Keywords of written reflection - a comparison between reflective and descriptive datasets

Thomas Daniel Ullmann

Institute of Educational Technology, The Open University,
Walton Hall, MK7 6AA Milton Keynes, United Kingdom
t.ullmann@open.ac.uk
http://www.open.ac.uk/iet/

Abstract: This study investigates reflection keywords by contrasting two datasets, one of reflective sentences and another of descriptive sentences. The log-likelihood statistic reveals several reflection keywords that are discussed in the context of a model for reflective writing. These keywords are seen as a useful building block for tools that can automatically analyse reflection in texts.

Keywords: reflection detection, thinking skills analytics, log-likelihood, keyword, key word

1 Introduction

Supporting learners with opportunities for reflective practice and fostering their reflective thinking are important educational goals. The UK Quality Assurance Agency for Higher Education (QAA), for example, recommends that all teaching and learning practices be informed by reflection [1]. The Organisation for Economic Co-operation and Development (OECD) places reflection at the 'heart of key competencies' [2], and furthermore, the Assessment and Analytical Framework for PISA sees reflection and evaluation as part of their assessment framework for reading literacy [3].

There are many ways for expressing reflective thoughts. A common representation is reflective writing (for example, see [4,5]). Reflective writing is a piece of text that contains the reflections of the writer. For example, reflective writing can be a journal, diary, blog post, or structured worksheet.

Researchers frequently analyse reflective writings to determine reflective writing quality and evaluate the success of academic writing programmes. This analysis usually follows a content analysis approach (for example, see [6,7]). Researchers use content analysis to systematically detect all textual evidence that belongs to model categories of reflective writing. However, the content analysis of reflective writing is a time-consuming process. Automated reflection analytics techniques have the potential of reducing the amount of time necessary to analyse reflective writings.

This paper contributes to the research of automated detection of reflection in texts (for example, see [8,9,10]). The automated detection of reflection is linked to one of the grand challenges of technology enhanced learning, which is 'e-assessment and automated feedback' [11,12]. In order to create automated systems to assist reflective writing assessment, techniques have to be developed first that can automatically detect reflection in texts. Once a system can detect reflection, this information can be used to automatically assist the assessment of reflection. Therefore, reflection detection is a base technology with the potential for several applications, for example, e-assessment and automated feedback.

For the automated detection of reflection, it is important to identify patterns or regularities found in reflective writings. These regularities bear the potential of being formalised in computer programs, which then can automatically detect these patterns of reflection in novel texts. This paper shows a method that allows the identification of reflection keywords based on the comparison of datasets with the log-likelihood statistic. It then discusses the keywords derived in the context of a model of written reflection.

2 Automated detection of reflection in texts

Research in the area of automated detection of reflection aims at the development of those techniques and technologies that can automatically identify the characteristics of reflection in texts. Three techniques have been identified that have been used to automatically analyse texts in respect to reflection [13]. They are dictionary-based, rule-based, and machine learning-based approaches. The dictionary-based approach makes use of lists/dictionaries of words. The words contained in a dictionary represent aspects of reflection. These dictionaries can be used to analyse texts with regard to the frequency of word occurrences in texts, or to visually highlight detected text passages (for example, see [8]). The rule-based approach makes use of a set of rules. Each rule captures an aspect of reflection. These rules, along with a rule-based system, allow drawing inferences from texts, and can be used to analyse reflective writings (for example, see [9]). The first two approaches make use of expert knowledge in order to construct the dictionaries or rules. The third approach makes use of machine learning. Machine learning algorithms learn regularities or patterns from many examples that represent facets of reflection [13]. The generated machine learning models classify unseen text into categories of reflection.

Although machine learning-based approaches can automatically build models to detect reflection, the first two approaches rely on explicit knowledge about either words or rules that represent aspects of reflection.

The literature of research that applied content analysis to investigate reflective writings indicated that reflective writings exhibit such textual patterns. Hatton and Smith [14] touched on language patterns that aided the coding of dialogic reflection. Fund et al. [15] noted the coordination between idea units as important for identifying reflection types. Poom-Valickis and Mathews [16] mentioned lists of keywords in order to code text units. Hawkes and Romiszowski [17] and Hawkes [18, 19] suggested an association between discourse markers and reflection. In that research, the guiding framework of the analysis was the model of reflection selected by these authors.

