Sentiment Analysis of the ICT4Rural Education Teacher Professional Development Course.

Coral FEATHERSTONE[1], Adele BOTHA[2]
[1]CSIR Meraka, PO Box 395, Pretoria, 0001, South Africa
Tel: +27 12 8414925 , , Email: cfeatherstone@csir.co.za
[2]CSIR Meraka, PO Box 395, Pretoria, 0001, South Africa
[2]UNISA, School of Computing College of Science, Engineering
& Technology University of South Africa
Tel: +27 12 841 3265, Email: abotha@csir.co.za

Abstract: This paper reports on an attempt to gauge the experiences of teachers participating in a teacher professional development course through sentiment analysis. The course formed part of a large scale implementation aimed at changing teachers' classroom practice to reflect a 21st century teaching and learning engagement. The course was purposely designed for a gameful educational user experience to address the design challenge of motivating the participants through a yearlong intervention. Although the actual design and implementation of the course is beyond the scope of this paper, we present results that indicate that the course was experienced mostly relevant and positive. This novel way of gauging the sentiment and hedonistic attributes of a course, holds much promise for similar applications.

Keywords: Sentiment Analysis; Mobile Learning; ICT4D

1 Introduction

The ICT for Rural Education Development (ICT4RED) project is part of a research programme (TECH4RED) initiated by the South African Department of Science and Technology (DST) in collaboration with the South African Department of Basic Education (DBE), the Eastern Cape Department of Education (ECDoE) and the South African Department of Rural Development and Land Reform (DRDLR). TECH4RED aims to contribute to the improvement of rural education via technology-led innovation. The learning from this programme will enable evidence-based policy development within government.

The research programme is currently in its third year, and has provided for a range of technology interventions including, amongst others, in ICTs, water and sanitation, health, nutrition and energy. This list is not exhaustive of the possibilities but is a considered selection of key interventions. [1].

Within TECH4RED the ICT4RED aims to investigate how the application and deployment of new and existing technologies (which include e-textbooks and other electronic resources) at schools in the Cofimvaba district of the Eastern Cape Province can assist to develop

a framework which can be replicated and scaled to other provinces and across the rural education system.

Each participating teacher received a tablet[1] and the learners, depending on their grade, either receive a tablet for use, or was part of a group that would share a tablet in class [2].

The *ICT4RED Teacher professional development* emerged as the primary driver within the initiative and aimed to guide the development of relevant teacher knowledge and proficiency to enable classroom practice to portray a 21st century teaching and learning engagement as outlined by Voogt and Odenthal [3]. The *ICT4RED TPD curriculum* is an instantiation of the Mobile learning curriculum [4] and is currently in its third iteration.

The curriculum was purposely designed to give a *Gameful Educational User Experience* [5]. Educational Gamification, viewed as the design strategy of using game design elements in educational contexts to support teaching and learning goals, is about learning and learning gains and ought to be grounded in best practice pedagogical principles. To this end, the *Gameful Educational User Experience* in this initiative is viewed as a purposely designed for gameful educational experience underpinned by the attributes and affordances of the technology and the pedagogy interaction, to realize and support the educational aim of the initiative. That aim being to meaningfully integrate technology as part of a change in classroom practice to reflect an emergent pedagogy. The design challenge that was addressed was the motivation of users [6].

This paper reports on an attempt to gauge the experience of the second iteration through sentiment analysis.

The rest of the paper is structured as follows. The following gives very high level overview of the ICT4RED course [2], followed by an overview of sentiment analysis and the basic techniques of opinion assessment. The method used is discussed and some limitations are outlined. This is followed by a presentation and discussion of the results where after a conclusion is presented.

2   ICT4RED Teacher Professional Development (TPD)

The course takes about a year to implement the 10 modules depending on the time constraints. The implementation takes place as illustrated in the diagram below:

Figure 1: TPD implementation flow

(1) The learning strategy, skills and other competencies are simulated during the TPD session. This provides an opportunity to experience an appropriate teaching strategy, learn about a topic and gain technology skills.

---

[1] The teachers would be required to show competence before ownership of the tablet was transferred.
[2] The actual design and implementation of the course is beyond the scope of this paper.

(2) Subsequent to the TPD session, the participating teachers have about 3 weeks to apply the strategy using technology in their own class. They need to record some evidence as outlined in the badge criteria.

