



How Well Do Student Nurses Write Case Studies? A Cohesion-Centered Textual Complexity Analysis

Mihai Dascalu, Philippe Dessus, Laurent Thuez, Stefan Trausan-Matu

► To cite this version:

Mihai Dascalu, Philippe Dessus, Laurent Thuez, Stefan Trausan-Matu. How Well Do Student Nurses Write Case Studies? A Cohesion-Centered Textual Complexity Analysis. 12th European Conference on Technology Enhanced Learning, EC-TEL 2017 Data Driven Approaches in Digital Education, 2017, Talinn, Estonia. pp.43-53, 10.1007/978-3-319-66610-5_4 . hal-01584796

HAL Id: hal-01584796

<https://hal.science/hal-01584796>

Submitted on 9 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

How well do Student Nurses Write Case Studies? A Cohesion-Centered Textual Complexity Analysis

Mihai Dascalu^{1,2,3}, Philippe Dessus³, Laurent Thuez⁴, & Stefan Trausan-Matu^{1,2}

¹ University Politehnica of Bucharest, Romania

{mihai.dascalu, stefan.trausan}@cs.pub.ro

² Academy of Romanian Scientists, Splaiul Independenței 54, 050094, Bucharest, Romania

³ Laboratoire des Sciences de l'Éducation, Univ. Grenoble Alpes, F38000 Grenoble, France
philippe.dessus@univ-grenoble-alpes.fr

⁴ IFSI, Centre Hospitalier Annecy-Genevois, Metz-Tessy 74374 Pringy,
France

lthuez@ch-annecygenevois.fr

Abstract. Starting from the presumption that writing style is proven to be a reliable predictor of comprehension, this paper investigates the extent to which textual complexity features of nurse students' essays are related to the scores they were given. Thus, forty essays about case studies on infectious diseases written in French language were analyzed using *ReaderBench*, a multi-purpose framework relying on advanced Natural Language Processing techniques which provides a wide range of textual complexity indices. While the linear regression model was significant, a Discriminant Function Analysis was capable of classifying students with an 82.5% accuracy into high and low performing groups. Overall, our statistical analysis highlights essay features centered on document cohesion flow and dialogism that are predictive of teachers' scoring processes. As text complexity strongly influences learners' reading and understanding, our approach can be easily extended in future developments to e-portfolios assessment, in order to provide customized feedback to students.

Keywords: Health Care, Nursing School, Textual Complexity, Infectious Diseases and Hygiene, Case Analysis

1 Introduction

The reflective turn in nurse training has gained popularity and interest, as in any professional fields pertaining to the “helping professions”, such as teachers, midwives, psychological counseling or social work [1]. The instructional models guiding their training have progressively abandoned the *apprenticeship* image, where the trainee has to do what the mentor does or tells. Even though simulations can be used to train nurses, higher-level mentoring models, involving either *reflections* – the trainees understand why they perform certain tasks, and which ones –, or *competencies* – the trainees do what they can, in reference to a set of “best practices”

or a competency framework, are most often promoted to support the building of sound nursing practices [2].

In consequence, and towards a more meaningful articulation between theoretical and practical knowledge, the assessment of complex professional skills is processed through critical thinking-based examinations of case studies [3], or creation of portfolios of actual competencies [4]. This approach of assessment aims at capturing the professional reflection of students when mastering their skills.

In addition, critical thinking has become a key skill in many professional training sectors [5], like nursing. This profession requires a wide range of skills (e.g., patient care, interpersonal skills, hygienic precautions, drug calculations, and safe lifting) [6]. Some of these skills are highly anchored in body and motor experience; others require accurate observations, analysis and problem-solving skills. For instance, the French curriculum of nursing schools requires students to write reflective essays, so-called “situation analyses”, which refer to their professional placements. The main pedagogical goal of this activity is to foster students’ abilities to extract the main variables of the situation, so that they solve problems and elaborate the most adequate solutions. In brief, they become able to use scientific, technical, procedural knowledge in order to develop fully professional nursing abilities. However, as many researchers pointed out [7], developing portfolios or critical thinking without mentoring is useless: students need guidance to extract and analyze relevant pieces of knowledge, manage plans for improvement, and link assessment and practice [8].

