Passive Versus Active: Frameworks of Active Learning for Linking Humans to Machines

Jaeseo Lim[*,1] **(jaeseolim@snu.ac.kr)** **Hwiyeol Jo**[*,2] **(hwiyeolj@gmail.com)**
Byoung-Tak Zhang[†,1,3] **(btzhang@bi.snu.ac.kr)** **Jooyong Park**[†,1,4] **(jooypark@snu.ac.kr)**

[1]Interdisciplinary Program in Cognitive Science, Seoul National University, Seoul, Republic of Korea
[2]Institute of Computer Technology, Seoul National University, Seoul, Republic of Korea
[3]Department of Computer Science and Engineering, Seoul National University, Seoul, Republic of Korea
[4]Department of Psychology, Seoul National University, Seoul, Republic of Korea

Abstract

There are numerous studies that show that the more learner actively participate in the learning process, the more they learn. Although the use of active learning to increase learning outcomes has been recently introduced in a variety of methods, empirical experiments are lacking. In this study, we introduce two frameworks of human active learning and then conducted two experiments to determine how these frameworks can be used as leaning tools. In experiment 1, we compared three types of active learning and passive learning in order to empirically confirm the effect of active learning. In experiment 2, based on the results of experiment 1, we explored through simulation on machine learning with the frameworks that the more active the learners are, the better outcomes can be obtained. Both experiments showed that active learning both also effective in human and machine learning. Therefore, our analyses of the two experiments fit within the taxonomy and classification of the frameworks of active learning. This result is further significant in that it gives practical implications on human and machine learning methods.

Keywords: active learning; DOLA framework; ICAP framework; knowledge transfer; learning outcome.

Educators and educational researchers have long perceived that students learn much better by learning actively (Chi & Wylie, 2014). Bonwell and Eison (1991) argued that active learning is defined as learning that requires students to engage cognitively and meaningfully with the learning materials. In particular, students are required to get involved with the learning material by considering, analyzing, synthesizing, and evaluating it rather than receiving it passively (King, 1993). Therefore, active learning refers to learner-centered instructional methods that dynamically engage learners in the learning process (Chi, 2009). The main constructs of active learning include the learner's cognitive engagement and participation with knowledge construction through meaningful learning experiences, and interaction between other learners during learning process (Chi, 2009).

Active learning has been studied in many fields, including education, learning sciences, educational psychology, and STEM education (Heller, Beil, Dam, & Haerum, 2010). However, although active learning has received considerable interest over the past several years, the current educational system, which emphasizes the amount of knowledge, naturally sets the delivery of knowledge as the goal of learning (Lim, Jo, Zhang, & Park, 2020). Accordingly, most schools, universities, and educational institutions educate students through lectures. While lecture-centered classes have the advantage of being able to deliver a lot of knowledge, such characteristic does not directly lead to students' achievements. Rather, there are many studies that show the problems of one-way lectures (Bonawitz et al., 2011; Wieman & Perkins, 2005; Mayer & Alexander, 2016).

This paper has three major sections. The first section describes why active learning is needed in the learning process. The second section outlines the framework of active learning based on the DOLA framework (Menekse, Stump, Krause, & Chi, 2013) and the ICAP framework (Chi & Wylie, 2014). In the third section, we show the results of the two experiments conducted for humans and machines, considering that application of the human active learning.

Being Passive Versus Active

Passive learning often means teacher-centered methods that favor direct instruction in which students usually learn through listening to and observing lectures presented by a professor or lecturer. Likewise, rehearsing or repeating words is rather passive because it leads to less learning and recalling as compared to deeper levels of processing (Chi, 2009). However, simply conveying a lot of information does not lead to learners' achievements. Based on research which compared lectures and other teaching methods, Bligh (1998) claims that traditional lecturing is as effective as other methods in relaying information, but not as effective in encouraging discussions, in changing attitude, and acquiring skills. Thus, it is the instructors' erroneous notion that lectures can draw students' attention, stimulate their motivation, or improve academic performance.

There is also an experimental comparison of students' learning outcome (Hrepic, Zollman, & Rebello, 2007; Hake, 1998). This study investigated the effect of listening to lectures with preliminary questions. As a result, although the lecturer delivered a lot of information, the students did not gain much knowledge. In response to the findings, Wieman and Perkins (2005) suggested that the amount of information may have been too overwhelming for students to acquire. There are also studies that support this opinion (Poh, Swenson, & Picard, 2010). They had participants wear a wristband which measures the sympathetic nerve system's activity through skin conductive response. While they had the device,

[*]Equal contribution
[†]Shared corresponding authors

they were told to record their own activity at a certain time. As a result of the analysis of the integrated data, the activity level was high when doing homework by themselves or taking tests, but the nerve activity level in listening to a lecture was as low as watching TV or doing house chores. The findings correspond to Bligh's claim that it is delusional to think one-way lectures are effective for drawing attention, or improving academic level.