(3) A badge evaluator evaluates the evidence provided and either award the badge or gives meaningful input on possible improvements. In this case the teacher can resubmit at any given time.

(4) If there are still modules left, another TPD session will be done and the process repeats.

(5) If all the modules have been presented the participating teacher has the opportunity to graduate and earn their device, should they have achieved the minimum criteria.

The TPD course narrative is one of changing classroom practice to portray more of an "emerging pedagogy for the information age" and less of a "traditional pedagogy" [7, 8].

Table 1: Towards Emerging pedagogy for the information age [7, 8]

Aspect	Less of "traditional pedagogy"	More of "emerging pedagogy for the information age"
Active	Activities prescribed by the teachers Whole-class instruction Little variation in activities Pace determined by program	Activities determined through negotiation Small groups Varied activities Pace determined by learners
Collaboration	Individual Homogeneous Groups Everyone for him/herself	Working in teams Heterogeneous groups Supporting each other
Creative	Reproductive learning Apply known solutions to problems	Productive Learning Find new solutions to problems
Integrative	No link between theory and practice Separate subjects Discipline-based Individual teaching	Integrating theory and practice Relations between subjects Thematic Teams of teachers
Evaluative	Teacher- directed Summative	Student directed Formative

The transformation was operationalized through a learning path that consists of 13 interim goals articulated as badges. The learning path changes focus from committing to the programme, to using the technology for personal use, evolving into using the technology for teaching and learning to ultimately using the technology to collaborate, communicate and share. This progression is detailed in the graphical image below:

Figure 2: Learning path Narrative

This is represented as follows in a sequential manner. Each of the learning goals articulated as badges in the course is represented by an image and is physically awarded in the form of a sticker to add to the learning path. Originally it was planned to award Mozilla open badges but the disconnected environment and the teachers not having a digital presence or email made the electronic badges irrelevant until much later in the intervention.
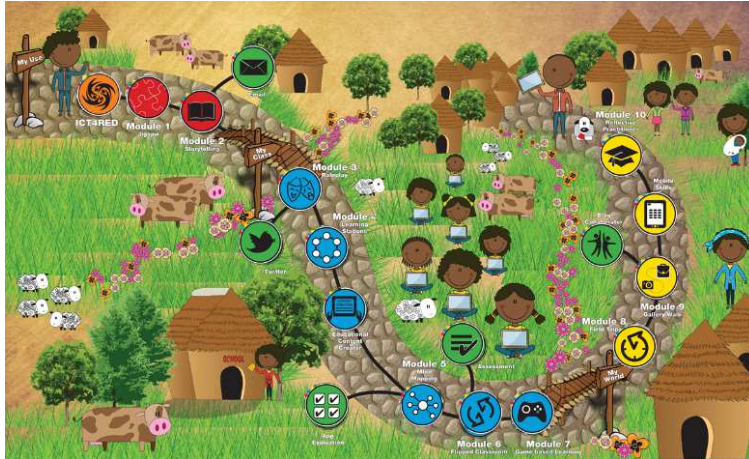

Figure 3: Sequential Learning Path represented by a road to be followed and badges to be earned.

There were 10 modules.

Module	Compulsory Badges
Module 1	Jigsaw strategy and getting to know your tablet
Module 2	Mind mapping strategy and implementing an acceptable user's policy in your class.
Module 3	Storytelling as a strategy
Module 4	Roleplay as a teaching strategy
Module 5	Learning stations as a teaching strategy and creating digital content.
Module 6	Game-Based Learning strategies and creating educational video's
Module 7	Scavenger hunt as a teaching strategy
Module 8	Gallery walk as a teaching strategy.
Module 9	Flipped classroom as a teaching strategy
Module 10	Reflecting on the journey.

Having briefly outlined the ICT4RED TPD course[3], the following section overviews Sentiment Analysis.

3   Sentiment analysis overview

Sentiment analysis, which is also known as opinion mining, is the use of Natural Language Processing (NLP), statistics or machine learning methods to identify or extract user opinions from text with the purpose of understanding how they are responding to a situation, such as an advertisement or learning program.