Another area of the research that investigates patterns of reflection makes use of systemic-functional linguistics [20–28]. This research has in common that it investigates text based on a linguistic framework in order to derive a link between the linguistic framework and reflection expressed in texts. The guiding framework for this type of research is the linguistic framework. Categories of the linguistic framework are then mapped to categories of reflective writing models in order to explore the relationship between linguistic resources and reflective writing.

The approach taken here makes use of the corpus linguistic keyword method using log-likelihood statistic as described by Rayson [29] to find reflection keywords. This method is based on the frequency analysis of two corpora/datasets in order to investigate words that occur significantly more frequently in one or another dataset. Here, the datasets consist of reflective and descriptive sentences. The keyword method is used to reveal words that are significantly more or less often used in the dataset of reflective sentences compared with the dataset of descriptive sentences.

This approach is different from the above outlined content analysis and systemic-functional approaches because it places data first, and not theory. The aforementioned approaches use theory to interpret data, whereas the approach taken here derives a set of keywords using a statistical method. These empirically derived keywords can then be interpreted in the context of theory. This is at first a data-driven approach that may inform theory [29].

The term 'keyword' has several notions, and within this paper it describes those words that occur significantly more frequently within one dataset than another [30]. Several statistical tests can be used to calculate the 'keyness' of words [31]. Here, the chosen test is the log-likelihood ratio test as described by Rayson [29] (see also [32] for a similar implementation of the log-likelihood test).

3 Models to analyse reflective writings

The datasets used to derive keywords were created according to a frequently found distinction made in research that analyses reflective writings: A text can be either descriptive/non-reflective or reflective. The lowest level is often described as descriptive, and it contains no reflection; on the other side of the scale are reflective texts, which can be further distinguished according to several levels of reflection (for example, see [33,34–38]). However, the common denominator of these models is the basic distinction between descriptive and reflective texts.

In addition to levels of reflection, research into the analysis of reflective writings proposed several models with various model categories that describe constituents of reflective writing. These model constituents describe the breadth, and not the depth, of reflection as the level models. Manual content analysis of reflective writings uses the categories that describe breadth facets of reflective writing, as well as levels of reflection as their coding category schema. Although the model categories vary from research to research, they do share some commonalities. The model used to aid in the interpretation of the keywords is based on the model for reflection detection described by

Ullmann [13]. An older version of this model can be found in the study by Ullmann et al. [9]. The breadth model categories are:

Experience: A reflective writing is often about experience or a personal matter. The description of what occurred and the capturing of the important characteristic of the situation provide the background and focal point for reflective writing. The description of the experience captures important parts of the experience, and provides the context and/or the reason for the writing. This category can be frequently found in models that analyse reflection (for example, see [6,7], [39]).

Personal: A reflective writing is often of a personal nature. This means that it is often about beliefs, personal assumptions, or knowledge about oneself. The text is written with a personal voice and shows the development of a perspective on the experience at hand. Several models describe this category (for example, see [20], [39,40]).

Feelings: Feelings can be the starting point of a reflection. Feelings often associated with reflection are the feeling of being concerned, having doubts, a feeling of uncertainty, frustration, but also feelings such as surprise or excitement. Whereas feelings can be the starting point of a reflection, they can also be the subject matter of the reflection, for example, reflections on the influence of feelings on thinking and action. Several models that analyse reflective writings contain references to this category (for example, see [7], [39,40]).

Critical stance: Expressing an alert or critical mindset is an important part of reflective writing. Having a critical stance involves being aware of problems and being able to identify or diagnose such problems. Being critical is about questioning assumptions and opinions, analysing and evaluating problems, judging situations, testing the validity of assumptions, drawing conclusions, and making decisions. This category is mentioned in many models (for example, see [6], [20], [39]).

Perspective: Although reflective writings are often written from the first person perspective, considering other perspectives is an important facet of reflective writing. Examples are the perspective of someone else, a theory; the social, historical, ethical, moral, or political context. Several content analysis models contain this category (for example, see [20], [39], [41]).