As an alternative to strategies of passive learning, various methods have been researched to increase learning outcomes. They are called active learning, which requires learners' cognitive intervention (Bonwell & Eison, 1991). According to Menekse et al. (2013), the main constructs of active learning are learners' engagement with concrete learning experiences, knowledge construction through meaningful activities, and some degree of interaction between learners during the learning process. Therefore, active learning means learner-centered instructional approaches that dynamically involve learners in the learning process.

Many research regarding active learning support that their effectiveness of these methods has to do with the learners' performance during learning process. For example, in the context of learning, these learning activities involve activities such as paraphrasing, manipulating objects and so forth (Igo, Bruning, & McCrudden, 2005; Chi, Roy, & Hausmann, 2008). In addition, Johnson and colleagues found that active learning such as cooperative, and collaborative activities during learning improved students' learning outcomes (D. Johnson, Johnson, & Smith, 1991; D. W. Johnson, Johnson, & Smith, 1998). Therefore, for various measures of learning outcome, active learning have been shown to exceed passive learning in many aspects (Chi & Wylie, 2014; Lambert & McCombs, 1998).

Frameworks of Active Learning

The concept of active learning tends to be used differently depending on the researchers. And the scope of active learning is also too comprehensive. To address the lack of taxonomy and framework about active learning, Chi and colleagues proposed the Differentiated Overt Learning Activities (DOLA) framework. The DOLA framework divided active learning methods into three modes: interactive, constructive, and active, depending on what learning activities students overtly show (Menekse et al., 2013). It is the characteristics of this framework which differentiates a variety of a learner's engagement in the learning process that have previously been considered by many research about active learning. This framework also insists that different modes of students' engagement have different learning outcomes. Because of the different features they have, these modes of learning activities involve different cognitive processes.

The framework assumes that being active is expected to involve learners' manipulation with the learning materials; being constructive is to facilitate the generation of novel thoughts, beyond those presented. And being interactive is expected to generate new ideas with peers, but both learners are contributing substantial mental efforts. These active modes of engagement, as defined, also predict learning outcomes. Interactive modes are superior to constructive activities, followed by active modes. All these active modes are better than the passive mode (Chi & Wylie, 2014; Menekse et al., 2013). This is in the same context as the Interactive-Constructive-Active-Passive framework below.

Chi and Wylie (2014) proposed the Interactive-Constructive-Active-Passive (ICAP) framework as a segmentation of active learning. The ICAP framework, referred to in previous publications as DOLA for Differentiated Overt Learning Activities (Chi, 2009), consists of a taxonomy (Chi & Wylie, 2014). Student learning is more effective in interactive learning than constructive learning, which is more effective than active learning activities, which are superior than passive learning (Chi & Wylie, 2014). Thus, the ICAP framework includes a taxonomy that differentiates three modes of active learning, referred to in the DOLA framework.

The core of the framework mainly insists that an interactive mode can enhance learning outcome more than a constructive mode, which is also superior to any active mode of engagement for learning. And these three types of active modes are better than the passive mode. All in all, the sequence I, C, A and P is due to the cognitive engagement of learners, and through the knowledge-change process in these orders of each mode, learners have hierarchical learning outcome.

Although these taxonomies and frameworks for active learning to increase learners' outcomes have been introduced in a variety of fields, there are not many attempts to analyze systematically. The reason why passive learning such as lectures is considered that not only do people think it is a more convenient method for both teachers and students, but also providing vast amount of information is effective in learning. Therefore, the current study is to examine to investigate the effectiveness of active learning based on these frameworks of active learning, and identify its impact on learners' learning outcomes. Thus, we compared passive, active, constructive, and interactive modes of learning: lecture group (Passive), self-study group (Active), lecture and review group (Constructive), and two discussion groups (Interactive). However, it is only a relative difference, not a discrete division. For example, self-study groups are more or less active than lecture groups. Nevertheless, in line with the previous findings that active learning will have a positive influence on actual learners' performance, the hypothesis of the present study was that the discussion groups would score higher than lecture and review group, which is better than the self-study group, which is better than the lecture group (Prince, 2004; Schroeder, Scott, Tolson, Huang, & Lee, 2007; Chi & Wylie, 2014).

For the next step, we explore through simulation on machine learning with frameworks that show the active the agents are, the better learning outcomes they can be obtained. Because the experiment with machines can complement the limitations of experiment with human such as sampling bias and subjectivity, thus, we aimed to maximize the effectiveness of by applying the human active learning framework to machines. In order to form active learning in machines, we have set up teacher models and student models for both passive and active learning.

