

Developing an Argumentative Writing Scale

Nimehchisalem, V., Mukundan, J.* and Rafik-Galea, S.

Department of Educational Studies, Faculty of Educational Studies, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ABSTRACT

Today, with more emphasis on the student writers' awareness of the genres of writing, there is a need for writing scales that are sensitive to the variation in the text types. Numerous writing scales have been developed over the past decades, but new scales are required to be developed as the testing situations vary. As a part of a project that aimed at developing a genre-specific writing scale to evaluate tertiary level argumentative writing, a focus group study was conducted. For this purpose, a pre-existing group of experienced English as a Second Language (ESL) lecturers (n=4) discussed what traits of the argumentative writing skill should be included in a writing scale. They also discussed how much weight should be assigned to each trait. As a result of the study, the subscales of 'task fulfilment', 'content', 'organization', 'vocabulary', 'style', 'grammar' and 'mechanics' (in the order of their importance) were proposed by these experts. The study has implications for teaching-testing of ESL writing skill.

Keywords: Assessing writing, scale development, evaluative criteria, argumentative writing

INTRODUCTION

It is interesting to know that over 30 percent of a typical language instructor's professional life is spent on assessment or assessment-related activities (Cheng, 2001). It is sad, however, that only a few

instructors are aware of the rules of efficient evaluation (Stiggins, 2007). With a growing emphasis on accountability, the significance of systematic and well-informed assessment is accentuated even more than before. Educational systems and their Stake-holders increasingly demand objective results of learners' improvement (Coombe, Al-Hamly & Troudi, 2009).

Likewise, in the area of English as a Second Language (ESL) writing, teachers as well as test developers have always needed

ARTICLE INFO

Article history:

Received: 20 November 2011

Accepted: 27 April 2012

E-mail addresses:

nimechie22@yahoo.com (Nimehchisalem, V.),

jaya@educ.upm.edu.my (Mukundan, J.),

shameemgalea@gmail.com (Rafik-Galea, S.)

* Corresponding author

assessment tools to measure students' performance accurately and consistently. Teachers require instruments that explicitly describe the sub-traits of the writing construct that need to be emphasized to improve their learners' writing performance. Research indicates that any mysterious evaluation of students' written samples increases students' test anxiety and can demotivate them (Brennan, Kim, Wenz-Gross & Siperstein, 2001).

Teaching is not separable from assessment. In Huot's words, "being able to assess writing is an important part of being able to write well" (Huot, 2002, p. 62). The reason is that students will find it hard to revise and edit well if they cannot differentiate between basic and competent writing. Despite its importance, assessment is not a well-liked part of most language instructors' professional career. Most writing teachers and lecturers despise assessment. The idea of 'teaching to the exam', in effect, irritates the most. However, tests can help in assessing learning if they are learner/individual/progress-focused (Hamp-Lyons, 2003). On the significance of writing assessment, Huot (2002) assumes, "in literate activity, assessment is everywhere" (p. 61). A lack of relevant scales, on the one hand, keeps writing lecturers away from professional testing of their learners' writing ability in English. A lack of professional handling of writing tests, on the other hand, discourages testing experts from developing writing scales that most probably will not be used by the writing lecturers. This has created a vicious cycle, which has resulted

in a gap between teaching and testing of writing. The problem can be solved by developing teacher-friendly writing scales and training teachers to use them effectively. ESL writing teachers need to be professional testers so they can teach more efficiently.

When the writing instructors do not have access to evaluative criteria checklists or writing scales, they commonly evaluate their learners' written works impressionistically. This method can be highly subjective and brings about the challenge of evaluators' idiosyncratic judgment (Cooper & Odell, 1999). In other words, two different raters may assign quite discrepant scores to the same piece. Scale-based assessment of writing can aid evaluators to score more reliably (Crusan & Cornett, 2002; Cooper & Odell, 1977).

Employing checklists and scales to evaluate written samples can help the reader guard against rater bias, or the rater's idiosyncratic beliefs about successful writing (Tedick, 2002). Research has indicated that when raters are untrained, they show a tendency to emphasize sentence level accuracy and language skills over other sub-traits of the writing skill, like content and organization (Sweedler-Brown, 1993). This suggests that scales and checklists can increase the validity of the evaluator's judgment by controlling the problem of rater bias.

Writing scales have been developed for decades to provide support for teachers. Writing instruments help instructors to evaluate their students' writing based on a set of descriptors and different levels of

performance (also known as bands). Each band descriptor examines papers focusing on a number of evaluative criteria (like grammar, vocabulary and content). With an eye on the descriptors, the evaluator examines and scores each paper. Although a vast number of writing scales are available in the literature, the researchers found no suitable writing scale to evaluate their students' argumentative pieces in their present context (Nimehchisalem & Mukundan, 2011). Therefore, with the objective of developing a writing scale they started a multi-phased project. In the first phase of this project, one of their main objectives was to determine the evaluative criteria of the scale. This paper presents a focus group study whose results helped the researchers gain an in-depth insight of the writing lecturers' views on the evaluative criteria. However, before the study and its findings are presented, a brief review of the literature will follow.

