1016/S0346-251X(01)00025-2.
- BAE, Jungok (2001) Cohesion and Coherence in Children’s Written English: Immersion and English-Only Classes. *Issues in Applied Linguistics* 12 (1), 51–88. <https://doi.org/10.5070/L4121005043>.
- BAŠ, Ivica/Saša BENULIČ/Margaret DALRYMPLE/Vineta ERŽEN/Soča FIDLER, Maja GRABAR/Meta GROSMAN/Aleša JUVANC/Smiljana KOMAR/Cvetka SOKOLOV/ Rastislav ŠUŠTARŠIČ (1996) *Angleščina pri maturi: Kako se uspešno pripravimo na preizkus znanja iz angleškega jezika*. Ljubljana: Državni izpitni center.
- BEAN, John C. (2011) *Engaging Ideas: The Professor’s Guide to Integrating Writing, Critical Thinking and Active Learning in the Classroom*. 2nd edn. San Francisco: Jossey-Bass.

- BRIESMASTER, Mark/Paulo ETCHEGARAY (2017) Coherence and cohesion in EFL students' writing production: The impact of a metacognition-based intervention. *Íkala, Revista de Lenguaje y Cultura* 22 (2), 183–202. DOI: 10.17533/udea.ikala.v22n02a02.
- BROWN, H. Douglas (2004) *Language Assessment: Principles and Classroom Practices*. New York: Longman.
- BROWN, H. Douglas/Priyanvada ABEYWICKRAMA (2019) *Language Assessment: Principles and Classroom Practices*. 3rd edn. Hoboken: Pearson Education ESL.
- BROWN, Gillian/George YULE (1983) *Discourse Analysis*. Cambridge: Cambridge University Press.
- CARREL, Patricia L (1982) Cohesion is not coherence. *TESOL Quarterly* 16 (4), 479–488. <https://doi.org/10.2307/3586466>.
- CARTER, Ronald (1993) *Introducing Applied Linguistics*. Harmondsworth, Middlesex: Penguin.
- CUSHING WEIGLE, Sara (1998) Using FACETS to model rater training effects. *Language Testing* 15, 263–287. <https://doi.org/10.1177/026553229801500205>.
- CUSHING WEIGLE, Sara (2010) *Assessing Writing*. Cambridge: Cambridge University Press.
- DOUGLAS, Yellowlees (2015) *The Reader's Brain: How Neuroscience Can Make You a Better Writer*. Cambridge: Cambridge University Press.
- ECKES, Thomas (2008) Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing* 25 (2), 155–185. <https://doi.org/10.1177/0265532207086780>.
- ELBOW, Peter (1996) Writing Assessment: Do It Better, Do It Less. E. M. White/W. D. Lutz/S. Kamusikiri (eds.), *Assessment of Writing: Politics, Policies, Practices*. New York: The Modern Language Association of America, 120–134.
- ENKVIST, Nils Erik (1990) Seven Problems in the Study of Coherence and Interpretability. U. Connor/A. M. Johns (eds.), *Coherence in Writing: Research and Pedagogical Perspectives*. Alexandria: TESOL, 11–28.
- GLENN, Cheryl/Melissa A. GOLDTHWAITE (2014) *The St. Martin's Guide to Teaching Writing*. 3rd edn. Boston/New York: St. Martin's.
- HALLIDAY, M. A. K./Ruqaiya HASAN (1976) *Cohesion in English*. London/New York: Longman.
- HOEY, Michael (1991) *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- HOLDSTEIN, Deborah H. (1996) Gender, Feminism, and Institution-Wide Assessment Programs. E. M. White/W. D. Lutz/S. Kamusikiri (eds.), *Assessment of Writing: Politics, Policies, Practices*. New York: The Modern Language Association of America, 204–225.
- HUGHES, Arthur (1989/2003) *Testing for Language Teachers*. 2nd edn. Cambridge: Cambridge University Press.

