Understanding the Difficulty Factors for Learning Materials: A Qualitative Study

Keejun Han, Mun Y. Yi, Gahgene Gweon, and Jae-Gil Lee

Knowledge Service Engineering Department,
Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea
{keejun.han,munyi,ggweon,jaegil}@kaist.ac.kr

Abstract. Difficult materials overwhelm learners whereas easy materials deter advanced knowledge acquisition. Toward the goal of automatic assessment of learning materials, we conducted a laboratory experiment involving 50 college students recruited from two universities in Korea using 115 PowerPoint files. On the basis of the qualitative analysis results, we propose a model of learning difficulty, distinguishing measurable factors from non-measurable factors. The most influential factors for the easiest and the hardest learning materials are also identified and compared. The study findings have implications for educational service providers who need to automatically classify learning materials based on their innate difficulties.

Keywords: Learning material difficulty, Recommender system, PowerPoint slides, Difficulty factors, Coding analysis.

1 Introduction

Toward the goal of automatic assessment of learning materials, this paper reports the findings of a laboratory experiment conducted to identify the factors that underlie the difficulty of PowerPoint slides from college students because of its commonality and popularity as educational source. In short, the objective of the study reported in this paper was to discover the factors that determine the difficulty of learning materials in general, and PowerPoint slides in specific, on the basis of user comments. The findings have significant implications for the development of an autonomous difficulty classifier, which can be easily incorporated into search engines and online learning service platforms.

The oldest method for measuring the difficulty of a document was to set up mathematical formulas that utilize lexical features of the document [1, 2]. There are also alternative approaches of applying machine learning techniques to estimate the difficulty of a document [3,4,5]. However, those approaches have limited value in assessing the difficulty of learning materials, particularly of PowerPoint slides, because they only focus on the textual sources of documents. Thus, the difficulty dimensions we propose in this paper can be considered more complete as they are applicable to the learning materials that consist of both textual and graphical sources.

K. Yacef et al. (Eds.): AIED 2013, LNAI 7926, pp. 615–618, 2013.

2 Method

We conducted a laboratory experiment involving 50 college students recruited from two universities in Korea. Each participant examined five PowerPoint files while thinking aloud about the difficulty aspects of each slide. In the end of each session, they were asked to choose the easiest and hardest learning materials out of five. All participants' utterances were recorded and transcribed.

The average age of the participants was 21.3. The youngest participant was 18 years old, and the oldest was 32. Participants have been using PowerPoint slides for 4.74 years on average. The shortest period of using PowerPoint slides was 1 year, and the longest was 12 years.

Thirty transcripts were analyzed by two coders; their inter-coder reliability was 0.87. Overall, the iterative coding process identified a total number of 41 difficulty factors out of 3150 initial units of utterances obtained from the 50 transcripts.

We further conducted a card-sorting study to empirically examine the mapping between the 41 difficulty factors, which were derived from the coding analysis, and the 7 principal categories (groups of similar factors), which were mainly theorized by the authors until that time with the help from prior research on difficulty. Two types of measurements, agreement and correlation proposed in [6], were calculated in order to evaluate the results of the card sorting study. The agreement scores for each category ranged from 0.5 to 0.84, with the average score of 0.64, showing that those categories were reasonably well understood across the participants. A correlational analysis conducted between the 41 difficulty factors and their corresponding category identified by the participants showed that 32 factors (78%) had a correlation value greater than 0.75, which is considered high [6]. The remaining 9 factors had a correlation value between 0.6 and 0.7. In addition, 39 out of 41 factors were placed in only one or two categories, implying that each category is highly distinct.