CONSTRUCT OF ARGUMENTATIVE WRITING SKILL

Before conducting the focus group study, the researchers went through the related literature to identify the evaluative criteria that had to be considered by the scale. A huge body of related literature is available. Different scholars regard varying components of the writing skill as important depending on the testing situations. While in some scales, content, organization, vocabulary, language and mechanics (e.g., Jacobs, Zingraf, Wormuth, Hartfiel and Hughey,

1981) are emphasized, in others, components like relevance and adequacy of content, compositional organization, cohesion, adequacy of vocabulary for purpose, grammar, punctuation and spelling (Weir, 1983) are distinguished.

Sometimes scholars seem to have different views. However, if we look closely at these views, it is evident that there is some sort of general agreement, despite the varying terms used to refer to the same components. To offer an example, content, purpose and audience, rhetorical matters (organization, cohesion, unity), and mechanics (sentence structure, grammar, vocabulary) are recognized as the major components of the ESL writing skill by Reid (1993), whereas Cohen (1994) considers content, organization, register (appropriateness of level of formality), style (sense of control and grace), economy, accuracy (correct selection and use of vocabulary), appropriateness of language conventions (correct grammar, spelling and punctuation), reader's acceptance (soliciting reader's agreement), and reader's understanding (intelligibility of the text) as the essential sub-traits.

Recent research in the area of determining the evaluative criteria for automated writing scales identifies grammar, usage, mechanics, style, organization, development, vocabulary and word length (Attali & Burstein, 2006). More recently and radically, organization and development have been replaced by essay length, a readily measureable dimension of writing by computers (Attali & Powers, 2008).

In the literature, other interesting dimensions of the writing skill can also be observed. However, there has been very little research on these dimensions. One of these dimensions is the writer's intellectual maturity (Odell, 1977). As Lee Odell proposes, it is possible to differentiate more from less mature writing through an investigation of the intellectual processes and their corresponding linguistic cues. The processes include focus, contrast, classification, change, physical context and sequence each of which is defined and provided with examples in Table 1.

Basic writers tend to shift the focus, use contrast, and classify less frequently than more mature writers. They also find it hard to describe changes using accurate

language. Describing the physical context or highlighting sequences throughout the passage would also be challenging for them. A scale can focus on these elements to differentiate between less and more competent student writers. In addition to the learners' ability to write well, their skill in inventing ideas to write their papers should also be considered in the development of a scale for argumentative writing. In this respect, the available models on argumentation can prove helpful. As one of the most practical and accurate models of argument, Toulmin's (1958/2003) Model of Argument, which comprises of six elements, is summarized in Table 2.

As the table shows, Toulmin's model makes it easy to analyze elements of a

TABLE 1
Intellectual maturity (Odell, 1977)

Intellectual process: definition	Linguistic cue	Example (cues underlined)
Focus: focus of attention in a sentence	Grammatical subject	<u>Sue</u> opened the window.
Contrast: discussing what something is not, or how it is different from other items	Although, but, not, despite, yet, etc.	Research does <u>not</u> mean reinventing the wheel.
Classification: showing similarities between two entities, feelings, etc. compared with others	Like, such as, for example, compare, as, etc.	Love is <u>like</u> a banana.
Change: indicating the transformations experienced by individuals	Become, change, turn, grow, etc.	She could see the truth as she <u>grew</u> older.
Physical context: the writer's precise description of a given setting	Nouns referring to geographical locations (cities, countries); objects in physical settings (tree, yard); sensory properties of physical settings (sound of wind in the trees)	The old <u>house</u> was filled with unforgettable memories.
Sequence: highlighting time sequences and logical sequences	Subsequently, consequently, etc.	<u>Consequently</u> , the change made a significant contribution to the development of the country.

TABLE 2
Elements of effective argumentation (Toulmin, 1958)

Element of argument	Description	Example (element underlined)
Claim [C]	statement of the thesis	<u>Smoking is dangerous.</u>
Data [D]	evidence providing proof for C	<u>The reason is that it is cancerous.</u>
Warrant [W]	the principle that bridges D to C implicitly/explicitly, proving the legitimacy of D	<u>Anything cancerous is dangerous.</u>
Qualifiers [Q]	the linguistic cues that show the strength of the C, D or W	Smoking is <u>very</u> dangerous.
Backing [B]	further support for W	<u>Cancer kills millions of people.</u>
Rebuttal [R]	response to the anticipated objections against the arguments	<u>Some may argue, however, that smoking gives them a good feeling.</u>

good argument. This facilitates evaluating argumentative papers as it can explicitly highlight the specific part of the argument that is problematic and needs revision.