The field of sentiment analysis covers a broad spectrum of perspectives and can be investigated at various levels of granularity [9, 10]. Free text that is mined for opinion can be investigated at sentence or word level. Individual words can be assessed for their degree of

---

[3] The ICT4RED Teacher Professional Development Course is licensed under the Creative Commons Attribution-Non-commercial-Share Alike 3.0 and can be downloaded at http://ict4red.blogspot.com/p/tpd-course.html

positivity or negativity or entire sentences can be analysed. The target of emotive words can also be established. The granularity investigated in this paper is at the word level.
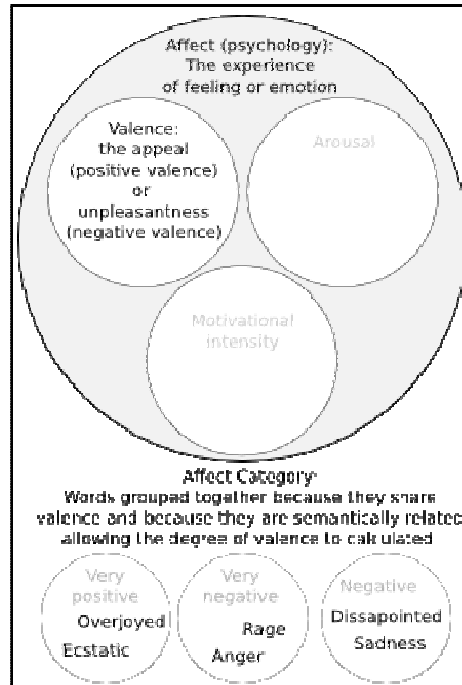

Figure 4: Venn diagram illustrating the relationships between the various sentiment analysis terms being discussed. Valence is the focus of the emotive content that is being extracted from the survey comments.

The sentiment which is the unit of analysis for this paper is a subset of sentiment known as affect. While affect is concerned with three concepts, namely, *valence*, *arousal* and *motivational intensity*, the scope of this paper concerns only one of these, valence. Valence is the measure of the positiveness or negativeness of words which refer to an emotional state. Figure 4 shows the relationships between these terms.

The specifics of the techniques used are discussed in more detail in the section below.

4 Basic techniques of opinion assessment

Teachers were asked to complete a "happy sheet" after each module training was completed. The purpose of the "happy sheets" was to obtain general information about the Module training and facilitator, the relevance of the training to the needs of the teachers and teacher confidence in using the skills learnt in the training. The "happy sheet" instrument included closed ended questions using a 4 point Likert rating scale that asked about general information of the training, relevance of the training and confidence in using the skills obtained in the training and open ended questions that elicited information concerning the experience of the participants. The instrument was administered to all teachers that attended each of the training. The instrument was administered after each of the modules and was collected by the facilitator, captured into a spreadsheet for analysis.

4.1 Likert responses

The closed ended questions were captured through Likert responses [11, 12]. One of the simplest techniques of opinion assessment is surveys. Likert [13] managed to show that a

relatively simple answer scoring mechanism whereby statements of degree of opinions rather than measurements of facts with a strongly negative opinion on one side and a strongly positive opinion at the other end of the scale with each item in the scale scored in increasing amounts could produce as accurate results as much more complex methods of attitude collection while at the same time solving some of the complexities introduced by statistics. It allowed the response from multiple questions to be aligned and compared with each other, thereby facilitating assessment of both the degree of individual answers, the range of multiple responses as well as the overall general attitude. The latter, called the Likert score, can be assessed by finding the average or sum of the resulting numerical values.

4.2   Data mining in free text

The comments of the open ended question section were explored through data mining. Data mining in free text requires an initial cleaning of the data to remove words, such as prepositions, that don't contribute in a meaningful manner to the significance of the text. This process is known as *stop word removal* [9].

Once the stop words have been removed, a lexical database such as WordNet [14, 15] or NRC Word-Emotion Association Lexicon [16, 17] can be consulted. These represent a database of English words grouped by various cognitive concepts, referred to as senses, that also provides word relationship information thereby allowing sentences to be analysed for meaning as well as structure [9].

4.2.1   Valence

Emotion words are the subset of words that are considered to be emotive. The NRC Word-Emotion Association Lexicon (WAL) [16, 17] expands on WordNet [14, 15] by providing additional labels specific to these words. They provide the valence of each of these words by labelling them with a number indicating how positive or negative the word is.

