

Determining progress in writing competency by assessing students' argumentation

Sofia Cassel Aletta Nylén
Department of Information Technology
Uppsala University
Sweden

Abstract—A problem when it comes to evaluating the quality of education in professional competencies, such as writing skills, is being able to detect and measure progression. We have previously defined course level based learning outcomes for academic writing competency in computer science; these are used in a writing across the curriculum (WAC), in the discipline (WID) program. However, in order to assess whether the program is effective, i.e., that the participating students' writing skills progress throughout the education, we need a different set of criteria. Such criteria must capture the quality of the text from an academic perspective. They must also be easy to evaluate, and it must be possible to compare evaluations of different texts.

There are many, sometimes conflicting, definitions of what 'good academic writing' or 'quality' in academic communication is. In this paper, we have defined it in terms of how the material is structured, how well arguments are presented, and how critical thinking is used to strengthen arguments. Following this definition, it is clear that argumentative skills can be used as an indicator of quality in academic communication. Our criteria for measuring writing competency are thus heavily based on assessing students' use of argumentative skills in written texts and are similar to criteria previously used to assess the quality of student participation in classroom discussions.

This paper presents a framework for quantitative and qualitative evaluation of texts written by computer science students. We have related the criteria in our framework to general definitions of academic writing, and to our previously defined goals for writing competencies. The framework provides a grading scheme that can be used to assign a score to a text, corresponding to the level of academic quality exhibited in that text. The results of our framework thus enables comparisons between different student texts. We have used the framework to evaluate writing progression for a group of IT engineering students over three years.

Index Terms—Proficiency assessment, Computer Science Education, Writing across the curriculum.

I. INTRODUCTION

Writing has been widely acknowledged as being one of the key competences that graduating students need to achieve during their studies. In order to enhance students' writing skills, many WAC and WID programs have been initiated in different institutions all over the world. This work aims at providing tools for evaluating the impact of WAC/WID programs by proposing a structured way to analyze student texts in order to observe signs of developing writing proficiency. To be able to compare the quality of texts, assessment criteria, or quality markers, for texts need to be identified.

However, there are many, sometimes conflicting, definitions of what 'good academic writing' or 'quality' in academic com-

munication is, which makes it difficult to evaluate progression in writing skills. In this paper, we have defined 'good academic writing' in terms of how the material is structured, how well arguments are presented, and how critical thinking is used to strengthen arguments. Following this definition, it is clear that argumentative skills can be used as an indicator of quality in academic communication. Our criteria for measuring writing competency are thus heavily based on assessing students' use of argumentative skills in written texts and are similar to criteria previously used to assess the quality of student participation in classroom discussions.

We use the framework presented in this paper to analyze texts written by students participating in the early stages of a WID program during a period of three years. Our analysis did not only show signs of progressive writing skills, but it also revealed some recurring development paths. Knowledge of these paths is interesting for further development of teaching in WID programs.

II. RELATED WORK

This work aims at capturing qualitative development (or lack of development), from an academics point of view, by evaluating evidence of the students' academic writing proficiency at different points in time, the texts they produced at those times. For example, even though many WAC/WID programs have been presented over the past decades, there are relatively few examples of evaluation of their impact on student performance. The success of a WAC/WID program is more often affirmed by faculty and student satisfaction [1] or faculty engagement [2].

Much of the work on development and assessment of writing skills are directed towards early development, in K-9 children [3], or towards second language learners. This work, however, focuses on writing skills development in higher education, where students go from having basic writing skills to developing their skills in writing scientific and academic texts. We believe that a significant factor in writing high quality academic texts is the ability to perform higher order reasoning and presenting this reasoning in text. Under this assumption, assessing writing skills has a lot in common with assessing other competencies where higher order reasoning is required, e.g., argumentation [4], constructive controversy [5] and reflection [6]. Some of these differ from writing due to the fact that they typically occur in dialogical contexts, but some of their objectives, such as convincing others through the

use of correct and valid evidence, are also goals of academic writing.