In addition to mature and skilful development of arguments, another dimension of argumentative writing that is also considered in the literature is the writer's awareness of the audience to whom the paper is addressed (Ryder, Lei & Roen, 1999). The audience can determine the style. A change in the audience may result in an entirely different paper. The writer's awareness of the audience will account for grounding, i.e., her written piece will cognitively, linguistically and socially be appreciated by her reader (Mäkitalo, 2006). It sounds particularly essential to consider audience awareness in the evaluation of argumentative pieces since it deals with the socio-cultural aspects of the pieces that may finally influence the reader's acceptance or rejection of the argument (Clark & Brennan, 1991). Ryder *et al.* (1999) mention four ways to account for the audience:

- i. Naming moves: addressing the reader using pronouns like 'you' or 'we' or placing them in certain groups like democrats
- ii. Context moves: sharing the background information based on the audience's prior knowledge.
- iii. Strategy moves: connecting to the audience by appealing to their interests, circumstances, emotions to ensure they will keep reading.
- iv. Response moves: anticipating the reader's probable responses and objections.

Because of the importance of audience awareness in argumentative writing, it seems necessary to include these moves in the evaluative criteria, in addition to the preceding dimensions of the writing construct.

Based on these theoretical foundations, the researchers developed a list of criteria (see Appendix 1) that would have to be evaluated by the focus group experts, the

process and outcome of which are discussed in this paper.

With the objective of developing a writing scale for assessing argumentative writing the following research questions were put forward:

1. Which evaluative criteria should the scale include?
2. How important is each evaluative criterion viewed by ESL writing experts?

In order to answer the research questions, the qualitative method was employed, which will be discussed in the next section.

METHOD

The qualitative method involved a focus group study in the form of a semi-structured interview. A feasible way to ensure validity is to have the scale and its criteria moderated before its administration. As Weir (1993, p. 19) points out, a “discussion of tasks and criteria of assessment is in fact a key contribution to achieving valid and reliable testing procedures.” Therefore, it is advisable to consult with the experts in the area to gain an understanding of certain important points that had probably been neglected. One systematic way to do this is through a focus group study. Such a group includes a number of individuals who are native to the research context. A trained researcher, who is also the group leader, elicits the group members’ interactive responses (Ary, Jacobs, & Razavieh, 2002).

Instrument

The evaluative criteria checklist for argumentative writing (Appendix 1) was developed based on a review of the related literature. Three experts were consulted to determine the adequacy of the checklist. As a result of this consultation, a section on personal information was added to the instrument. In addition, the researchers were advised to leave the end of each domain open to explicitly show the participants that they could add as many components as they wished to in any part of the checklist.

The final instrument was a six-point scale Likert style instrument including a brief part on the experts’ personal information and two major sections on ‘lexico-grammatical elements of language’ and ‘content’. The part on personal information elicited information on the experts’ name, rating experience, phone number and email address. The first section was composed of the following components: syntax, usage, mechanics, style, essay length, intellectual maturity. The second section consisted of coherence, cohesion, effective argumentation, and audience. These components were divided into further subcategories. In front of each component, there were numbers that would be marked by the experts to indicate the level of significance of each criterion by assigning it a score from zero to five. Zero signified ‘unimportant’ while five meant ‘very important’. The end of the checklist was also left open where any further evaluative criteria that the experts thought had been neglected could be added.

The first five items, including syntax, usage, mechanics, style and essay length as well as their sub-categories were taken from other similar studies like Attali and Powers (2008). While these items focused on the form, the criteria in items 6 to 11 emphasized the meaning domain of the writing ability. The item on intellectual maturity came from Odell (1977). A review of Harmer (2004) and similar literature resulted in the next two criteria, cohesion and coherence. The next item, i.e. effective argumentation, represented Toulmin's (1958/2003) model. The last two items were concerned with audience awareness and invocation (Ryder *et al.*, 1999).

Focus Group

According to the literature, the size of the focus group relies on the scope of the study and the available resources (Ary *et al.*, 2002). Bloor, Frankland, Thomas and Robson (2002, p. 26) recommend focus groups of between "six and eight participants as the optimum size." Nevertheless, they also mentioned that some studies consisted of 3 to 14 participants. Morgan (1995) asserts focus group discussions with a small number of participants are suitable when the research is of a complex nature and when the participants are experts, who may be offended if they are not granted the freedom to talk as long as they wish. In this study, since the topic is complex and the researchers dealt with experts, only four experts participated in the focus group discussions.

The participants in the focus group knew one another. For a number of reasons a pre-existing group of colleagues was selected for the focus group. According to Bloor *et al.* (2002, p. 22), "Research participants who belong to pre-existing social groups may bring to the interaction comments about shared experiences and events and may challenge any discrepancies between expressed beliefs and actual behaviour and generally promote discussion and debate." They give an example of an excerpt from Kitzinger (1994) where one of the participants reminds the other participant of an experience that is related to the discussion. They argued if these participants were strangers similar interactions would be overlooked. Pre-existing groups can also have practical advantages. For instance, it is relatively easier to bring the group together (Bloor *et al.*, 2002).