- HYLAND, Ken (2002) *Teaching and Researching Writing*. Harlow/London/New York: Longman.
- HYLAND, Ken (2003) *Second Language Writing*. Cambridge: Cambridge University Press.
- IB (1992) *Workshops in English B for New Teachers: Working Materials*. Cardiff: International Baccalaureate Europe.
- ILC, Gašper/Alenka KETIŠ/Aleksandra KOMADINA/Ana LIKAR/Simona MEGLIČ/Irena ZORKO NOVAK (2018) *Angleščina: Predmetni izpitni katalog za splošno maturo*. Ljubljana: Državni izpitni center.
- JACOBS, Holly L./Stephen A. ZINKGRAF/Deanna R. WORMUTH/V. Faye HARTFIEL/ Jane B. HUGHEY (1981) *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.
- KNOCH, Ute (2007) 'Little coherence, considerable strain for reader': A comparison between two rating scales for the assessment of coherence. *Assessing Writing* 12 (2), 108–128. <https://doi.org/10.1016/j.asw.2007.07.002>.
- McNAMARA, Tim (2000) *Language Testing*. Oxford: Oxford University Press.
- NUNAN, David (1993) *Introducing Discourse Analysis*. Harmondsworth: Penguin.
- PARKS, A. Franklin/James A. LEVERNIER/Ida MASTERS HOLLOWELL (1991) *Structuring Paragraphs: A Guide to Effective Writing*. 3rd edn. New York: St. Martin's Press.
- RICHARDS, Jack C./John TALBOT PLATT/Heidi PLATT (1992) *Dictionary of Language Teaching and Applied Linguistics*. 2nd edn. Harlow: Longman.
- RICHARDS, Keith (2011) Case Study. E. Hinkel (ed.), *Handbook of Research in Second Language Teaching and Learning*. 2nd edn. New York/London: Routledge, 207–221.
- SHAW, Stuart D./Cyril Weir (2007) *Examining Writing*. Cambridge: Cambridge University Press.
- SKOUFAKI, Sophia (2020) Rhetorical Structure Theory and coherence break identification. *Text & Talk* 40 (1), 99–124. <https://doi-org.nukweb.nuk.uni-lj.si/10.1515/text-2019-2050>.
- SOKOLOV, Cvetka (1999) *Pisni sestavek pri študentih angleščine* (MA Thesis). Ljubljana: Filozofska fakulteta.
- SOKOLOV, Cvetka (2000) Pisna zmožnost – motnje v koherenci. In: Meta GROSMAN (ed.) *Angleščina – prenovi na pot*. Ljubljana: Zavod RS za šolstvo, 97–136.
- SOKOLOV, Cvetka (2013) *Pomen standardizacije ocenjevanja pisni sestavkov pri poučevanju angleščine kot tujega jezika/The Role of Standardization in Assessing Writing in Teaching English as a Foreign Language* (PhD Thesis). Ljubljana: Filozofska fakulteta.
- SOKOLOV, Cvetka et al. (2014) *Written Composition: Analytic Scoring Scales*. Unpublished. Ljubljana: Filozofska fakulteta.

- SOKOLOV, Cvetka (2018) Analiza zgledov meril za razčlenjevalno ocenjevanje razlagalnih in utemeljevalnih pisnih sestavkov. *Vestnik* 10 (1), 169–186. DOI: 10.4312/vestnik.10.169-186.
- SPANDEL, Vicki/Richard J. STIGGINS (1990) *Creating Writers: Linking Assessment and Writing Instruction*. New York/London: Longman.
- VOGRINC, Janez (2008) *Kvalitativno raziskovanje na pedagoškem področju*. Ljubljana: Pedagoška fakulteta.
- WEIR, Cyril/Jon ROBERTS (1994) *Evaluation in ELT*. Oxford UK/Cambridge USA: Blackwell.
- WHITE, Edward M (1996) Power and Agenda Setting in Writing Assessment. E. M. WHITE, W. D. Lutz/S. Kamusikiri (eds.), *Assessment of Writing: Politics, Policies, Practices*. New York: The Modern Language Association of America, 9–24.
- WIDDOWSON, Henry (1978) *Teaching Language as Communication*. Oxford: Oxford University Press.
- WIKBORG, Eleanor (1990) Types of Coherence Breaks in Swedish Student Writing: Misleading Paragraph Division. U. Connor /A. M. Johns (eds.), *Coherence in Writing: Research and Pedagogical Perspectives*. Alexandria: TESOL, 133–148.
- WILSON, Gerald L./Michael S. HANNA (1993) *Groups in Context: Leadership and Participation in Small Groups*. 3rd edn. New York: McGraw-Hill.
- YU, Guoxing (2007) Student's Voices in the Evaluation of Written Summaries. *Language Testing* 24 (4), 239–572. <https://doi.org/10.1177/0265532207080780>.