Experiment 1

In experiment 1, we aimed to find out what learning methods produce better outcomes in class. Thus, based on the DOLA and ICAP framework, we have set five groups: lecture group (P), self-study group (A), lecture and review group (C), and two discussion groups (I).

Method

Participants and Design. Undergraduate students in a selective university participated in this experiment. A total of 138 students participated, with 63 females. Students were assigned to each group randomly: the lecture group (L group, *n* = 25), the self-study group (S group, *n* = 25), the lecture and review group (LR group, *n* = 30), the lecture and discussion group (LD group, *n* = 29), and the self-study and discussion group (SD group, *n* = 29). Three or four students formed a discussion group. There were no significant differences in age among the groups.

Learning Material. The instructional lecture used in the experiment was a monologue-style. It was a lecture on law available and dealt with accusation, charge, and recognition as criminal procedure code. This is because these topics were less likely to be affected by prior knowledge since that are not available for undergraduate course. Written learning material was created from the lecture, and was seven-page long.

Test Items. The final test items were categorized into three: (1) Verbatim type items. It consisted of short answer questions, requiring memorization for the material; (2) Paraphrased type items. They are questions where participants have to explain a concept from the material; (3) Transfer type items. It included all the concepts in the material, and one's understanding on the whole learning contents to apply in a new situation and explain (i.e., questions with novel scenarios).

Procedures. The students first took a survey to check their levels of knowledge and interest on the topics. The detailed progress of each group was as follows: the L groups listened to a lecture without any physical manipulation for 36 minutes. The S groups studied the learning material by themselves for 36 minutes. The S groups was also free of physical manipulations such as underlining during the learning process. The LR groups listened to a lecture for 18 minutes and then studied the provided written learning material by themselves without any physical manipulation for 18 minutes. Students of LD and SD groups listened to lecture or studied the written learning materials by themselves for 18 minutes and then discussed in groups of three or four for another 18 minutes. In fact, the total amount of learning time for both groups was the same. Lastly, all two groups took a twenty-minute final test.

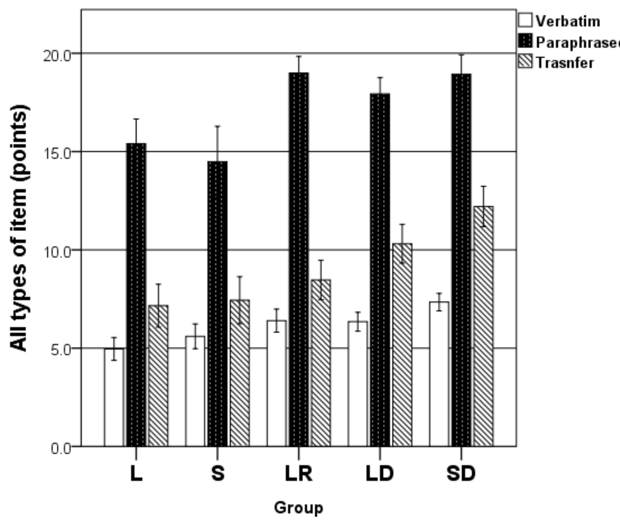

Figure 1: Comparisons of the final test of experiments 1 as a function of different learning. Note. L = lecture; S = self-study; LR = lecture and review; LD = lecture and discussion; SD = self-study and discussion. All types of item consist of verbatim, paraphrased, and transfer type items. Error bars are ± 2SE.

Result and Discussion

Prior to full-scale analysis, there were no significant difference in background knowledge survey between the five groups. That is, most of the students had no prior knowledge on the learning topics.

The mean values and standard deviations of the test are shown in Figure 1. We carried out an analysis of covariance (ANCOVA) and post-hoc using Tukey to find out any difference between the five groups. The total mean values of the L group (M = 27.52, SD = 5.03) and the S group (M = 27.52, SD = 5.89) were significantly lower than those of other three groups, $p < .001$. The L group (M = 4.96, SD = 1.46) scored significantly lower than other four groups in verbatim type items, $ps < .01$, and the S group (M = 5.60, SD = 1.58) scored only lower than the SD group, $p < .001$. The L group (M = 15.40, SD = 3.14) and the S group (M = 14.48, SD = 4.52) scored significantly lower than the other three groups in the paraphrased type items, $ps < .05$. For transfer type items, the L group (M = 7.16, SD = 2.73) and the S group (M = 7.44, SD = 3.00) scored significantly lower than the two discussion groups, $ps < .01$, but showed no significant difference from the LR group.