Despite its interest in developing professional expertise, the assessment of portfolios or essays stemming from case studies is seldom performed for two reasons. First, the cognitive processes engaged by teachers during assessments are subject to little research [9]. Second, essay grading is time-consuming and there is a limited set of potential computer-based procedures to support this demanding process. Recent advances in Natural Language Processing (NLP) make it possible, at least partially, to automatically assess students’ skills through some proxies, like the textual formulation of their abilities or reflective thoughts on a professional situation. Teachers would use these proxies, once identified, to assess the quality of essays in large-scale educational contexts, like university exams or MOOCs. Moreover, this would encourage course designers to progressively abandon the frequently-used Multiple Choice Questionnaires (also used in nurse training [10]), which are less prone to capture higher-level thinking processes.

Thus, our aim is to create and validate an extensible and adaptive automated method of evaluating student’s case studies. More specifically, our approach is to consider that the analysis of the students’ textual production can predict their teachers’ grades. This approach is in line with the reflective approach, which prescribes that professionals are able to verbalize their thoughts and decisions, and that, in turn, their verbalizations are subject to a fine-grained analysis to predict which competence is acquired. Therefore, our research question is to examine to what extent an automated assessment approach of nurse students’ essays can help teachers assess their professional abilities. Within the conducted analyses, we used *ReaderBench*, a multi-language and multi-purpose system to assess the textual complexity of the students’ essays [11, 12]. Moreover, we chose to focus in this study on the domain of

infectious diseases and hygiene, of crucial importance in nurse training. This domain is closely related to the quality of the care persons receive, their health and their well-being, as well as biology (relationships with infectious agents).

In the rest of this paper, we focus on ways to automatically assess health care training (medical and nurse studies), as well as on textual complexity measures to quantify students' essay quality. Afterwards, we introduce to the main components of our study, followed by results, discussions, and conclusions.

2 Automated Assessment Approaches in Health Care

A posteriori semi-automated e-portfolios assessments are frequent in the literature [13], but they rely on qualitative research-focused systems like *NVivo* [14]. However, systems that rely on more integrated, automated, and quantitatively oriented data are considerably scarcer. *CONSPECT* [15] is a blog-based automated assessor which uses NLP and Network Analysis techniques to evaluate the conceptual development of medical students. The system takes as input students' blog writings and displays a network of the main concepts they used. It also can automatically compare the evolution of the terms used by a given learner to other students, or domain experts.

A more recent study [16] aimed at devising an LMS-based system to provide an automated assessment of e-portfolios, upon raw statistical features like word count or number of images. A first comparison of human vs. machine grades of 12 e-portfolios yielded promising results ($r = .67$). Another study [17] argues that e-portfolios enhanced with learning analytics can potentially increase the quality and efficiency of workplace-based assessment and feedback in professional education.

However, none of the previous approaches models the extent to which teachers are sensible to textual features encountered while reading, nor accounts for more sophisticated and semantically-related textual features.

3 Textual Complexity and Assessment

The complexity of texts, or their level of sophistication, is an important educational issue, either for the selection of texts for reading purposes [18], for understanding academic materials [19], or merely for assessing text difficulty [20]. Despite some attempts [21], little has been done so far to uncover the relationships between the students' writings (e.g., essays, reflective thoughts, portfolios) and the grades that were given by teachers or experts.

Seminal research [22, 23] showed that very shallow textual features of a document (e.g., number of characters, words, sentences, paragraphs and length of words and sentences) are good predictors of human grades. More extensive research on lexical, syntactic, and semantic levels [24] showed that essay quality increases as both lexical and syntactic text levels increase, whereas semantic-based cohesion indices (word or sentence-based) are negatively correlated with essay quality. Moreover, a recent research [25] processed about 560 master and bachelor theses, analyzing a wide range of textual complexity features (from lexical to semantic levels), and linking them to

their assigned grades. The results showed that the correlation between these two variables was low, but this was mainly due to the skewed grade distribution and to the difficulty in selecting the most adequate criteria beforehand, which would best predict the assigned grades.