Outcome: There can be several outcomes from reflective writing. An outcome from reflective writing can be a description of lessons learned, better understanding of the situation or context, new insights, change of perspective or behaviour, and the awareness of one's way of thinking. An outcome can be also an intention to do something or any planning for the future. The category outcome is also frequently mentioned in content analysis models used to analyse reflective writings (for example, see [39], [41,42]).

These six categories, which stem directly from the research on manual content analysis of reflective writing, form the guiding framework for the interpretation of the results of keyword analysis.

4 Dataset generation process and datasets

The datasets of reflective and descriptive sentences were obtained from research that investigated the automated detection of reflection using machine learning (details are found in [13]). These two datasets are mostly based on a sample of the British Academic Writing English Corpus (BAWE) [43,44]. The sampled texts are mostly from the disciplines of health, business, and engineering.

A sentence splitter divided each sample text from the BAWE text collection into sentences. Seven to ten raters ranked each sentence on a six-point Likert scale as to whether the sentence is descriptive or reflective. A crowdsourcing solution[1] was used to distribute the sentences to the raters. An even-numbered scale was used so that the raters had to decide whether the sentence is reflective or descriptive, and to avoid misusing the neutral point of an odd-numbered scale as the 'don't know' category. The ratings on the six-point Likert scale were then dichotomised into the class reflective and descriptive.

The aim was then to generate two datasets of approximate equal size to aid the comparison of both datasets. A sentence was only included into the dataset if it received a 4/5 majority of ratings for either belonging to the reflective or descriptive classes. The decision on which aggregation strategy to choose was based on the 4/5 majority because it represents a more strict quality standard compared with the more lenient simple majority vote. This ensures that only those sentences that received substantial support as belonging to one of the two categories were included in this study. For example, a sentence that received ten ratings was included if eight or more of the ten ratings ranked the sentence as reflective (or descriptive). Reliability estimates of the ratings aggregated with majority and 4/5 majority vote were reported by Ullmann [13], who found as substantial for the majority vote a Cohen's kappa of 0.62, and almost perfect for the 4/5 majority vote a Cohen's kappa of 0.92, according to the benchmark of Landis and Koch [45]. From this annotated dataset of highly agreed sentences, a random sample of 500 reflective sentences and 500 descriptive sentences was drawn.

All sentences from the two datasets were pre-processed with the same data generation process. This involved the removal of any punctuations, numbers, and superfluous whitespaces, sentence tokenisation to words, and word conversion to lower case. The R environment for statistical computing and graphics [46,47] was used to develop the scripts for data processing and calculation of the log-likelihood ratio.

The dataset of reflective sentences contains a total of 12,697 words (2,200 unique words). The average sentence length is 25.39 words. The dataset of descriptive/non-reflective sentences contains a total of 10,284 words (2,800 unique words). The average sentence length is 20.57 words.

5 Results

The frequency of each word of each dataset was counted and compared. Word comparison is based on the log-likelihood of the two terms [29]. The log-likelihood considers the frequency of the two terms compared with the size of the entire datasets. Table 1 lists the log-likelihood of all words with a log-likelihood higher than 10.83, which represents a p-value < 0.001[2], and an effect size calculated with the Bayes Factor[2] of > 2

[1] CrowdFlower (http://www.crowdflower.com/)

[2] http://ucrel.lancs.ac.uk/llwizard.html

[48]. Word pairs below these thresholds are not listed. Table 1 is sorted by the log-likelihood with the highest log-likelihood at the top, and the lowest at the bottom. Furthermore, the table indicates for each term, the frequency of occurrence in the datasets of reflective and descriptive sentences. The column 'Use' indicates with a + and - sign whether the term is overused ('+') in the reflective dataset, which means that it has a higher relative frequency in the reflective dataset, or underused ('-'), which means that it is more frequently used in the dataset of descriptive sentences.