The members of the focus group were initially selected from among lecturers of a reputable Malaysian public research university. However, since they were too occupied to agree on a time to participate in the meeting, the researchers were urged to look for ESL writing experts elsewhere. Finally, the participants were chosen from among the writing lecturers a branch campus of a large teaching university located in the state of Melaka, Malaysia. They volunteered to participate in the focus group meetings. They were all females of around 40 years of age, and were all senior lecturers.

Procedure

The focus group was briefed on the project and asked for advice on the criteria. Each participant was given a copy of the checklist (Appendix 1) which was presented to them by a table leader, the first author. Then, they discussed their views on the appropriateness of the criteria and the level of importance of each criterion. The discussion lasted around two hours and was recorded using a high quality voice recorder. Next, the discussion was transcribed. Berk (2006) argues that transcripts be reviewed by the participants and verified as a necessary part of the validation process. Therefore, after it had been transcribed, the focus group discussion was sent to each participant. They read the transcripts, made the necessary revisions and, reverted them to the researchers. These transcripts were read closely and indexed for ease of interpretation. According to Bloor *et al.* (2001, p. 63), analysis makes data “manageable for interpretation ... [by merging] all extracts of data that are pertinent to a particular theme, topic or hypothesis.” The indexed data were interpreted and used by the researchers to further refine the checklist, resulting in a new checklist (Appendix 2).

RESULTS AND DISCUSSION

This section presents the results of the focus group discussion, that are divided into two major parts, each part discussing the findings of one of the two research questions.

Changes to the Evaluative Criteria

As expected, the list of criteria was heavily modified and went through several changes. Some of the criteria were modified; others were discarded while others were added to the list. Each of these changes is discussed respectively in this part.

Some of the terms used in the checklist were modified; for example, the term ‘lexico-grammatical elements’ was changed to ‘grammar’. One of the participants argued the term was too technical. It was likely that novice raters would have problem understanding it. Furthermore, ‘audience’ and its two sub-categories were regarded as appropriate for public speech evaluation rather than argumentative writing rubrics. Therefore, its components went under ‘style’. The first sub-category of audience, ‘demonstrating an awareness of the audience by basing the argument on their...’, was reworded as ‘creating interest in audience’, and together with all its three sub-categories was moved under style.

The component of ‘style’ and its sub-categories also underwent changes. It was defined more broadly as ‘skilful weaving of language’ and was modified to an independent category. It was further divided into ‘creating interest in audience’ and ‘appropriate register’. Further, ‘avoiding repetitious words’ was modified to ‘variety of simple and complex words’ and together with ‘appropriate word/phrase use’ was classified under vocabulary. ‘Avoiding unnaturally long/short sentences’ was also considered irrelevant under style. It

was changed to 'variety of structures' and moved under 'grammar'. Similarly, the subcategories of 'using complex structures' under 'syntax' were simply reworded as 'variety of structures'. In fact, these subcategories, including 'modifying nouns', 'nominalization', 'reduced clauses' and 'inverted sentences', came from James Moffett's (1992) concept on syntactic growth. These components were all condensed into 'variety of structures' since the word 'complex' could have a negative connotation for the rater. The experts also agreed that 'essay length' was irrelevant under 'grammar' and that it had to be moved under another category that was added to the list, called 'task fulfilment'. The term 'essay length' was modified to 'writing over word limit'.

Some of the items were modified to contribute to the flexibility of the checklist. The item 'using complex structures' and its four sub-categories could give the impression to particularly the novice raters that in order to be excellent writers students have to use only complex structures. The item was, therefore, modified to 'variety of structures' and all its subcategories were eliminated.

Because of its importance, 'vocabulary' was categorized as an independent category and its sub-categories were all modified into the more relevant items of 'appropriate word/phrase use,' 'collocations, idioms, figures of speech' and 'variety of simple and complex words'. An independent category was also added for 'mechanics', but its subcategories did not undergo any change.

The first two criteria under 'content', 'coherence' and 'cohesion' were moved to be under a new independent category 'organization'. Meanwhile, the third sub-category of 'content', that is, 'effective argumentation', was reworded as 'development of ideas'. It was also suggested that all the criteria under 'effective argumentation' be dropped because the participants commented the terms in the list were highly technical and would confuse the raters more than helping them. This led the researchers to reduce the six types of appeals. However, the elements of argumentation (claim, data, etc.) were not removed as the researchers regarded them as important elements of the checklist.

Finally, Lee Odell's (1977) 'intellectual maturity' was assumed as more appropriate for text analysis purposes rather than descriptors of a rating scale. In fact, one of the participants considered it more appropriate for narrative and not argumentative mode of writing. Therefore, even though the model has a strong theoretical basis, its subcategories were dropped. They would make the scale complicated, undermining its economy and rater-friendliness. 'Intellectual maturity' was modified to 'maturity of ideas'.