4.2.2   Affect Categories – Emotion beyond positive and negative

Valance can also be addressed on more than one level. Positive, negative or neutral valence can be scored as a whole, but these concepts can be extended by grouping words into sentiment categories around a predefined set of words such as happy, sad and angry. While positive/negative valence is an active area of research, Strapparava and Mihalcea [18] point out that a finer grained annotation could increase the effectiveness of applications of emotion classification. Words grouped together because they share valence and because they are semantically related are known as affect categories and they allow the degree of valence to be calculated.

Ekman [19] points out that a range from pleasant to unpleasant is not sufficient to differentiate between differing emotions since the basic emotions differ in various aspects such as expression, behavioural response and other aspects. For this reason the focus on just positive and negative cannot describe the full picture. There are authors that recognize different emotion groupings, but Ekman [19], who is frequently cited, considers only a basic set of emotion words which share properties. He argues that all emotive words fall in one way or another into a basic set of six emotions consisting of fear, anger, enjoyment, disgust, sadness, surprise and possibly also contempt, shame, guilt, awe and embarrassment.

These basic properties are chosen in acknowledgement of the fact that emotions occur during life tasks and frequently with encounters with people, and by their ability to differentiate emotions from each other as well as from other non-emotional states such as mood and attitude.

Their correlation to individual facial expressions is also noted [19]. It is the shared characteristics of a particular word's emotive content that groups it into one of these categories.

While there can be a multitude of words expressing anger there can be variations within the words in the fear family, so grouping the words into basic families this way does not dismiss the nuances of words within the emotive set. What is does do is organize them in a manner that makes them easier to compare and analyse [19].

Strapparava et.al [18] analysed news headlines both for basic degrees of positiveness as well as adding Ekman's word families pointing out that news headlines were a good target for this analysis since they tend to contain words intended to get attention and therefore frequently contain emotive words.

Two projects were identified as suitable for the task of categorising the affect in the text namely, WAL and the WordNet-Affect [20]. Both projects associate the words with the six affect categories (joy, sadness, anger, fear, disgust, and surprise) identified by Ekman. WAL was given preference due to its larger size of tagged words and due to it being more recent.

5 Method

There were 136 teachers taking part in the second iteration and they were mostly second language English speakers with their primary language being isiXhosa. After each TPD session the facilitator handed out a survey to the participating teachers that they had to fill in. Analysis was done in three parts from the data collected in these surveys.

Firstly, questions using a Likert scale were used to assess the relevance of the course materials, course content and the software applications used during training as well as a rating of how well the course was facilitated and organized. There are two scales which are measured, namely rating and relevance. The data was graphed for easy assessment of the results.

Secondly a free comment field in the survey was investigated at word level granularity for degree of positive or negative feedback. WAL was used to collect the positiveness and negativeness of the comments and to establish which emotional categories predominate. Standard text pre-processing tasks, such as the removal of stop words was done up front and comments that indicated "no comment" were also removed. The removed comments are: No, None, No comment, Not at all, Not for now, Not yet, Nothing and N/A. The following section outlines the limitations that were imposed on the study.

6 Limitations

There are a number of limitations concerning the data that were noted and are outlined:

- The link between language proficiency and use of vocabulary and the resulting range of emotive words are acknowledged but not directly addressed in the study.
- The comments from the survey are very sparse with an average of eight words although one comment contained forty four words.
- Due to the sparseness some techniques are not suitable. In particular, no attempt is made to identify the target word of emotive words and no attempt is made to identify the intended sense of words with multiple senses.
- Surrounding terms, which can increase the positive or negativeness of the emotive words [21] are also not considered.
- Complexity such third person references and words whose changing semantics depending on the surrounding sentence is also not taken into account.
- Some of the "happy sheets" were mislaid and resulted in a sparse dataset for some of the modules.

Having outlined the limitations of the dataset, the results obtained are given in the following section.

7 Results

In these sections we present results of the analysis. The results were taken from 599 completed questionnaires.

7.1 Overall Satisfaction

The literature identifies transactional and overall satisfaction (or cumulative satisfaction), where transactional satisfaction can be regarded as the post choice evaluative judgment of a specific occasion and cumulative satisfaction, an overall evaluation based on the total experience [22]. Broadly seen this implies the level to which the participant expectations were met.