For example, the word 'i' is frequently present in the dataset of reflective sentences. It occurs 700 times in the reflective dataset and 105 times in the descriptive dataset. It has the highest log-likelihood ratio of 376.07, which means that the word 'i' occurs unusually often (the column use has a '+'-sign) in the dataset of reflective sentences compared to the dataset of descriptive sentences. The word 'he' is underused in the reflective dataset (see the '-'-sign), which means it appears unusually often in the descriptive dataset according to the used log-likelihood test.

Table 1. Log-likelihood of the datasets words.

Word	Reflective dataset	Descriptive dataset	Log-likelihood	Use
i	700	105	376.07	+
have	191	35	88.09	+
me	107	8	81.75	+
my	201	56	59.11	+
feel	68	4	56.23	+
felt	61	4	48.76	+
not	117	29	39.91	+
that	285	130	31.25	+
more	78	18	28.85	+
better	30	1	28.37	+
is	72	123	26.41	-
this	157	63	24.11	+
believe	26	1	23.91	+
now	29	2	22.80	+
he	5	26	20.35	-
by	36	71	20.23	-
future	17	0	20.17	+
of	285	329	19.24	-
was	181	84	18.84	+
situation	31	4	18.34	+
think	31	4	18.34	+
are	30	61	18.32	-
if	47	11	17.12	+
would	83	29	17.01	+
and	369	402	16.92	-
but	51	13	16.82	+

Word	Reflective dataset	Descriptive dataset	Log-likelihood	Use
never	14	0	16.61	+
bit	13	0	15.43	+
system	1	13	14.89	-
their	11	31	14.60	-
could	56	17	14.55	+
ae[3]	0	9	14.47	-
knowledge	5	21	14.25	-
aware	12	0	14.24	+
hindsight	12	0	14.24	+
learnt	17	1	14.06	+
although	20	2	13.54	+
probably	11	0	13.05	+
it	160	80	12.98	+
place	2	14	12.83	-
myself	22	3	12.58	+

6 Discussion

Table 1 lists the dataset words with the highest log-likelihood. Their relative frequency differs between datasets, which makes them distinctive. They are words for which it is unlikely that the null hypothesis, where their relative frequencies are the same, is true. These are the keywords defined by the statistical procedure. Their p-value and effect size act as inclusion criteria. An additional criterion could have been used to exclude words that occur relatively infrequently [30]. For example, the word 'ae' for 'A&E' (see footnote 3) occurs nine times in both datasets, which makes it the word with the least occurrences in Table 1.

In the following subsections, some of the keywords are discussed, and a link between the keywords and their belongings to one of the categories of the model of reflective writing is established. Several of the keywords are illustrated with sentences obtained from the datasets. The keywords within each sample sentence are highlighted in bold.

6.1 Experience

The description of an 'Experience' often entails the description of a situation that occurred in the past. One of the keywords directly addresses a 'situation'. An example of a sentence with this key word is, 'On the whole I felt I and the other members of staff did all they could to manage a difficult **situation** and gave Joseph more than ample

[3] The word 'ae' represents 'A&E', which refers to the Accident and Emergency service. As all punctuations have been removed during the data generation process, also the '&' of A&E was removed.

opportunity to cooperate however, in **hindsight** I feel some aspects could have been handled differently'. Another example is the sentence, 'My reaction to her and the **situation** surprised me as I became quiet agitated and in **hindsight** was probably just as unaccommodating as she was'. The two sentences also contain the keyword 'hindsight', which is used to express a retrospective understanding of a situation.

6.2 Personal

It is notable that first person singular pronouns, such as 'I', 'me', 'my', and 'myself', are keywords overused in the reflective dataset. This may indicate that reflective writings are frequently written from the first person perspective, which is also in line with the category 'Personal' of the reflection model as outlined above. The third person singular pronoun 'he' is more associated with a descriptive text. The category 'Personal' is also about own beliefs. The keyword 'believes' may be indicative for expressing beliefs. The following sentence is an example of this: 'Reflecting about conflicts I had in the past, I **believe** that I could have handled some of them better'. Another example of this is the sentence, 'When the time came to allocate work for the plan I **believe** the team dynamics were developed enough to assess accurately everybody's strengths and apportioned work accordingly'. The modal verb 'would' can also refer to beliefs. An example is the sentence, 'I felt that as team leader I **would** have control in the group and I **would** have more say in the way our team was run, little did I know then'. Another example is the sentence, 'Knowing about the tradition, I **would** definitely have acted differently, hopefully being in the position to speak at least a bit of the language'.