Some of the criteria were eliminated since they either overlapped with others or seemed too technical. 'Avoiding run-on sentences' was discarded because it was the same as 'punctuation'. The component of 'usage' was deleted as it would make the rubrics unnecessarily complicated, but its sub-categories were added to those of

‘grammar’. ‘Invoking the audience’ and its sub-categories were also removed because they sounded more like the criteria used for oral speech evaluation.

Two of the subcategories of ‘vocabulary’, namely ‘word length’ and ‘correct use of confusable words’, were also dropped. Word length was considered an irrelevant element for evaluative purposes of the present scale. This criterion seemed more relevant for automated machine scoring of essays. Confusables were also removed because ‘appropriate word/phrase use’ already included ‘correct use of confusable words’.

There were a few new criteria that were added to the checklist. A criterion that the group considered an integral part of any writing scale was ‘task fulfilment’. As it was suggested, test takers would not succeed in a writing test if they fail to take the task into consideration. Besides ‘writing over the word limit’ another sub-category, ‘covering all the task’ was added under ‘task fulfilment’. In addition, four new categories were added under ‘content’, which included ‘relevance’,

‘development of ideas’, ‘maturity of ideas’, and ‘consistency of stance’, as important features of successful writing.

To sum up, as a result of the focus group discussion, the checklist went through a good deal of change. The new version included seven main categories of task fulfilment, content, organization, vocabulary, style, grammar, and mechanics according to their levels of significance, each of which was further divided into its own subcategories. Appendix 2 presents the second version of the list of the evaluative criteria after the first round of focus group meetings.

The Importance of the Criteria

The focus group also indicated the level of importance for each of the evaluative criteria while they were discussing the necessary changes to the checklist. Table 3 presents the list of criteria as determined by the focus group as well as the means and percentages of their importance levels.

According to the focus group participants, among the seven aspects of writing skill ‘task fulfilment’, ‘content’ and ‘organization’ were all checked as highly

TABLE 3
Focus group results on importance of each criterion

Criteria	Participant				Mean	%
	1	2	3	4		
Task fulfilment	5	5	5	5	5	17
Content	5	5	5	5	5	17
Organization	5	5	5	5	5	17
Vocabulary	4	3	4	5	4	13
Style	4	4	5	3	4	13
Grammar	4	3	3	4	3.5	12
Mechanics	4	3	3	3	3.25	11

important. In other words, it was believed that 17 percent of the total score should go to each of the three criteria. On the other hand, 'vocabulary' and 'style' were considered important, and each would cover 13 percent of the total mark. Finally, 'grammar' and 'mechanics' were regarded as fairly important. This suggests that focus group participants would dedicate 12 and 11 percent of the total score to 'grammar' and 'mechanics', respectively.

An interesting point that should be noted is that the checklist that was given to the focus group (Appendix 1) started with a focus on 'form' and lexico-grammatical elements of language. However, as a result of the focus group discussion, 'content' moved before 'form' (Appendix 2). It can therefore be concluded that, at least for these participants, content and meaning out-weighs form in argumentative writing. This was also observed in the results of the survey where the respondents rated factors like content and task fulfillment as more important than language skills.

CONCLUSION

The paper began with a review of the evaluative criteria to be considered in developing an argumentative writing scale. Based on this review, a checklist was developed. A group of ESL lecturers' views on the appropriateness and importance of the criteria was investigated. The results made the criteria less ambiguous, more relevant and more economical. Neglecting these important data could reduce the validity of

the scale that was to be developed based on the criteria.

The findings of this study can be useful for test and scale developers. They can follow the same procedure to find what counts in the evaluation of a particular area of language ability from the viewpoint of the practitioners in their testing situation. Similarly, writing instructors may apply the list of the criteria to develop checklists to assess their learners' argumentative writing. By doing so, they can systematically diagnose the particular problem areas of their student writers. They may also introduce the criteria to their learners and help them use the checklist as a guide for a peer feedback activities or self-assessment purposes. Research shows that most Malaysian students are unaware of the criteria according to which their written pieces are scored (Ahour & Mukundan, 2009). There is empirical evidence that an implicit method of evaluation can increase learners' test anxiety and lower their motivation (Brennan *et al.*, 2001). Checklists of this type can, therefore, improve the quality of teaching-testing ESL writing (Campbell, 1998).

There is an emphasis on language teachers' professionalization in testing (Bachman, 2000). It is expected that the present study could turn the focus on language teachers' professional development in assessment literacy. Such an improvement can enhance learners' academic achievements (Coombe *et al.*, 2009).