Since teachers, while scoring an essay, have access to the reading material assigned through the reading task, it is now well documented that its textual features may likely influence their scoring. So far, lexical and syntactic levels' quality is known to positively influence human judgments; more investigations are to be performed on semantic levels (i.e., cohesion-based).

4 Research Question

While perusing students' essays for assessment and scoring purposes, teachers are mostly focused on the usage of domain concepts and the manner in which they are related to the task at hand. Our research question is to understand to what extent teachers are also sensitive to other features, like textual complexity at several levels (lexical, syntactic, semantic, dialogical). To that aim we first computed a wide range of complexity indices, followed by a Discriminant Function Analysis (DFA) to analyze to which extent our model can classify students' grades. As Attali [26] put it, we can consider this large number of complexity indices as "black boxes" that are related to essay quality, though not individually interpretable per se.

5 Method

5.1 Participants

Forty essays written by 1st-year nurse students as case studies of 'infectious diseases and hygiene' were randomly selected. For homogeneity purposes, we excluded essays from repeating students and essays from students having completed medicine studies during the previous year.

5.2 Textual Complexity Assessment with *ReaderBench*

We used *ReaderBench* [11], a multi-language and multi-purpose NLP framework, designed to be an educational helper for students, teachers, and tutors. *ReaderBench* takes as input a wide range of educational productions (e.g., essays, explanations, discussions) and automatically assesses features, like the main concepts used, knowledge-building contributions, comprehension prediction, topic extraction, or textual complexity assessment. *ReaderBench* makes use of Cohesion Network Analysis [27] which harmoniously integrates semantic distances from WordNet with similarity measures derived from semantic models (i.e., Latent Semantic Analysis, LSA, and Latent Dirichlet Allocation, LDA), trained on our custom text corpora. Thus, we gathered a nurse-centered corpus for the analyses to account for the specificity of the vocabulary usage. We selected 9 documents on infectious diseases

and hygiene, of about 273 pages comprising of 133,000 words, compliant with the French nurse training competencies framework. This corpus was added on top of a more general corpus (one-year issues of the French newspaper *Le Monde*; <http://lsa.colorado.edu/spaces.html>), and was used to train new semantic models integrated in the *ReaderBench* framework.

Of particular importance to the rest of this paper is the measure of *document flow*, coined in [28]: a “measure of a document’s structure derived from the order of different paragraphs and of the manner in which they combine to hold the text together”. (id., p. 765) This is an aggregated measure based on the identification of paragraph relationships in terms of semantic relatedness that captures global cohesion. Besides a wide variety of textual complexity indices presented in detail in previous papers [11, 29], *ReaderBench* integrates specific measures derived from the polyphonic model [30], inspired from Bakhtin’s dialogism [31]. According to this model, interanimating ‘voices’, in a generalized way, are coherent points of view over semantically related concepts. Therefore, these indices take into account the distribution of ‘voices’ as well as their co-occurrence patterns [32]. Derived from dialogism, voices are operationalized as semantic chains and can be perceived as recurrent points of view or emerging topics that span throughout the document.

We ran on *ReaderBench* a multi-dimensional analysis of textual complexity indices adapted for French language, integrating classic surface metrics derived from automatic essay scoring techniques, morphology and syntax factors [33], as well as semantics and discourse factors [11]. In the end, subsets of factors were aggregated through a Discriminant Function Analysis in order to predict student performance.

5.3 Procedure

The main characteristics of students’ selected essays are as follows: mean length: 1,342 words ($SD = 293$ words); minimum length: 680 words; maximum length: 2,179 words. Each essay was distributed randomly to one teacher who graded it. Afterwards, the essays were typed and corrected for spelling, followed by their automated assessment with *ReaderBench*.

Table 1. Grader allocation and information on essay grades.