A. Argumentative competence

In [4], the authors discuss the nature and definition of argumentative competence and present a selection of evaluation criteria for this competence. These criteria have been collected from 97 articles where academics have described assessment of argumentative competence. The criteria are, following their definition, divided into three groups: *metacognitive*, *metastategic*, and *epistemological* criteria. Metacognitive criteria deals with how students are able to use their knowledge, both subject knowledge and knowledge about argumentation. Metastrategic criteria deals with how students are able to use different argumentation elements to support their argument, how well they understand distinctions between types of elements, and how well developed the strategy of this use is in relation to the task. Epistemological criteria deal with students' understanding of norms and objectives at different levels. A more detailed presentation of these criteria is presented in Table I. The criteria are similar to those used in [7] to assess quality of online forum discussions and are also relevant in the context of academic writing, some after minor adjustments.

B. Constructive controversy

Constructive controversy [5] is a type of argumentation where two parties are in disagreement and the two seek to reach an agreement. In educational settings, it is mainly used in connection with group work and the utilisation of the dialogical context is even more prominent than in arguments in general. The authors of [5] discuss "Conditions Determinating the Constructiveness of Controversy". Though many of these conditions are focused on interpersonal interaction, a few are more general and interesting even in the context of writing. These include perspective-taking, described as "reflected in one's capacity to phrase messages so that they are easily understood by others..." which corresponds to, e.g., 3.1.3 in Table I, and rational argumentation.

C. Reflection

The authors of [6] discuss reflection competence from different perspectives. The aim is to strengthen the development of reflection competence in teacher education. They present criteria for recognising four different types of reflection in student texts: descriptive writing, descriptive reflection, dialogic reflection and critical reflection. Descriptive writing is not reflective, the students simply describe events that have occurred or report of literature without providing reasons or justifications. In descriptive reflection, the students provide reasons or justifications, but in a descriptive way. The reflection occurs either based only on one perspective or based on the recognition of multiple perspectives. Dialogic reflection is characterized by the students entering into a personal discourse to explore possible reasons for observed outcomes. Critical reflection is not limited to exploring actions and events, but

also takes into consideration the multiple contexts in which the actions and events are situated.

III. FRAMEWORK FOR WRITING ASSESSMENT CRITERIA

Taking stance in the observation that consistent assessment relies on having well-defined criteria, we define a framework for assessing students' writing. The authors of this work have previously defined learning outcomes for academic writing in computer science. These learning outcomes, however, are not concrete or detailed enough for allowing comparisons of the level of writing proficiency between texts, especially when the texts are assessed by different teachers. In this work, a framework which allows for more detailed criteria is presented with the intention to enable comparisons within a given set of criteria.

We describe academic writing competence as consisting of two types of skills. The first type is of a more technical nature and includes criteria such as "the spelling is correct" and "the report has an introduction", criteria that are either fulfilled or not. In this work, these criteria are referred to as *practical criteria*. The second type of criteria can be fulfilled to different extents. An observation is that these criteria often have a counterpart in criteria for argumentative competence as described above. They are therefore referred to as *argumentative criteria*.

Furthermore, in this framework, assessment of student texts is based on three basic aspects: *structure*, *language*, and *referencing*. Each of the three basic aspects consists of a number of more detailed criteria. As described, the structure aspects, which deal with structure of texts at different levels, consist of both practical and argumentative criteria. Examples of practical structure criteria are "the report has a conclusion" and "the conclusion summarizes the work", while argumentative structure criteria include "each section has a clear story to tell" and "the connection between one paragraph and the next is obvious". Table II shows examples of criteria of both types for the different aspects. A reference to related criteria in Table I appears after each argumentative criteria. Note that this is just a sample of criteria and in no way a complete set defining academic writing competence.

Using our framework, progression can be defined as increasing text quality in at least one of the following dimensions: quantitatively, i.e., increasing the number of criteria that are fulfilled, or qualitatively, i.e., increasing the quality in satisfaction of argumentative criteria.

IV. WRITTEN MATERIAL

Our sample consists of reports collected from three different courses in computer science: one first-year course, one second-year course, and one third-year course. We briefly describe each type of report.