Further research following the same method, but with a different focus group, could shed light on the findings of this research. The focus group participants in this study came from a pre-existing group. A similar study with a purpose-constructed group of participants may lead to different findings. In this regard, Wilkinson (1998) argues that focus group studies with participants that do not know one another may sometimes yield richer results since strangers often speak out more freely about issues that may sound embarrassing for friends. Therefore, it would be interesting to compare the findings of this study with another focus group study with a purpose-constructed group. The findings of this study can also be useful for further investigation in the area of ESL writing instruction. The checklist can be adapted slightly considering the level of students. Then, the effects of using such a checklist on the quality of their writing can be studied.

REFERENCES

- Ahour, T., & Mukundan, J. (2009). Analytic assessment of writing: diagnosing areas of strength and weakness in the writing of TESL undergraduate students. *Iranian Journal of Language Studies*, 3(2), 195-208.
- Ary, D., Jacobs, L. C., & Razavieh, A. (2002). *Introduction to research in education* (6th Ed.). Belmont, CA: Wadsworth.
- Attali, Y., & Burstein, J. (2006). Automated essay scoring with *e-rater* V.2. *Journal of Technology, Learning, and Assessment*, 4(3). Retrieved from <http://escholarship.bc.edu/jtla/vol4/3/>
- Attali, Y., & Powers, D. (2008). *A developmental writing scale*. ETS RR-08-19, Princeton, NJ. Retrieved from <http://www.ets.org/Media/Research/pdf/RR-08-19.pdf>
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1-42.
- Berk, R. A. (2006). *Thirteen strategies to measure college teaching*. Sterling, Virginia: Stylus Publishing.
- Bloor, M., Frankland, J., Thomas, M., & Robson, K. (2002). *Focus group in social research*. London: Sage.
- Brennan, R., Kim, J., Wenz-Gross, M., & Siperstein, G. (2001). The relative equitability of high-stakes testing versus teacher-assigned grades: An analysis of the Massachusetts Comprehensive Assessment System (MCAS). *Harvard Educational Review*, 71(2), 173-216.
- Campbell, C. (1998). *Teaching second language writing: Interacting with text*. Boston: Heinle & Heinle Publishers.
- Cheng, L. (2001). An investigation of ESL/EFL teachers' classroom assessment practices. *Language Testing Update*, 29, 53-83.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: American Psychological Association.
- Cohen, A. D. (1994). *Assessing language ability in the classroom*. Boston: Heinle & Heinle Publishers.
- Coombe, C., Al-Hamly, M., & Troudi, S. (2009). Foreign and second language teacher assessment literacy: Issues, challenges and recommendations, *Cambridge ESOL Research Notes* 38, 14-18. Retrieved on February 12, 2010 from http://www.cambridgeesol.org/rs_notes/rs_nts38.pdf

- Cooper, C. R., & Odell, L., (Eds.). (1977). Evaluating writing: Describing, measuring, judging. III: National Council of Teachers of English.
- Cooper, C. R., & Odell, L. (1999). Introduction: evaluating student writing, what can we do, and what should we do? In C.R. Cooper & L. Odell, (Eds.), *Evaluating writing: The role of teacher's knowledge about text, learning, and culture* (pp. vii-xiii). Urbana, Illinois: National Council of Teachers of English (NCTE).
- Crusan, D., & Cornett, C. (2002). The cart before the horse: Teaching assessment criteria before writing. *The International Journal for Teachers of English Writing Skills*, 9, 20-33.
- Hamp-Lyons, L. (2003). The impact of testing on teaching: Ideologies and alternatives. In G. Poedjosoedarmo (Ed.), *Teaching and assessing language proficiency* (pp.1-20). Singapore: RELC.
- Harmer, J. (2004). *How to teach writing?* Petaling Jaya, Malaysia: Longman Pearson Education Limited.
- Huot, B. (2002). *(Re)Articulating writing assessment for teaching and learning*. Logan, Utah: Utah State University Press.
- Jacobs, H., Zingraf, S., Wormuth, D., Hartfiel, V. F., & Hughey, J. (1981). *Testing ESL composition: A practical approach*. MA: Newbury House Publishers.
- Kitzinger, J. (1994). The methodology of focus groups: the importance of interaction between research participants. *Sociology of Health & Illness*, 16(1), 103-121.
- Mäkitalo, K. (2006). *Interaction in online learning environments: How to support collaborative activities in higher education settings*. Research Reports. Jyväskylä, Finland: Institute for Educational Research. University of Jyväskylä.
- Morgan, D. (1995). Why things (sometimes) go wrong in focus groups? *Qualitative Health Research*, 5(4), 516-523.
- Moffett, J. (1992). *Detecting growth in language*. Portsmouth: Boynton/Cook-Heinemann.
- Nimehchisalem, V., & Mukundan, J. (2011). Determining the evaluative criteria of an argumentative writing scale. *English Language Teaching*, 4(1), 58-69.
- Odell, L. (1977). Measuring changes in intelligence processes as one dimension of growth in writing. In C.R. Cooper and L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 107-132). III: National Council of Teachers of English.
- Reid, M. J. (1993). *Teaching ESL writing*. New Jersey: Regents/ Prentice Hall.
- Ryder, P. M., Lei, E. V., & Roen, D. H. (1999). Audience considerations for evaluating writing. In C. Cooper and L. Odell (Eds.), *Evaluating writing: The role of teacher's knowledge about text, learning, and culture* (pp. 53-71). Urbana, Illinois: National Council of Teachers of English (NCTE).
- Stiggins, R. J. (2007). Conquering the formative assessment frontier, in McMillan, J. (Ed.), *Formative classroom assessment* (pp. 8-28). New York, NY: Colombia University Teachers College Press.
- Sweedler-Brown, C. O. (1993). ESL essay evaluation: The influence of sentence-level and rhetorical features. *Journal of Second Language Writing*, 2(1), 3-17.