6.3 Feelings

The words 'feel' and 'felt' are at top of the list, and they occur relatively more often in the dataset of reflective sentences. This is in line the category 'Feelings' of the reflection model. Expressing feelings is often mentioned as part of reflective writing.

6.4 Critical stance

Several words can be associated with the category 'Critical stance', for example, the keywords: 'more', 'better', 'if', 'but', 'never', 'could', and 'although'.

The word 'more' could relate to the critical thought of a writer that something is lacking and that more of something would be better, or it could relate to the realisation that there is now more of something that was previously not there. For example, the sentence, 'I should be **more** aware about the power issues and how they silence patients', expresses the first sense, which is the realisation that something is still lacking. The sentence, 'I noticed **more** discussions taking place after the first couple of sessions, and I felt our group was **more** established as we began to get to grips with what the vignette would entail and felt comfortable with each other', refers to the second meaning—the realisation of a change.

The word 'better' could refer to the critical awareness of the writer that something is now better, as expressed in the sentence, 'I hadn't really thought of it like this before but by empathising with Mary's situation I **better** appreciate the importance of the patient's perspective'. It could also express that something should have been better, as in the sentence, 'I might have **better** explored Jim's internal thoughts and his wondering about what he finds in the cave'.

The conjunction 'if' could express a premise followed by a conclusion as part of reasoning about something. For example, the sentence, 'It would have been helpful **if** I had shared my concerns about the group with the LSA to start with'. If he had shared his concerns (premise), then it would have been helpful (conclusion).

The conjunction 'but' could express a contrasting thought. An example is the sentence, 'Looking at it from Marissa's point of view, she may have known that I was on the wrong track, **but** she probably would not have been able to do anything about it because I am a doctor'. Another example is the sentence, 'I did not do all that well in the exam so maybe I need to prepare differently - **but** I really don't know how to do it'.

A writer could flag with the adverb 'never' a realisation that something was never experienced or that something never happened before. The following sentence is indicative for this: 'In the topic I found it most interesting about the lack of invariance problem as I have **never** realised the fact before although the point is reasonably understandable'. Another example is the following sentence: 'I had **never** experienced those same feelings of disconnect in real time though, **never** felt as though the person talking was somehow not me, until last Tuesday'.

The verb 'could' might indicate the awareness of a possibility or alternative. The sentence, 'On the whole our group worked well but **could** have been improved by more openness and discussion about issues affecting the group, such as social loafing', shows that the writer describes a realisation that there is a real possibility for improvement. Another example is the sentence, 'Reflecting about conflicts I had in the past, I believe that I **could** have handled some of them better'.

The word 'although' could be used in a contrasting way. An example is the following sentence: '**Although** throughout my training to date, I have dutifully reflected on various clinical situations and considered learning objectives within the practice portfolio; I can not say that I had actually fully taken on the implications of what it is to be a truly reflective practitioner'. The writer describes the contrast between the perception of reflection in previous trainings and the current perception. Another example is the sentence, 'Applying the learning cycle proved to be a useful tool, **although** I was very sceptic at the beginning'. This sentence describes a contrast in perception. The sceptical few dissolved over time.

6.5 Perspective

The keyword list from Table 1 does not contain a keyword that allows us to establish a strong link between a keyword and the category 'Perspective'.

6.6 Outcome

The 'Outcome' category of the model contains a retrospective dimension that entails, for example, the description of lessons learned, but also a prospective dimension directed to statements about what to do in future. The keyword list contains the word 'learnt'. This word could express that something was learned. An example of this outcome facet is the sentence, 'I have **learnt** that when I am requesting something different, I must explain my needs fully, and communicate the message more effectively'. Another example is the sentence, 'I have also **learnt** from participating in this group that I was, in this case, one of the more dominant group members, and felt confident in expressing my views and ideas'. Another keyword is the noun 'future'. This keyword could be used to express future intentions. An example is the sentence, 'Although this situation didn't have a satisfactory outcome I hope to have learned from the experience and aim to use my new insight to develop my **future** practice'. Another example is the sentence, 'This makes you realise that you could come across this within the profession and luckily from this activity I am now aware of this and I can now make the most of any opportunities that a rise to enable me to take this into account and maybe build up my confidence so in the **future** I can maybe go onto stand up for what I believe in and also get my opinions noticed if I feel this necessary'.