First-year report: The students are instructed to write a report about a given topic. There are several topics to choose from, and they can be broadly categorized as either *technical* or *nontechnical*. The students are introduced to the traditional three-part structure of academic texts, i.e., introduction, 'main

Metacognitive	Structure	1.1.1 length (no of arguments) 1.1.2 complexity (argument structures) 1.1.3 clarity/coherence
	Conceptual quality	1.2.1 conceptual relevance 1.2.2 knowledge integration 1.2.3 originality
	Epistemic quality	1.3.1 use of predef arg structures 1.3.2 use of correct and valid evidence, avoidance of pseudo-evidence 1.3.3 explicit relations between arg elements or premises (diagramming args or verbal justification)
Metastrategic	Meta-element presence	2.1.1 warrant/backing supporting claims 2.1.2 counter-argument or rebuttal for objections raised 2.1.3 qualifiers/meta 2.1.4 clarification 2.1.5 question 2.1.6 explanation 2.1.7 challenge 2.1.8 evaluation 2.1.9 introduction (preparation for arguments that follow) 2.1.10 example (analogy or counterexample) 2.1.11 hypothesis
	Meta-element type	2.2.1 distinction among different types of argument elements
	Meta-task understanding	2.3.1 avoidance of “my-bias” perspectives 2.3.2 theory-evidence coordination 2.3.3 use of strategic sequences of moves
	Meta-task non argument	2.4.1 narration 2.4.2 explanation 2.4.3 fallacies
	Epistemological	Understanding of norms and objectives

TABLE I
EVALUATION CRITERIA FOR ARGUMENTATIVE COMPETENCE

part’, and conclusion. The instructions explicitly state that at least two ‘types’ of references are to be used; examples of such types are web pages, books, and popular science articles. The reports are approximately 3-4 pages in length and written individually.

Second-year report: This report is written by students after having carried out a programming project within the context of a course. The students are instructed to follow the traditional three-part structure of academic texts, i.e., write an introduction, a ‘main part’ describing their project, and a conclusion. In the conclusion, students are additionally required to reflect

	Practical	Argumentative
Structure	three-part structure	each section has a clear story to tell (1.1.3)
	work is introduced in the introduction	the connection between one paragraph and the next is obvious (1.3.3)
	conclusion summarizes the work	
Language and style	spelling is correct	language style is appropriate for the field, the purpose of the text and the intended reader (3.1.3)
	grammar is correct	perspective (subjective/objective) (2.3.1)
Referencing	references contain enough information to allow the reader to find the source	the referred work is of sufficiently high quality (1.3.2)
	references are correctly formatted according to one referencing style	references used to support claims, not just exemplify things (2.1.1)

TABLE II
EXAMPLES OF CRITERIA FOR ASSESSING PROGRESSION IN WRITING

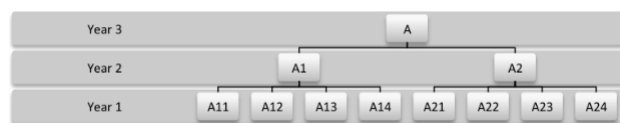


Fig. 1. Example of hierarchies for student group selection.

on their group work. References are not explicitly required, but encouraged. Students work in groups of four or five and write group reports.

Third-year report: This report is the equivalent of a bachelor thesis. It documents a more extensive project than the second-year report, and there are also more requirements on structure and quality of the contents, including a section about related work. This makes it absolutely unavoidable to use references, and also to choose wisely what sources to consult. Students work in groups of two to four students when writing these reports. In this course, the students have access to more scaffolding than in the two previous courses.

We have studied 39 reports written by 28 students, 22 male and 6 female. The students wrote individual reports in year 1. In the second year, the students formed 8 groups writing one report per group and in the third year, we have selected 8 students (5 female) that formed 3 groups writing one report per group. One of the third year groups was all male, one was all female and one was mixed. Hence, the reports from the different years are connected by their authors to form three hierarchies as illustrated in Figure 1. We call these hierarchies Group A, Group B and Group C.

The reports were collected prior to the launch of a WAC program and the analysis hence reveals problems that are now addressed. The analysis is still interesting as it establishes a

baseline that can be used for comparisons with student groups that have taken part in the WAC program.