- Tedick, D. J. (2002). Proficiency-oriented language instruction and assessment: Standards, philosophies, and considerations for assessment. In Minnesota Articulation Project, D. J. Tedick (Ed.), *Proficiency-oriented language instruction and assessment: A curriculum handbook for teachers (Rev Ed.)*. CARLA Working Paper Series. Minneapolis, MN: University of Minnesota. The Center for Advanced Research on Language Acquisition. Retrieved from http://www.carla.umn.edu/articulation/polia/pdf_files/standards.pdf
- Toulmin, S. (1958/2003). *The uses of argument*. Cambridge: Cambridge University Press.
- Weir, C. J. (1983). *Identifying the language needs of overseas students in tertiary education in the United Kingdom*. Unpublished PhD Thesis, University of London Institute of Education.
- Weir, C. J. (1993). *Understanding and developing language tests*. Hampshire, UK: Prentice Hall International.
- Wilkinson, C. (1998). Focus groups in feminist research: power interaction, and the co-construction of meaning. *Women's Studies International Forum*, 21(1), 111-125.

APPENDIX 1

Evaluative Criteria Checklist for ESL Argumentative Writing

Personal information

<i>Name:</i>	<i>Rating experience:</i>	<i>(years)</i>
<i>Tel:</i>	<i>Email:</i>	

- **Kindly check ✓ the detailed criteria below according to their importance (0-5) in scoring university students’ argumentative essays. You may add any other descriptors/ sub-descriptors you find necessary to the list.**

- 0: Unimportant
- 1: Not very important
- 2: Almost important
- 3: Fairly important
- 4: Important
- 5. Very important

Argumentative writing evaluative criteria	Level of importance					
	0	1	2	3	4	5
1 Lexico-grammatical elements of language						
a. syntax						
i. correct use of pronouns						
ii. correct use of verb forms						
iii. avoiding fragments						
iv. correct use of possessives						
v. avoiding plural/singular noun problems						
vi. avoiding run-on sentences (two or more sentences connected together only with commas)						
vii. avoiding garbled sentences (sentences with confusing meanings due to their disorganized forms)						
viii. subject-verb agreement						
ix. using complex structures						
(1) modifying nouns by adjective/relative clauses (e.g. Travelling is a hobby <u>that can teach great lessons.</u>)						
(2) nominalization by using noun clause for subject (e.g. <u>What really hurts</u> is his ignorance.)						
(3) reduced sentences (e.g. ‘Having entered the room, she turned on the light.’ OR ‘The lamp, one of Edison’s best known inventions, changed the face of the world.’)						

(4) inverted sentences (e.g. 'Little is known about mysteries of the outer space.' OR 'Were one interested, one could try it.')						
(5)						
x.						
b. usage						
i. avoiding wrong/missing articles						
ii. correct preposition						
iii. avoiding wrong word forms (e.g. 'Her father is a <u>cook</u> .')						
iv. correct use of confusable words (e.g. 'advise' and 'advice')						
v. avoiding faulty comparisons						
vi. word length						
iv.						
c. mechanics						
i. spelling						
ii. capitalization						
iii. punctuation						
iv.						
d. style						
i. avoiding repetitious words						
ii. avoiding unnaturally long/short sentences						
iii. appropriate word/phrase use						
iv. wrong use of passive voice						
v.						
e. essay length						
f. intellectual maturity						
i. focus: frequent shifting of the focus, i.e., the grammatical subject						
ii. contrast: focusing on what something is not or how different it is from other things						
iii. classification: labeling people, actions, feelings or ideas compared with other entities						
iv. change: showing how the course of action changes						
v. physical context: describing the setting						
vi. sequence: describing the order in which events occur						
g.						
i.						