Table 2 and Figure 5 show the satisfaction measurement across all of the learning modules. The Likert score was calculated as the average of the rating multiplied by the score of the rating. This allows us to compare responses from different questions with each other in order to compare satisfaction ratings. It is observed from the results given below that there is a significantly greater "good" and "excellent" rating.

Table 2: Overall rating

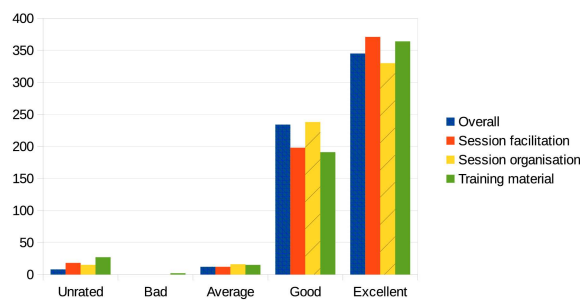| Rating	Overall	Session facilitation	Session organisation	Training material
Unrated (0)	8	18	15	27
Bad (1)				2
Average (2)	12	12	16	15
Good (3)	234	198	238	191
Excellent (4)	345	371	330	364
Likert score	3.52	3.5	3.45	3.45

Figure 5: Overall rating

The satisfaction measurements show that in general it can be stated that most of the participants implied an above average satisfaction with the session organisation, training material and general experience.

7.2 Overall relevance

Table 3 and Figure 6 show the feedback on the perceived relevance of the course content. In this table it can be observed that the Likert score indicates that the relevance is biased towards the rating of "relevant" rather than to "very relevant".

Table 3: Overall relevance

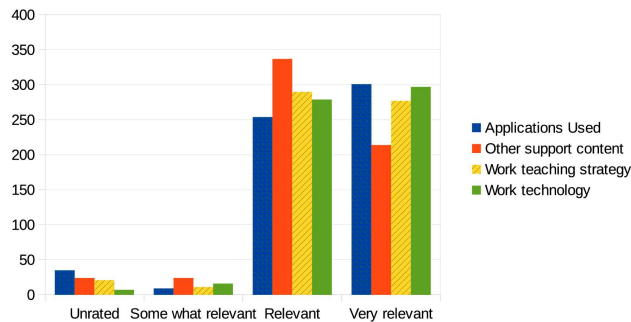| Relevance	Applications Used	Other support content	Work teaching strategy	Work technology
Unrated (0)	35	24	21	7
Somewhat relevant (1)	9	24	11	16
Relevant (2)	254	337	290	279
Very relevant (3)	301	214	277	297
Likert score	2.37	2.23	2.41	2.45

Figure 6: Overall Relevance

The results imply that the participants found the TPD session relevant to very relevant. The technology, content, application was seen as relevant for practice and can be assumed suitable in various degrees to their teaching and learning engagement.

7.3 Overall Rating by Module

Comparing ratings by module required some adjustment since the amount of responses received for each module is inconsistent. In order to compare them more accurately the count of the responses was converted to a percentage by dividing the responses by the module response count. Table 4 shows the actual response count and Figure 7 shows the graphed percentages. In the following the Module 6 should be disregarded as there were some feedback sheets that were misplaced. It can be noted that the response count is varied which indicates that there was not a sufficiently stringent control on the collection and logging of the feedback. This might have been because of the very distributed nature of the TPD iteration. The Likert scale indicates that Module 2, 3 and 4 were considered to be the best received.

Table 4: Responses by module

Module Rating	0	1	2	3	4	5	6
Unrated (0)	4	1		3			
Bad (1)							
Average (2)		3	1	5		3	

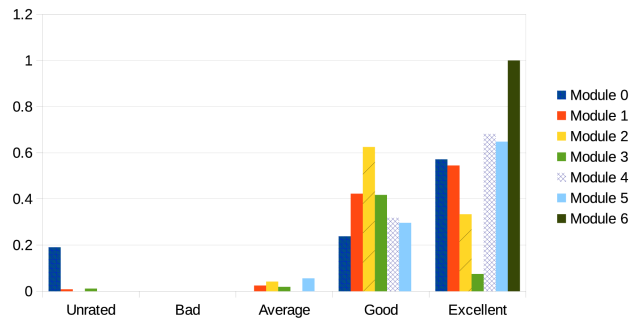|							
Good (3)	5	52	15	111	35	16	
Excellent (4)	12	67	8	147	75	35	1
Response count	21	123	24	266	110	54	1
Likert score	3	3.5	3.63	3.68	3.68	3.59	4

Figure 7: Overall Rating by Module

It is significant that there are no "Bad" ratings. In general it is observed that the modules were well received.