3 Results

After card-sorting, we further distinguished the 41 factors into those that are automatically measurable by computer versus not. This distinction was made so that researchers who are building automatic classifier for learning material difficulty could consider using these measurable factors. The distribution of the automatically measurable factors and non-measurable factors over the 7 principal categories are presented as follows:

- Detailedness: Factors that represent how comprehensible and concrete the slides are.
    - Measurable: Highlighting important terms, Presence of examples, Presence of formula, Presence of tables, Presence of visual materials, Presence of external links, Brief summary for visual materials
    - Non-measurable: Detailedness of visual materials, Detailedness of text, Presence of animation effects

- Structural Completeness: Factors that represent how comprehensible & concrete the slides are.
    - Measurable: Presence of a summary, Presence of sub-titles, Presence of bullets, Presence of numbering, Presence of grocery terms, Presence of a table of contents, Presence of Q&A
- Relevancy: Factors that capture how appropriate the slide components are.
    - Measurable: Title relevancy, Visual material relevancy, Similarity between slides and its origin
    - Non-measurable: Animation effect relevancy
- Flow: Factors that represent how logically coherent the slides are.
    - Measurable: Similarity between adjacent slides
    - Non-measurable: A logical order of contents
- Readability: Factors that indicate how well the text is comprehensible.
    - Measurable: Term difficulty, Topic difficulty in a domain
- Length: Factors that capture the size of the presentation.
    - Measurable: The length of slides, The number of words in a page, The number of tables, The number of formula, The number of examples, The number of external links, Topic coverage, The number of visual materials
    - Non-measurable: The number of animation effects
- Formatting Style: Factors that capture the appearance of slides.
    - Measurable: Font size, Language used
    - Non-measurable: The number of colors used, Background color, Text color, Visual attractiveness of visual components (figures, graphs, animations), Visual attractiveness of non-visual components

We further examined the factors that were most frequently mentioned regarding whether a given PowerPoint slide material was easy or difficult. Table 1 shows the top factors that contributed in determining each difficulty level, as well as the frequency of each factor. Recall that we had fifty participants. Therefore, a frequency of 15 for a given factor means that 30% of the participants listed that factor as a determinant.

Certain factors are listed as being influential for both easy and difficult levels of learning materials. Such factors differed in terms of its value. For example, the top factor for both levels of difficulty is "topic difficulty in a domain." For the "easy" list, this means that the topic itself was not difficult, whereas for the "difficult" list, the topic itself was difficult. Another example is "presence of visual materials." In the "easy" list, this factor tells us that if there are visual elements in a learning material, it tends to be easy. However, in the "difficult" list, this factor means that absence of visual elements makes a learning material difficult. Half of the factors that made the top lists are unique to each difficulty level. Therefore, different factors should be accounted for depending on the difficulty level.

Table 1. Top N Influential Difficulty Factors for the Easiest and Hardest Learning Materials

Difficulty factor for **easiest** learning material	Freq	Difficulty factor for **most difficult** learning material	Freq
Topic difficulty in a domain	15	Topic difficulty in a domain	14
Presence of visual materials	14	Number of words in a page	14
Summary for visual materials	13	Presence of visual materials	10
Presence of examples	13	Highlighting important terms	8
Highlighting important terms	8	Number of visual materials	7
Presence of Q&A	8	Term difficulty	7
The number of example	7	Summary for visual materials	6

4 Conclusion

In this research, we conducted a qualitative study to identify the factors that affect the difficulty of learning materials, in particular PowerPoint slides. Going through the coding and card-sorting processes, we developed a model of difficulty factors over the seven principal categories of learning difficulty. Further, through the difficulty factor comparison analysis, we identified top influential factors for determining whether a given learning material is relatively easy or difficult. Our proposed model of difficulty factors can benefit online educational service providers who want to automatically sort their learning materials in terms of the material's innate difficulty.

Acknowledgement. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0029185).

References

1. Flesch, R.: A new readability yardstick. Journal of Applied Psychology 32, 221–233 (1948)
2. Dale, E., Chall, J.S.: A formula for predicting readability. Educational Research Bulletin 27, 1–20 (1948)
3. Peterson, S.E., Ostendorf, M.: A machine learning approach to reading level assessment. Computer Speech and Language 23, 89–106 (2008)
4. Chen, W., Persen, R.: A Recommender System for Collaborative Knowledge. In: Artificial Intelligence in Education (2009)
5. White, S.: Mining the Text: 34 Text Features That Can Ease or Obstruct, Text Comprehension and Use. Literacy Research and Instruction 51(2), 143–164 (2012)
6. Spencer, D.: Card sorting: Designing usable categories. Rosenfield Media (2009)