Developing an Argumentative Writing Scale

2. content						
a. coherence (internal logic helping readers follow the writer's purpose and line of thought)						
i. transition						
ii. organization (the overall conceptual structure of a text and is related to the <i>effect</i> of the text on the language user. Some organizational conventions include topic sentence, first primary support sentence, secondary support sentences, conclusion or transition sentences)						
iii.						
b. cohesion (linguistic techniques to make sure our prose 'sticks together' and help to bind elements of a text together.; comprises ways of marking semantic relationships such as reference, substitution, ellipsis, conjunction, and lexical cohesion as well as conventions such as those governing the ordering of old and new information in discourse)						
i. repetition of words						
ii. lexical set chains (words in the same topic area)						
iii. grammatical cohesion (pronoun and possessive reference, article reference, tense agreement, linkers, substitution and ellipsis)						
c. effective argumentation						
i. making a claim (or thesis, perhaps with accompanying qualifiers limiting the scope of the argument; e.g., 'Morphine can be dangerous.')						
ii. taking a position (e.g. 'People should be warned against dangers of addiction to Morphine.)						
iii. providing data to support the argument through appeal to						
(1) conventional wisdom; i.e., facts (e.g., 'Morphine is addictive.')						
(2) personal experience (e.g., 'My friend's wife left him because of his addiction to morphine.')						
(3) authority (e.g., 'According to researchers, neonates of mothers addicted to morphine will suffer from cardiac problems.')						
(4) analogy/figurative language (e.g., 'Addiction is like cancer that comes painlessly but kills in the long run.')						
(5) history (e.g., 'Abundant historical evidence indicates addiction is common in societies where disorder prevails.')						
(6) legal rights (e.g., 'Addiction is banned in most of the countries all over the world.')						
(7)						
iv. providing warrants; i.e., bridging claim to data to show the connection between them (e.g., 'Morphine is dangerous because it is addictive.')						
backing to show the logic used in the warrants is good in term of realism as well as theory (e.g., 'There is empirical proof that morphine is strongly addictive.')						
v. providing rebuttal by accounting for counter-arguments (e.g., Morphine is dangerous unless it is used for medical purposes.)						

vi.						
d. audience						
i. demonstrating an awareness of the audience by basing the argument on their						
(1) values and perceptions						
(2) attitudes						
(3) background knowledge						
(4)						
ii. invoking the audience through						
(1) naming moves (addressing the reader using pronouns like 'you' or 'we' or placing them in certain groups e.g., democrats)						
(2) context moves (limiting the given background information based on the audience's prior knowledge)						
(3) strategy moves (connecting to the audience by appealing to their interests, circumstances, emotions to ensure they will keep reading since they will find the text engaging)						
(4) response moves (anticipating and accounting for the reader's probable responses and objections)						
(5)						
e.						

APPENDIX 2

Evaluative Criteria Checklist for ESL Argumentative Writing after Focus Group Discussion

Criteria after the focus group meeting (version 2)	Level of importance					
	0	1	2	3	4	5
1. task fulfilment						5.0
a. covering all the task						✓
b. writing over the word limit						✓
2. content						5.0
a. relevance						✓
b. development of ideas						✓
i. making a claim (or thesis, perhaps with accompanying qualifiers limiting the scope of the argument; e.g., 'Morphine can be dangerous.')						✓
ii. taking a position (e.g. 'People should be warned against dangers of addiction to Morphine.')						✓
iii. providing data to support the argument						✓
iv. providing warrants; i.e., bridging claim to data to show the connection between them (e.g., 'Morphine is dangerous because it is addictive.')						✓
v. providing rebuttal by accounting for counter-arguments (e.g., Morphine is dangerous unless it is used for medical purposes.)						✓
c. maturity of ideas						✓
d. consistency of stance						✓
3. organization						5.0
a. coherence						✓
i. transition						✓
ii. rhetorical organization						✓
b. cohesion						✓
i. lexical set chains						✓
ii. grammatical chains						✓
(1) pronoun and possessive reference						✓
(2) article reference						✓
(3) tense agreement						✓
(4) linkers						✓
(5) substitution						✓
(6) ellipsis						✓
4. vocabulary					4.0	
a. appropriate word/phrase use						✓
b. use of collocations, idioms, figures of speech				✓		
c. variety of simple and complex words					✓	

5. style (skilful weaving of language)					4.0	
a. creating interest in the audience					✓	
i. values and perceptions						
ii. attitudes						
iii. background knowledge						
b. appropriate register					✓	
6. grammar				3.45		
a. pronouns				✓		
b. verb forms (tenses, passive voice)						✓
c. avoiding fragments					✓	
d. Possessives				✓		
e. plural/singular forms (qualifiers and determiners)				✓		
f. avoiding garbled sentences (sentences with confusing meanings due to their disorganized forms)					✓	
g. subject-verb agreement						✓
h. variety of structures				✓		
i. articles			✓			
j. prepositions				✓		
k. adjectives (comparatives and superlatives)				✓		
7. mechanics				3.33		
a. spelling (errors resulting in readers' confusion and hinder communication)					✓	
b. capitalization			✓			
c. punctuation					✓	