Grader	No. Graded Essays	Grade Range (max: 20)	Mean	<i>SD</i>
A	14	[5.0; 16]	11.3	2.5
B	9	[8.0; 17]	13.5	2.1
C	5	[10.5; 19]	16.4	2.3
D	12	[5.3; 18]	12.6	2.5
Overall	40	[5.0; 19]	12.9	3.5

6 Results

We split the students into two equal-sized groups, namely high-performance students with scores greater or equal to 13 (in France, a [1; 20] scale is used), while the rest were catalogued as low-performance students (see Figure 1 for correspondent frequency histogram). The textual complexity indices from *ReaderBench* that lacked normal distributions were discarded. Pearson correlations were then calculated for the remaining indices to decide whether there was a statistical ($p < .05$) and meaningful relation (at least a small effect size, $r > .3$) between the selected indices and the dependent variable (the students' essay scores). Indices that were highly collinear ($r \geq .90$) were flagged, and the index with the strongest correlation with the essay scores was kept, while the other indices were removed. The remaining indices were included as predictor variables in a stepwise multiple regression to explain the variance in the students' essay scores, as well as predictors in a Discriminant Function Analysis used to classify students based on their performance.

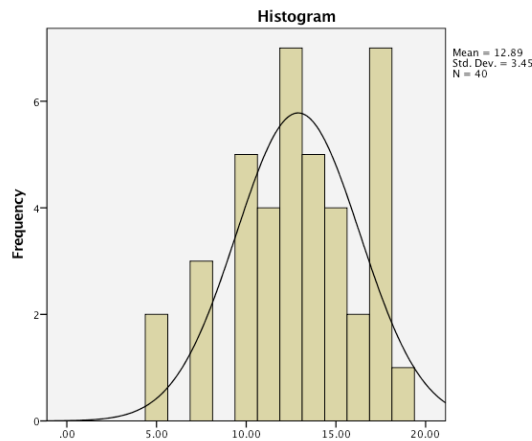


Fig. 1. Essay scores distribution.

Medium sized effects for Pearson correlation coefficients ($.3 < |r| < .5$) were found for *ReaderBench* textual complexity indices, as presented in Table 2 and relating to: document cohesion flow (e.g., adjacent accuracy), global cohesion (e.g., paragraph-document and start-middle relatedness) and dialogism (e.g., 'voice' entropy as a measure of diversity in terms of semantic chains that contain related concepts). The effects of each index are presented in detail in the next section. The negative correlations denote a wider range of introduced topics, a more diverse vocabulary for essays with higher scores, thus a lower average global cohesion while relating each paragraph to the entire document.

We conducted a stepwise regression analysis using the first three most significant indices as the independent variables. This yielded a significant model, $F(1, 38) = 12.367$, $p < .001$, $r = .496$, $R^2 = .246$. One variable was selected as a significant and positive predictor of essay scores: document cohesion flow adjacent

accuracy using Wu-Palmer distance and maximum criterion. This variable explained 25% of the variance in the students' essay scores.

Table 2. Correlations between *ReaderBench* textual complexity indices and essay scores.

Indices	<i>r</i>	<i>p</i>
Document cohesion flow adjacent accuracy using Wu-Palmer distance and maximum criterion	.496	.001
Document cohesion flow adjacent accuracy using path distance and above plus standard deviation criterion	.451	.004
Content words (i.e., nouns, verbs, adjective and adverbs that are not considered stop-words by providing contextual information)	.448	.004
Average start-middle cohesion using path distance	-.446	.004
Average paragraph-document cohesion using path distance	-.436	.005
Average 'voice' paragraph entropy	.431	.005
Average paragraph-document cohesion using Wu-Palmer distance	-.405	.010

Afterwards, a multivariate analysis of variance (MANOVA) was conducted to examine whether the lexical and semantic properties differed between high and low performing students. For all the variables presented in Table 3, Levene's test of equality of error variances was not significant ($p > .05$); thus, the MANOVA assumption that the variances of each variable are equal across the groups was met. There was a significant difference among the two groups, Wilks' $\lambda = .295$, $p < .001$ and partial $\eta^2 = .705$. The textual complexity indices from Table 3 present the effect sizes of the variable introduced in Table 2; all indices were significantly different between the two groups of students.