6.7 Summary

Overall, this discussion showed that for several keywords, the log-likelihood statistic can derive words that are in line with the categories of the chosen model of reflection.

The keywords listed in Table 1 can be seen as good candidate words for the construction of dictionaries. With the shown approach, we can form a set of words, such as a dictionary, that can be used to automatically summarise texts with regard to the frequency of occurrence for each category. However, Table 1 also shows that words, although they are found frequently in one dataset, cannot be used to distinguish completely between reflective and descriptive use. For example, the keyword 'i' is found 700 times in the reflective dataset, but 'i' is also used 105 times in the dataset of descriptive sentences. A definite classification of text passages based on single words is also not the aim of dictionary-based approaches. One of the use cases there is that the dictionaries are used to predict important outcomes. This is a quantitative indicator that can be used to corroborate the findings of a study (for example, see [49]).

7 Conclusion and outlook

This paper demonstrated the application of the keyword method on a dataset of reflective and descriptive sentences. The log-likelihood statistics was used to determine words with high 'keyness' in either the dataset of reflective sentences or that of descriptive sentences. The words derived with the described method represent words that occur with unusual relative frequency in the datasets. In the discussion of the results, several of these keywords were assigned to categories of a model of reflective writing. This step illustrated that the investigated keywords can be associated with the reflection model categories. This supports the applicability of the keyword method to derive words important for reflection.

An extension of the shown approach is to investigate the 'keyness' of the categories of the reflection model. For example, a comparison of a dataset that describes the outcomes of reflective writings with a reference dataset would allow us to derive keywords of reflection outcomes by adding an in-depth study of keywords for this category.

Furthermore, reflection dictionaries can be combined with rule-based systems [9]. Rule-based systems provide more control in modelling relationships between dictionary words, which could add to the precision of the automated method.

As outlined, the automated detection of reflection in writings relies on patterns of reflection, because these patterns can be codified and used for the automated analysis of writings. The study showed that reflective writings contain such patterns at the word level, because there are words that occur significantly more often in the dataset of reflective sentences than in the dataset of descriptive sentences.

8 References

1. QAA: UK quality code for higher education. Part B: Assuring and enhancing academic quality. Chapter B3: Learning and teaching. (2012).
2. Rychen, D.S., Salganik, L.H.: The Definition and Selection of Key Competencies: Executive Summary. OECD (2005).
3. OECD: PISA 2012 Assessment and Analytical Framework. OECD Publishing (2013).
4. Dyment, J.E., O'Connell, T.S.: The Quality of Reflection in Student Journals: A Review of Limiting and Enabling Factors. Innovative Higher Education. 35, 233–244 (2010).
5. Moon, J.A.: Learning Journals: A Handbook for Reflective Practice and Professional Development. Routledge (2006).
6. Poldner, E., Van der Schaaf, M., Simons, P.R.-J., Van Tartwijk, J., Wijngaards, G.: Assessing student teachers' reflective writing through quantitative content analysis. European Journal of Teacher Education. 37, 348–373 (2014).
7. Prilla, M., Renner, B.: Supporting Collaborative Reflection at Work: A Comparative Case Analysis. In: Proceedings of the 18th International Conference on Supporting Group Work. pp. 182–193. ACM Press, New York, NY, USA (2014).
8. Ullmann, T.D.: An Architecture for the Automated Detection of Textual Indicators of Reflection. In: Reinhardt, W., Ullmann, T.D., Scott, P., Pammer, V., Conlan, O., and Berlanga, A. (eds.) Proceedings of the 1st European Workshop on Awareness and Reflection in Learning Networks. pp. 138–151. CEUR-WS.org, Palermo, Italy (2011).
9. Ullmann, T.D., Wild, F., Scott, P.: Comparing Automatically Detected Reflective Texts with Human Judgements. In: Moore, A., Pammer, V., Pannese, L., Prilla, M., Rajagopal, K., Reinhardt, W., Ullmann, T.D., and Voigt, C. (eds.) 2nd Workshop on Awareness and Reflection in Technology-Enhanced Learning. CEUR-WS.org, Saarbruecken, Germany (2012).