7.4 Positive, negative and neutral sentiment of comments

When considering the open ended questions that elicited information concerning the experience of the participants, the non-emotive words were not removed from the comments.

WAL consists of a text file containing 36632 lines of words that are considered to be emotive words. Each line contains three items, namely a word, a keyword indicating affect or valence, and a zero or one which indicates if the keyword is true or false. Each word appears in the file on nine lines, once for each keyword. An example of the lines for the word "excellent" is shown in table 5.

Table 5:Example for the word "excellent" from NRC Word-Emotion Association Lexicon (WAL) [16, 17]

Word	Key word for affect category from the NRC Word-Emotion Association Lexicon	True (1) or False (0) for keyword
Excellent	anger	0
Excellent	anticipation	0
Excellent	disgust	0
Excellent	fear	0
Excellent	joy	1
Excellent	negative	0
Excellent	Positive	1
Excellent	Sadness	0
Excellent	surprise	0

From Table 5 we can see that the word "excellent" belongs in the "joy" affect category and that the word is "positive" in valence. A word that is neither positive nor negative is considered to be objective or neutral.

In order to determine the degree of valence of the comments, all of the words in the comments for each module were looked up in WAL. The positive, negative and objective matches were counted and then divided by the total number of matched words. This effectively

provides the ratio of positive to negative words. In statistics this is referred to as the arithmetic mean. The results are shown in Table 5. It is clear from the table that while most of the emotive words in the comments were emotionally neutral for all of the modules, the remaining emotive words were mostly positive.

Table 6: Positive, negative and neutral sentiment of comments

Module	0	1	2	3	4	5	6
Positive	0.1623	0.1623	0.1637	0.1383	0.1602	0.1270	0.2110
Negative	0.0444	0.0444	0.0299	0.0131	0.0508	0.0492	0.0327
Objective	0.7931	0.7931	0.8062	0.8485	0.7888	0.8237	0.7561

A graph of just the positive and negative of the results in Table 5. is presented in Figure 8. This clearly show that modules 4 and 5 received the most negative comments while Module 3's comments were the most positive..
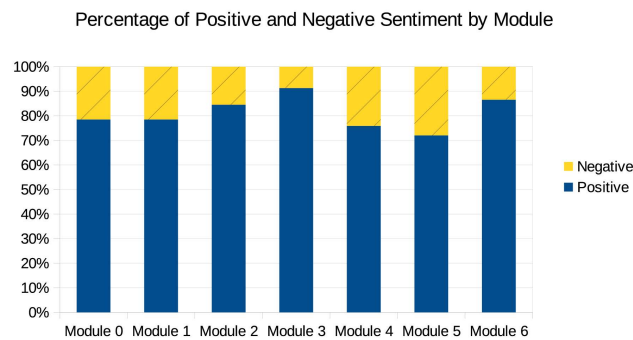

Figure 8: Positive, negative and objective sentiment of comments

In general the feedback again indicates a majority positive sentiment with the valence skewed towards the positive.

7.5 Affect categories

Once stemmed there were 897 words remaining in the comments of which 140 conveyed affect and 187 were categorised by WAL as non-neutral. The manner in which this was done was discussed in detail in section 7.4. In addition, Table 7 also lists the words with the highest occurrences in the affect category. The fact that only 23% of the words are in the negative affect categories once again supports the premise that the learning programme was well received.

Joy and surprise were the two most predominant affect categories respectfully and the valence was predominantly positive. The affect categories, word counts and example words are detailed in Table 7.

Table 7: Count of words matching the six Ekman categories

Affect	Word count	Example words
joy	71	excellent, good, happy, lovely, happy, gain, excellent, successful, helpful, promise, fun, proud, powerful, engaging, outstanding
surprise	36	good, lovely, exciting, chance
fear	14	problem, difficult, change, powerful

sad	11	problem, difficulty, disappointing, tough
anger	6	barrier, difficulty, battery, challenge
disgust	2	interested, powerful

The participants implied that they experienced affect "joy" and "surprise" in the course. This confirms the previously collected data from the rest of the survey and allows us to triangulate the results.