The stepwise Discriminant Function Analysis (DFA) retained two variables as significant predictors (Content words, and Document cohesion flow adjacent accuracy using path distance and above plus standard deviation criterion) and removed the remaining variables (Document cohesion flow adjacent accuracy using Wu-Palmer distance and maximum criterion) as non-significant predictors. These two indices correctly allocated 33 of the 40 students, $\chi^2(df = 2, n = 40) = 19.015$, $p < .001$, for an accuracy of 82.5% (the chance level for this analysis is 50%). For the leave-one-out cross-validation (LOOCV), the discriminant analysis allocated 31 of the 40 students for an accuracy of 77.5% (see the confusion matrix reported in Table 4). The measure of agreement between the actual student performance and that assigned by the model produced a weighted Cohen's Kappa of .652, demonstrating substantial agreement.

Table 3. Tests of between-subjects effects for significantly different indices.

Dependent Variable	Mean (SD) low	Mean (SD) high	F	Sig.	Partial η^2
Document cohesion flow adjacent accuracy using Wu-Palmer distance and maximum criterion	0.98 (0.61)	1.74 (0.54)	17.33	<.001	0.313
Document cohesion flow adjacent accuracy using path distance and above mean plus standard deviation criterion	1.07 (0.63)	2.07 (0.85)	18.07	<.001	0.322
Content words	472.79 (122.10)	655.24 (139.67)	19.16	<.001	0.335
Average start-middle cohesion using path distance	0.48 (0.07)	0.41 (0.07)	9.03	.005	0.192
Average paragraph-document cohesion using path distance	0.76 (0.02)	0.74 (0.02)	6.27	.017	0.142
Average ‘voice’ paragraph entropy	1.14 (0.11)	1.26 (0.10)	15.15	<.001	0.285
Average paragraph-document cohesion using Wu-Palmer distance	0.863 (0.015)	0.855 (0.010)	4.18	.048	0.099

Table 4. Confusion matrix for DFA classifying students based on performance.

		Predicted Performance group		Total
		Low	High	
Whole set	Low	17	2	19
	High	5	16	21
Cross-validated	Low	17	2	19
	High	7	14	21

7 Discussion and Conclusions

The results of this study shed light on the essay features, in terms of complexity, that influence teachers’ scoring processes of nurse students’ case studies. First, we showed that one discriminant function, based on document cohesion flow using Wu-Palmer distance, significantly differentiated the two student groups (of low and high performance). The correlation between this variable and the teachers’ scoring is moderate (.50), and higher than the values found in another study [28] with regards to the process of scoring the overall quality of essays.

Moreover, the analysis of textual complexity indices that correlate the most with human scores brings added information on teachers’ focus. Essays with higher scores tend to be longer and contain more content words. They inherently introduce more varied concepts, additional ideas (thus, more ‘voices’ are encountered), which determines a decrease in global cohesion perceived in terms of paragraph-document cohesion, start-middle cohesion (i.e., the semantic similarity between the introduction

versus the essay body), as well as a higher entropy determined by the presence of additional semantic chains. Essays that received higher scores have a better organization in terms of paragraph structure, and a more suitable cohesion flow among adjacent paragraphs with two distance functions and both criteria; this leads to a more coherent discourse.

As a consequence, this study showed that human categorization of professional case studies can be partly predicted in analyzing document flow features. This study leads to the use of systems that would help teachers assess students' portfolios or case studies; in a parallel way, students would benefit from an automated support during writing. We strongly believe that the series of activities case studies promote can be supported by systems like *ReaderBench*: help students make connections to content, let them focused on the grade-influential textual features, collect and analyze data, write multiple drafts against standards towards the development of contextual features, prompt specific and timely feedback [34].

However, this study has some limitations. First, the number of essays is rather low, though comparable with that of other studies [35], and each essay is assessed by only one rater. Second, the way specific words are used in essays should have been subject to a more detailed analysis; for instance, the Age of Exposure model [36] would account for a more developmental view of word acquisition. Unfortunately, the model is not currently available for French language. Although semantic models like LSA and LDA were trained on specific corpora that were designed to properly conceptualize nurses' vocabulary, in further studies, we plan to adopt a more developmental view, capturing students' reflection evolution in assessing, for each student, a set of essays along the university year, independently assessed by at least two raters. We also plan to undertake a study in which students can freely assess their essays upon a series of textual complexity features, concurrently trying to improve their writing skills. Eventually, this approach might be applied to other domains and contexts, like teacher training, where reflective written accounts on activity foster professional development as well.