10. Ullmann, T.D., Wild, F., Scott, P.: Reflection - quantifying a rare good. In: Kravcik, M., Krogstie, B.R., Moore, A., Pammer, V., Pannese, L., Prilla, M., Reinhardt, W., and Ullmann, T.D. (eds.) Proceedings of the 3rd Workshop on Awareness and Reflection in Technology-Enhanced Learning. pp. 29–40. CEUR-WS.org, Paphos, Cyprus (2013).
11. Sutherland, R., Eagle, S., Joubert, M.: A vision and strategy for Technology Enhanced Learning: Report from the STELLAR Network of Excellence. (2012).
12. Fischer, F., Wild, F., Sutherland, R., Zirn, L. eds: Grand Challenges in Technology Enhanced Learning - Outcomes of the 3rd Alpine Rendez-Vous. Springer International Publishing (2014).
13. Ullmann, T.D.: Automated detection of reflection in texts - A machine learning based approach. PhD thesis The Open University (submitted).
14. Hatton, N., Smith, D.: Reflection in teacher education: Towards definition and implementation. Teaching and Teacher Education. 11, 33–49 (1995).
15. Fund, Z., Court, D., Kramarski, B.: Construction and Application of an Evaluative Tool to Assess Reflection in Teacher-Training Courses. Assessment & Evaluation in Higher Education. 27, 485–499 (2002).
16. Poom-Valickis, K., Mathews, S.: Reflecting others and own practice: an analysis of novice teachers’ reflection skills. Reflective Practice. 14, 420–434 (2013).
17. Hawkes, M., Romiszowski, A.: Examining the reflective outcomes of asynchronous computer-mediated communication on inservice teacher development. Journal of Technology and Teacher Education. 9, 285–308 (2001).
18. Hawkes, M.: An analysis of critically reflective teacher dialogue in asynchronous computer-mediated communication. In: IEEE International Conference on Advanced Learning Technologies, 2001. Proceedings. pp. 247 –250 (2001).
19. Hawkes, M.: Linguistic Discourse Variables as Indicators of Reflective Online Interaction. American Journal of Distance Education. 20, 231–244 (2006).
20. Birney, R.: Reflective Writing: Quantitative Assessment and Identification of Linguistic Features, http://repository.wit.ie/2658/, (2012).
21. Forbes, A.: Evidence of learning in reflective practice: a case study of computer-assisted analysis of students’ reflective blogs. New Zealand Association for Cooperative Education 2011 Conference Proceedings. 11–14 (2011).
22. Luk, J.: Assessing teaching practicum reflections: Distinguishing discourse features of the “high” and “low” grade reports. System. 36, 624–641 (2008).
23. Reidsema, C., Mort, P.: Assessing reflective writing: Analysis of reflective writing in an engineering design course. Journal of Academic Language and Learning. 3, A117–A129 (2009).
24. Ryan, M.: Improving reflective writing in higher education: a social semiotic perspective. Teaching in Higher Education. 16, 99–111 (2011).
25. Ryan, M.: Conceptualising and teaching discursive and performative reflection in higher education. Studies in Continuing Education. 34, 207–223 (2012).
26. Ryan, M.: Reflexive writers: Re-thinking writing development and assessment in schools. Assessing Writing. 22, 60–74 (2014).