V. ANALYSIS

To illustrate how the framework can be used to evaluate students' progression in writing proficiency, we used a subset of the criteria in Table II to assess practical and argumentative features of student texts. Six features were selected. These are the practical features *introduction* (work is introduced in the introduction) and *conclusion* (conclusion summarizes the work) and the argumentative features *inter-paragraph flow* (the connection between one paragraph and the next is obvious), *style* (language and style is appropriate for the field, the purpose of the text and the intended reader), *perspective* (subjective/objective) and *references* (the referred work is of sufficiently high quality). Each feature was scored individually and the overall score for a report obtained by adding all of its feature scores. Below, we describe how each feature was scored and observations of different stages of accomplishment occurring in the texts.

A. Scoring

Since the aim of this work is to explore the applicability of the framework rather than to perform an advanced analysis of texts, we have chosen to use rather simplistic scoring criteria. For the practical features, the score is 1 or 0 depending on whether the feature is present or not. The argumentative features are scored according to the level to which they are fulfilled with scores starting at 0 if the feature is not present at all.

Practical features: In the introduction, different levels of academic writing proficiency can be seen in the way students present their topic, in whether they consider it a problem to be solved or simply the heading of a report. We have observed three stages in students' problem formulations. In the first stage, students either do not formulate the problem to be investigated at all, or they formulate it as one or more direct questions. These questions can be rhetorical or directed at the reader. Often, subproblems occur later in the text and they are also formulated as questions to be answered. The text thus acquires a question/answer structure, as if the student is writing down his or her inner dialog when trying to think about and explain the problem.

In the second stage, students can formulate the problem they are interested in, and often also state why it is interesting. Typically, however, they do not state why *that particular problem* would be interesting to solve; they do not take into account other possible ways to achieve the same result.

In the third stage, students state both why the results would be interesting, and why the problem is an interesting one to solve; perhaps even why it is the right problem to solve in order to obtain the desired results. This obviously requires more knowledge about the problem in itself, but more importantly about other investigations into solving similar problems, i.e., an understanding of related work. These three stages are very similar to the first three types of reflection

described in Section II-C meaning that the students reaching the third stage are approaching the competence of critical reflection.

In this work, we have given the score 1 if the introduction includes any description of the context in which the work is relevant, the problem to be solved, the presented solution, or motivation to why the problem or the solution are interesting. For conclusions, the score 1 is given if there is a section towards the end of the report where the work is summarized.

Argumentative features: The feature *inter-paragraph flow* is given the score 0 if there is no flow between paragraphs or if the paragraph splits are arbitrary or nonexistent. In some text, the author tries to remedy the arbitrary structure by adding an (unnecessarily) informative heading to each paragraph. Higher scores are given to texts where each paragraph has one (or two) main topic(s). The score 1 is given if there are references to earlier paragraphs that are not clear. The score 2 is given when there is mostly a flow between paragraphs and the highest score, 3, is used for texts with mostly flow between paragraphs and (albeit colloquial) phrases that push reader to next paragraph, e.g., "Now let's go to..."

The feature *language style* focuses on the level of formality in the text and use of academic wordings and subject appropriate language. The score 0 is used for informal texts. The author writes comfortably about his/her topic in a casual style that resembles spoken language. Colloquial expressions are used, as well as imprecise expressions ('really' hard), and references to unspecified 'things' etc. Sentence boundaries are not important and very long sentences (with several main clauses) are mixed with very short sentences (sometimes only consisting of a lonely sub-clause). The text is difficult to read, but typically makes sense if spoken. Texts that are somewhat informal or that contain attempts at more academic wordings (sometimes resulting in awkward formulations) are given the score 1. These texts are typically still very mixed in terms of style. The highest score, 2, is used for texts that are appropriately formal, where the authors are comfortable with a more academic vocabulary, and use precise wording.