To our knowledge, this study is one of the few in which cohesion-centered indices proved to be predictive of human grading scores. Similarly, *ReaderBench* is one of the rare tools that provide as many and as varied textual complexity indices for languages other than English (in this study, French).

Acknowledgements. The authors wish to thank Patrice Lombardo, head of the IFSI, Centre Hospitalier Annecy-Genevois, who helped make this research possible, and Jean-Luc Rinaudo, University of Rouen, for his valuable input all along the phases of this research. This research was partially supported by the FP7 2008-212578 LTfLL project, by the 644187 EC H2020 *Realising an Applied Gaming Eco-system* (RAGE) project, as well as by University Politehnica of Bucharest through the "Excellence Research Grants" Programs UPB-GEX 12/26.09.2016.

References

1. Powell, J.H.: The reflective practitioner in nursing. *Journal of Advanced Nursing*, 14, 824–832 (1989)
2. Maynard, T., Furlong, J.: Learning to teach and models of mentoring. In: Kerry, T., Mayes, A.S. (eds.) *Issues in mentoring*, pp. 10–24. Routledge, New York (1995)
3. Simpson, E., Courtney, M.: Critical thinking in nursing education: Literature review. *International Journal of Nursing Practice*, 8(2), 89–98 (2002)
4. Jasper, M.A.: The potential of the professional portfolio for nursing. *Journal of Clinical Nursing*, 4(4), 249–255 (1995)
5. Popil, I.: Promotion of critical thinking by using case studies as teaching method. *Nurse Education Today*, 31, 204–207 (2011)
6. Schober, J., Ash, C. (eds.): *Student nurses' guide to professional practice and development*. CRC Press, Boca Raton (2005)
7. Driessen, E.: Do portfolios have a future? *Advances in Health Sciences Education*, 22(1), 221–228 (2016)
8. Eva, K.W., Bordage, G., Campbell, C., Galbraith, R., Ginsburg, S., Holmboe, E., Regehr, G.: Towards a program of assessment for health professionals: From training into practice. *Advances in Health Sciences Education*, 21(4), 897–913 (2016)
9. Brooks, V.: Marking as judgment. *Research Papers in Education*, 27(1), 63–80 (2012)
10. Green, S.M., Weaver, M., Voegeli, D., Fitzsimmons, D., Knowles, J., Harrison, M., Shephard, K.: The development and evaluation of the use of a virtual learning environment (Blackboard 5) to support the learning of pre-qualifying nursing students undertaking a human anatomy and physiology module. *Nurse Education Today*, 26(5), 388–395 (2006)
11. Dascalu, M., Dessus, P., Bianco, M., Trausan-Matu, S., Nardy, A.: Mining texts, learner productions and strategies with ReaderBench. In: Peña-Ayala, A. (ed.) *Educational Data Mining: Applications and Trends*, pp. 345–377. Springer, Cham, Switzerland (2014)
12. Dascalu, M.: Analyzing discourse and text complexity for learning and collaborating. *Studies in Computational Intelligence*, Vol. 534. Springer, Cham, Switzerland (2014)
13. Nielsen, K., Pedersen, B.D., Helms, N.H.: Reflection and learning in clinical nursing education mediated by ePortfolio. *Journal of Nursing Education and Practice*, 5(12), 63 (2015)
14. QSR International Pty Ltd.: NVivo. Vol. 2017 (2017)
15. Wild, F.: *Learning Analytics in R with SNA, LSA, and MPIA*. Springer, Berlin (2016)
16. Müller, W., Rebholz, S., Libbrecht, P.: Automatic inspection of e-portfolios for improving formative and summative assessment. In: *International Symposium on Emerging Technologies for Education (SETe 2016)*, pp. 480–489. Springer, Berlin (2016)
17. van der Schaaf, M., Donkers, J., Slof, B., Moonen-van Loon, J., van Tartwijk, J., Driessen, E., Badii, A., Serban, O., Ten Cate, O.: Improving workplace-based assessment and feedback by an E-portfolio enhanced with learning analytics. *Educational Technology Research and Development*, 65(2), 359–380 (2017)
18. Fitzgerald, J., Elmore, J., Koons, H., Hiebert, E.H., Bowen, K., Sanford-Moore, E.E., Stenner, A.J.: Important text characteristics for early-grades text complexity. *Journal of Educational Psychology*, 107(1), 4–29 (2015)
19. Frantz, R.S., Starr, L.E., Bailey, A.L.: Syntactic complexity as an aspect of text complexity. *Educational Researcher*, 44(7), 387–393 (2015)
20. Collins-Thompson, K.: Computational assessment of text readability: A survey of current and future research. *International Journal of Applied Linguistics*, 165(2), 97–135 (2014)