27. Shaheed, N., Dong, A.: Reflection and analysis in design student blogs. DS 36: Proceedings DESIGN 2006, the 9th International Design Conference, Dubrovnik, Croatia. (2006).
28. Wharton, S.: Presenting a united front: assessed reflective writing on a group experience. Reflective Practice. 13, 489–501 (2012).
29. Rayson, P.: From key words to key semantic domains. International Journal of Corpus Linguistics. 13, 519–549 (2008).
30. Scott, M.: PC analysis of key words — And key key words. System. 25, 233–245 (1997).
31. Lijffijt, J., Nevalainen, T., Säily, T., Papapetrou, P., Puolamäki, K., Mannila, H.: Significance testing of word frequencies in corpora. Digital Scholarship in the Humanities. fqu064 (2014).
32. Dunning, T.: Accurate methods for the statistics of surprise and coincidence. Computational linguistics. 19, 61–74 (1993).
33. Ballard, K.K.: Using Van Manen's model to assess levels of reflectivity among preservice physical education teachers, https://repository.tamu.edu/handle/1969.1/4373, (2006).
34. Ip, W.Y., Lui, M.H., Chien, W.T., Lee, I.F., Lam, L.W., Lee, D.: Promoting self-reflection in clinical practice among Chinese nursing undergraduates in Hong Kong. Contemporary Nurse. 41, 253–262 (2012).
35. Kember, D., Jones, A., Loke, A., McKay, J., Sinclair, K., Tse, H., Webb, C., Wong, F., Wong, M., Yeung, E.: Determining the level of reflective thinking from students' written journals using a coding scheme based on the work of Mezirow. Int. J. of Lifelong Educ. 18, 18–30 (1999).
36. Lai, G., Calandra, B.: Examining the effects of computer-based scaffolds on novice teachers' reflective journal writing. ETR&D-Educ. Tech. Res. Dev. 58, 421–437 (2010).
37. McCollum, S.: Insights Into the Process of Guiding reflection During an Early Field experience of Preservice Teachers. (1997).
38. Sparks-Langer, G.M., Simmons, J.M., Pasch, M., Colton, A., Starko, A.: Reflective Pedagogical Thinking: How Can We Promote It and Measure It? Journal of Teacher Education. 41, 23 –32 (1990).
39. Wald, H.S., Borkan, J.M., Taylor, J.S., Anthony, D., Reis, S.P.: Fostering and Evaluating Reflective Capacity in Medical Education: Developing the REFLECT Rubric for Assessing Reflective Writing. Academic Medicine. 87, 41–50 (2012).
40. Wallman, A., Lindblad, A.K., Hall, S., Lundmark, A., Ring, L.: A Categorization Scheme for Assessing Pharmacy Students' Levels of Reflection During Internships. Am J Pharm Educ. 72, (2008).
41. Plack, M.M., Driscoll, M., Marquez, M., Cuppernull, L., Maring, J., Greenberg, L.: Assessing Reflective Writing on a Pediatric Clerkship by Using a Modified Bloom's Taxonomy. Ambulatory Pediatrics. 7, 285–291 (2007).
42. Plack, M.M., Driscoll, M., Blissett, S., McKenna, R., Plack, T.P.: A method for assessing reflective journal writing. Journal of allied health. 34, 199–208 (2005).
43. Gardner, S., Nesi, H.: A Classification of Genre Families in University Student Writing. Applied Linguistics. 34, 25–52 (2013).

44. Nesi, H., Gardner, S.: Genres across the disciplines: Student writing in higher education. Cambridge University Press (2012).
45. Landis, J.R., Koch, G.G.: The Measurement of Observer Agreement for Categorical Data. Biometrics. 33, 159–174 (1977).
46. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2014).
47. Feinerer, I., Hornik, K.: tm: Text Mining Package. (2014).
48. Wilson, A.: Embracing Bayes factors for key item analysis in corpus linguistics. In: Bieswanger, M. and Koll-Stobbe, A. (eds.) New approaches to the study of linguistic variability. pp. 3–11. Peter Lang, Frankfurt (2013).
49. Chung, C.K., Pennebaker, J.W.: Linguistic Inquiry and Word Count (LIWC): Pronounced "Luke" and Other Useful Facts. In: McCarthy, P.M. and Boonthum, C. (eds.) Applied Natural Language Processing and Content Analysis: Advances in Identification, Investigation and Resolution. pp. 206–229. Information Science Reference (an imprint of IGI Global), Hershey, United States of America (2012).