For the *perspective* feature, we consider two perspectives: the subjective and the objective, where the objective scores higher than the subjective. Texts written from an entirely personal viewpoint, often with explicit references to the author as 'I' (or 'we' for group reports) score 0. In these texts, the inclusive pronoun 'we' is sometimes used to indicate the author and any reader of the report. They typically contain unsupported, generalizing statements about how 'we' usually do things, or how 'we' think. The highest score, 2, is used for reports that are written in an objective perspective, with students using references rather than generalizing statements in order to make their point. In texts that score 1, students mix perspectives. Typically, the main part of the report is written from an impersonal viewpoint, while sections that require opinions, arguments, or reflection (such as the introduction, discussion of results, etc.) are still based on personal opinions and sometimes anecdotes.

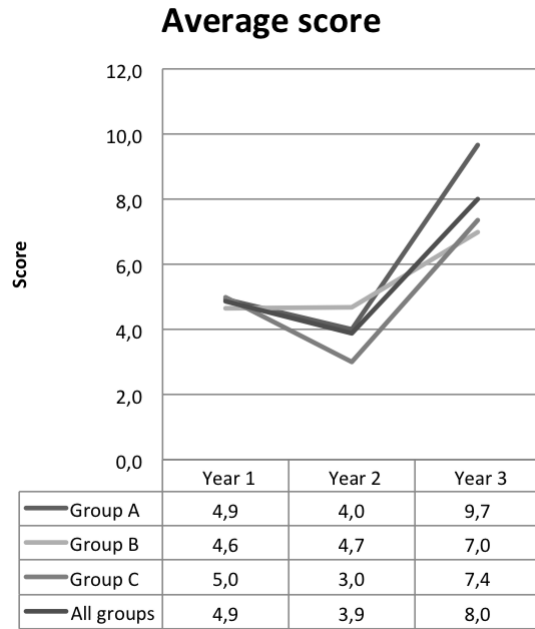


Fig. 2. Average scores for the student groups.

The quality of the *references* is crudely scored by the types of texts referred to. The highest score, 3, is used for scientific work such as books, journals and conference proceedings. Course literature and student theses are given the score 2 and web pages, technical specifications and popular science literature, the score 1. The average score for all work cited in a report is used as the reference score for the report.

VI. RESULTS

The average score for reports in each year is presented in Figure 2. The maximum score for a report is 12. The average score for the year 3 reports are, as can be expected, higher than for the year 1 reports. The year 2 reports, however, have an unexpectedly low score, in most cases lower than year 1. We believe that this can partly be explained by the fact that in year 2, the students were not specifically instructed to cite their sources, resulting in all reports lacking references. As a comparison, Figure 3 shows that when the reference score is omitted, the average score is indeed higher in year 2 than in year 1 in most cases.

The fact that no student group included references without a prompt to do so indicates that the students have neither conceptualized nor internalized referencing. Using references is not yet part of their writing habit or their way of expressing themselves, and they have not understood why it is used. This problem is now targeted in the WAC program.

VII. CONFOUNDING FACTORS IN THE ANALYSIS

There are some possibly ‘confounding factors’ to be considered when analyzing our results. Here, we mention the most important ones.

Average score (without "references")

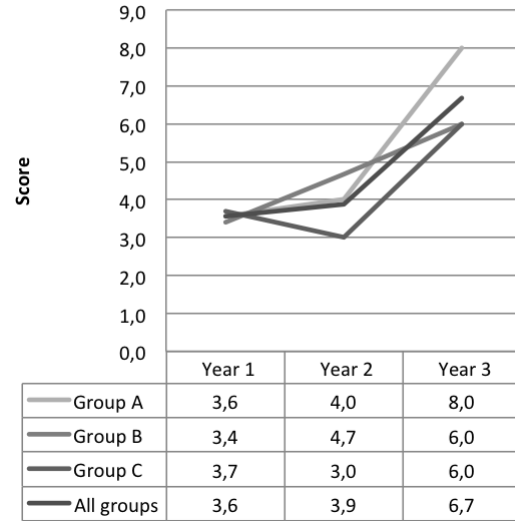


Fig. 3. Average scores for the student groups, excluding the feature “references”.