POVZETEK

IZZIV OCENJEVANJA VSEBINE IN KOHERENCE

Pri analitičnem ocenjevanju pisnih sestavkov je ocenjevanje vsebine in koherence velik izziv. Bralci besedila namreč interpretirajo po svoje, v skladu s svojim znanjem, izkušnjami, vrednotnim sistemom in drugimi vidiki svoje osebnosti, ki povzročajo pristranskost, pa naj se ocenjevalci še tako trudijo, da bi pisne izdelke vrednotili objektivno. Pod vplivom navedenih dejavnikov berejo tudi opisnike, ki so sami po sebi niz kratkih besedil. Enotnost razumevanja opisnikov otežujeta njihova splošnost in nedoslednost pri njihovem uvrščanju v ustrezne ocenjevalne kategorije, ki sta delno posledica narave pisanja, delno pa neusklajenosti strokovnjakov na raziskovalnih področjih diskurza, teorije pisanja in ocenjevanja pisanja. Tako so ocenjevalci, ki iščejo ustrezna merila za ocenjevanje pisnih sestavkov, soočeni s pojmom, kot je na primer “relevantnost/pri-mernost utemeljitev/dokazov”, ki je v nekaterih merilih za ocenjevanje pisnih sestavkov uvrščen v ocenjevalno kategorijo vsebine, v drugih pa v ocenjevalno kategorijo koherence. Kljub temu je veljavnost in zanesljivost ocenjevanja mogoče izboljšati. Članek se posveča odnosu med vsebino in koherenco, ki ga odražajo definicije obeh pojmov v relevantnih virih in opisniki v dveh sklopih

meril za ocenjevanje pisnih sestavkov v rabi v Sloveniji. Empirični del vsebuje študijo primera, ki vključuje 46 učiteljic in učiteljev na slovenskih srednjih šolah, katerih odzivi na vprašalnik potrjujejo subjektivno razumevanje opisnikov and potrebo po izobraževanju učiteljic in učiteljev na področju uporabe analitičnih ocenjevalnih meril, potrebo po redni standardizaciji ocenjevanja na šoli, kjer poučujejo, in po razmisleku, ali bi bilo opisnike v rabi priporočljivo izboljšati, na primer prilagoditi in/ali dopolniti.

Ključne besede: ocenjevanje pisanja, analitično ocenjevanje, vsebina in koherenca, ocenjevalna merila, opisniki

ABSTRACT

THE CHALLENGE OF ASSESSING CONTENT AND COHERENCE

Content and coherence are the categories most difficult to evaluate fairly when raters use analytic scoring scales. Readers inevitably interpret texts in their own idiosyncratic ways, depending on their knowledge, experience, ethical considerations, and other personal biases that they cannot completely set aside when grading a text. This is also true for descriptors, which are themselves short texts. To make matters worse, due to the very nature of writing but also the lack of consensus among experts in discourse research, writing theory, and writing assessment, descriptors are categorized vaguely and inconsistently. As a result, raters seeking useful evaluation criteria are confronted with descriptors that cover the same concept, such as “relevance”, being categorized in one set of criteria as relating to the content of the written text and in another as belonging to the category of coherence. Nevertheless, the objectivity of the evaluation of written work can be increased. The article examines the relationship between content and coherence, which is reflected in the way the two concepts are defined in the relevant literature, as well as in some descriptors used in two grading scales used in Slovenia. The empirical part of the paper presents a case study involving 46 secondary school teachers, whose responses to a questionnaire confirm the subjectivity of the understanding of individual descriptors and the need for adequate training of teachers in the use of analytic scoring scales, regular standardization in the schools where they work, evaluation of the assessment scales they use and their possible adaptation.

Keywords: writing assessment, analytic scoring, content and coherence, scoring scales, descriptors