21. Page, E.B.: The imminence of... grading essays by computer. *Phi Delta Kappan*, 47, 238–243 (1966)
22. Larkey, L.S.: Automatic essay grading using text categorization techniques. In: *Proc. SIGIR'98*, Melbourne (1998)
23. Page, E.B., Paulus, D.H.: The analysis of essays by computer. U.S. Department of Health, Education, and Welfare, project No. 6-1318, Washington (1968)
24. Crossley, S.A., McNamara, D.S.: Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life-Long Learning*, 21(2/3), 170–191 (2011)
25. Mosallam, Y., Toma, L., Adhana, M.W., Chiru, C.-G., Rebedea, T.: Unsupervised system for automatic grading of bachelor and master thesis. *Proc. Int. Conf. eLearning and Software for Education (eLSE 2014)* (2014) 165–171
26. Attali, Y.: Validity and reliability of automated essay scoring. In: Shermis, M.D., Burstein, J. (eds.) *Handbook of automated essay evaluation: Current applications and new directions*, pp. 181–198. Routledge, New York (2013)
27. Dascalu, M., McNamara, D.S., Trausan-Matu, S., Allen, L.K.: Cohesion Network Analysis of CSCL Participation. *Behavior Research Methods*, 1–16 (2017)
28. Crossley, S.A., Dascalu, M., Trausan-Matu, S., Allen, L., McNamara, D.S.: Document Cohesion Flow: Striving towards Coherence. In: 38th Annual Meeting of the Cognitive Science Society, pp. 764–769. Cognitive Science Society, Philadelphia, PA (2016)
29. Dascalu, M., Stavarache, L.L., Trausan-Matu, S., Dessus, P., Bianco, M.: Reflecting Comprehension through French Textual Complexity Factors. In: 26th Int. Conf. on Tools with Artificial Intelligence (ICTAI 2014), pp. 615–619. IEEE, Limassol, Cyprus (2014)
30. Trausan-Matu, S.: A Polyphonic Model, Analysis Method and Computer Support Tools for the Analysis of Socially-Built Discourse. *Romanian Journal of Information Science and Technology*, 16(2-3), 144–154 (2013)
31. Bakhtin, M.M.: The dialogic imagination: Four essays. The University of Texas Press, Austin and London (1981)
32. Dascalu, M., Allen, K.A., McNamara, D.S., Trausan-Matu, S., Crossley, S.A.: Modeling comprehension processes via automated analyses of dialogism. In: *Proceedings of the 39th Annual Meeting of the Cognitive Science Society (CogSci 2017)*. Cognitive Science Society, London, UK (In press)
33. Dascalu, M., Trausan-Matu, S., Dessus, P.: Towards an integrated approach for evaluating textual complexity for learning purposes. In: Popescu, E., Klamma, R., Leung, H., Specht, M. (eds.) *Advances in web-based learning (ICWL 2012)*, Vol. LNCS 7558, pp. 268–278. Springer, New York (2012)
34. Darling-Hammond, L., Hammerness, K.: Toward a pedagogy of cases in teacher education. *Teaching Education*, 13(2), 125–135 (2002)
35. Crossley, S.A., McNamara, D.S.: Say more and be more coherent: How text elaboration and cohesion can increase writing quality. *Journal of Writing Research*, 7(3), 351–370 (2016)
36. Dascalu, M., McNamara, D.S., Crossley, S.A., Trausan-Matu, S.: Age of Exposure: A Model of Word Learning. In: 30th AAAI Conference on Artificial Intelligence, pp. 2928–2934. AAAI Press, Phoenix, AZ (2016)