A. Detailed instructions or progression?

In the second-year reports, students are required to reflect on how the group has functioned, and include these remarks in the conclusion of the report. We believe that these instructions have affected not only the conclusion but also the *introduction* of these reports. Both introduction and conclusion are often written from a subjective perspective (using ‘we’). In addition, students tend to motivate why the problem is interesting in terms of the group members’ (sometimes the group’s, as one entity) personal development. This is worth noting, because the instructions only state that the problem must be clearly formulated. We hypothesize that the instructions signal to the students that successful and conscious working in groups is something generally valued (e.g., by teachers) and that the problem can thus be motivated in terms of group development. This is perhaps further supported by the fact that some groups motivate their problem in terms of ‘learning new things’, which in the context of university education can also be considered generally valued.

In the first-year reports, motivating why the problem is interesting is not mandatory. The instructions state that students are to write a report about a topic, not solve a problem and write a report about the solution. However, some students still appear to (either consciously or subconsciously) treat this task as a kind of problem-solving: we believe that the question/answer structure that some first-year reports exhibit may be the result of this internal process.

B. Report topics

The instructions for the first-year report gave students the option to choose between writing about a more technical subject, a more theoretical subject, or a third subject that allowed

for personal opinions (and could be considered nontechnical). These choices will of course affect the evaluation results. For example, it is natural to assume that a student writing a report about a subject that he/she has very strong opinions about, is more likely to use a subjective perspective, and perhaps try to explicitly draw the reader in by using the inclusive 'we' pronoun.

It is also natural that reports about such different subjects will contain very different types of references. For example, it is more likely that a student writing about a technical subject will use a technical report as a reference.

VIII. CONCLUSIONS AND FUTURE WORK

We have presented a framework for assessing student writing in a structured way, in order to enable text comparison at a group level, focused on quality. These comparisons can then be used to detect progression in writing proficiency. The framework has been used to analyze a sample set of student texts produced by a group of students over a period of three years. Our results indicate progression within the student group as a whole, and also revealed some recurring development paths. Information about in which areas progression occurs and in which areas it does not is valuable for educators seeking to improve students' writing skills. Insights in common development paths also gives indications on how to structure courses to facilitate students' learning to write well.

Assessment of writing in higher education with the aim to discover progression in writing skills is a topic that has not been explored much. More data needs to be analyzed in order to be able to make definitive statements about how students' writing skills can and should progress over the course of their education. Further research is also needed on which assessment criteria are useful for determining progression in writing skills, and on how a suitable set of criteria that define good-quality text should be composed. Our work should be seen as a pilot study in these respects.

REFERENCES

- [1] R. Ochsner and J. Fowler, "Playing devil's advocate: Evaluating the literature of the wac/wid movement," *Review of Educational Research*, vol. 74, no. 2, pp. 117–140, 2004.
- [2] C. O. Erwin and T. M. Zappile, "Organizational response to a university writing initiative: Writing in the disciplines (wid) in an interdisciplinary department," *Double Helix*, vol. 1.
- [3] C. S. Puranik, L. J. Lombardino, and L. J. Altmann, "Assessing the microstructure of written language using a retelling paradigm," *American Journal of Speech-Language Pathology*, vol. 17, no. 2, pp. 107–120, 2008.
- [4] C. Rapanta, M. Garcia-Mila, and S. Gilabert, "What is meant by argumentative competence? an integrative review of methods of analysis and assessment in education," *Review of Educational Research*, vol. 83, no. 4, pp. 483–520, 2013. [Online]. Available: <http://rer.sagepub.com/content/83/4/483.abstract>
- [5] D. W. Johnson and R. T. Johnson, "Energizing learning: The instructional power of conflict," *Educational Researcher*, vol. 38, no. 1, pp. 37–51, 2009. [Online]. Available: <http://edr.sagepub.com/content/38/1/37.abstract>
- [6] N. Hatton and D. Smith, "Reflection in teacher education: Towards definition and implementation," *Teaching and Teacher Education*, vol. 11, no. 1, pp. 33 – 49, 1995. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0742051X9400012U>
- [7] P. Häkkinen, "Multiphase method for analysing online discussions," *Journal of Computer Assisted Learning*, vol. 29, no. 6, pp. 547–555, 2013. [Online]. Available: <http://dx.doi.org/10.1111/jcal.12015>