8   Conclusion

This paper reported on an attempt to gauge the experience of the second iteration of the ICT4RED TPD course through sentiment analysis. The results show that teachers were happy with the training and felt strongly that the content and skills provided in the Course are relevant for them. Teachers also indicated that they are confident in using the skills obtained in the classroom. The high level of confidence revealed by teacher's show that they believe that they have obtained skills and are positive that they can implement the skills obtained in the classroom. The actual flow through of the skills to the classroom practice was tracked through the obtainment of the relevant badges as micro-accreditation of the interim goals. All the participating teachers obtained the badges needed and graduated having demonstrated their application of the skills.

9   References

[1].   DST, *The Cofimvaba Schools District Technology Project: Status Report 2*, 2013, Department of Science and Technology: Pretoria. p. 12.
[2].   Ford, M., *eTexbook Initiative*, 2012, CSIR.
[3].   Herselman, M. and A. Botha, *ICT4REDResearch Framework: September 2013*, 2013, CSIR Meraka.
[4].   Botha, A., et al. *Towards a Mobile Learning Curriculum*. in *IST-Africa 2012*. 2012. Dar es Salaam, Tanzania: IIMC International Information Management Corporation.
[5].   Botha, A., M. Herselman, and M. Ford. *Gamification beyond badges*. in *IST Africa*. 2014. Mauritius.
[6].   Deterding, S., *Eudaimonic Design, or: Six Invitations to Rethink Gamification*, in *Rethinking Gamification*, M. Fuchs, et al., Editors. 2014, Meson press Lüneburg.
[7].   Voogt, J. and G. Knezek, *IT in primary and secondary education: emerging issues*, in *International handbook of information technology in primary and secondary education*, J. Voogt and G. Knezek, Editors. 2008, Springer: New York.
[8].   Peffers, K., et al., *A design science research methodology for information systems research*. Journal of Management Information Systems, 2007. **24**(3): p. 45 - 77.
[9].   Martin, J.H. and D. Jurafsky, *Speech and language processing*. International Edition2000, Upper Saddle River, New Jersey: Prentice Hall, Pearson Education International}.
[10].   Baccianella, S., A. Esuli, and F. Sebastiani. *SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining*. in *LREC*. 2010.
[11].   Matell, M.S. and J. Jacoby, *Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties*. Journal of Applied Psychology, 1972. **56**(6): p. 506.
[12].   Garland, R., *The mid-point on a rating scale: Is it desirable*. Marketing Bulletin, 1991. **2**(1): p. 66-70.
[13].   Likert, R., *A technique for the measurement of attitudes*. Archives of psychology, 1932.

[14]. Miller, G.A., *WordNet: a lexical database for English.* Communications of the ACM, 1995. **38**(11): p. 39-41.

[15]. Fellbaum, C. *WordNet: An Electronic Lexical Database*. 1998; Available from: http://wordnet.princeton.edu.

[16]. NRC Word-Emotion Association. *NRC Word-Emotion Association Lexicon (Version 0.92)*. 2011 [cited 2014 December]; Available from: http://www.saifmohammad.com/WebPages/lexicons.html

[17]. Mohammad, S.M. and P.D. Turney, *Crowd-sourcing a word–emotion association lexicon.* Computational Intelligence, 2013. **29**(3): p. 436–465.

[18]. Strapparava, C. and R. Mihalcea. *Learning to identify emotions in text*. in *Proceedings of the 2008 ACM symposium on Applied computing*. 2008. ACM.

[19]. Ekman, P., *An argument for basic emotions.* Cognition & Emotion, 1992. **6**(3-4): p. 169-200.

[20]. Strapparava, C. and A. Valitutti. *WordNet Affect: an Affective Extension of WordNet*. in *LREC*. 2004.

[21]. Kennedy, A. and D. Inkpen, *Sentiment classification of movie reviews using contextual valence shifters.* Computational Intelligence, 2006. **22**(2): p. 110-125.

[22]. Oliver, R.L., *Satisfaction: A behavioral perspective on the consumer*2010: ME Sharpe.