Therefore, these results suggest that interactive modes of learning enhance scores for transfer type items. In particular, students' learning performance was high in the order of passive < active < constructive < interactive. Consistent with the DOLA framework and ICAP framework, the findings showed the better learning outcomes in active learning. Although it varies slightly depending on the test items, the finding that a large difference was found in transfer type

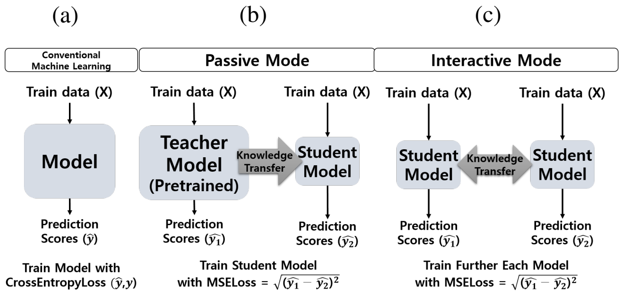

Figure 2: Illustration of the training frameworks. (a) the conventional training framework. `Pretrained` means the model is trained like (a) before using it. (b) passive learning using knowledge distillation (in this case, transfer), which minimizes the loss between the prediction scores of the two models instead of the loss between the prediction scores and the true labels. (b) and (c) imitate passive mode and interactive mode, respectively.

questions is the most important practical implication. Transfer is very hard to achieve (Muldner, Lam, & Chi, 2014), and because transfer is the ability to apply existing knowledge to a new situation or environment, it is also the ultimate goal of education.

Subsequently, we compared active learning with passive learning in machines in order to further extend and even validate the results of human active learning. In order to examine the extensibility of the frameworks of active learning, experiment 2 was conducted.

Experiment 2

Active learning has often been used in machine learning but with different concepts. The method in machines, which query datasets to be labeled for training by an oracle, may get better accuracy (Settles, 2010). However, it has limitations that many of the queries generated by the learner included no recognizable symbols, no semantic meanings (Settles, 2011).

In experiment 2, we simulated on machine learning with the frameworks of 'human' active learning in order to verify the result that the more active in the learning process the better learning outcomes can be obtained (Chi & Wylie, 2014; Chi, 2009; Menekse et al., 2013). Like experiment 1, we implemented the human learning process but the interactive mode, the most effective learning method, and passive mode were only compared.

Method

Task, Datasets, and Classifiers. Although the task of experiment 1 was open-ended question answering, we set experiment 2 as text classification. This is because open-ended question answering is still hard problem to be solved and the model to solve the problem should be complex and large. Therefore, we attempt to use simple and basic problem.

We used 5 benchmark text classification datasets. 3 are topic classification: DBpedia ontology (Lehmann et al., 2015), YahooAnswers (Chang, Ratinov, Roth, & Srikumar, 2008), AGNews and 2 are sentiment classification: Yelp reviews (Zhang, Zhao, & LeCun, 2015), IMDB (Maas et al., 2011). For the details on the datasets, refer to the original papers.

Next, we used TextCNN (Kim, 2014) and LSTM (Hochreiter & Schmidhuber, 1997) as classifiers but we made the difference between passive learning and active learning on model capacity. Passive learning required a teacher model (M_T), which is able to learn from the data fully. On the other hand, student models (M_S) only represented novice learners. In TextCNN classifiers, the teacher model architecture consisted of 2 convolution layers, which had 32 and 16 channels, respectively. We also utilized the multi-kernel approach, which kernel sizes were 2, 3, 4, and 5. The student model architecture consisted of 1 convolution layer, which had 32 channels only. And, its kernel size were 2 and 3. In LSTM classifiers, the teacher model architecture consisted of forward and backward LSTM layers with 300 hidden units. In contrast, the student model architecture had forward LSTM only with 150 hidden units. We used Adam optimizer (Kingma & Ba, 2014) and Cross Entropy Loss to train the models. The other hyperparameters (e.g. learning rate, batch size, etc.) were the same.

Implementation. Knowledge transfer, for example teaching (lecturing), and discussion in experiment 1, was implemented by knowledge distillation (Hinton, Vinyals, & Dean, 2015). The method did not use training data directly, instead it used other models to train a model. Specifically, a model can learn the other models' prediction scores on the training data. The transfer was implemented by mean squared error loss between two model predictions as the original paper suggested. By using the idea, the training frameworks were illustrated in Figure 2. Passive learning (b) used a teacher model (M_T) and a student model (M_S). Both models are trained on the conventional training framework (see (a)) and then knowledge transfer occurred from M_T to M_S; it imitates "the teacher provides knowledge to the student." Lastly, (c) imitates active learning used in experiment 1. Limiting the time (36 mins in passive learning vs. 18 mins in active learning) in experiment 1 corresponds to constraint the training capacity for machine. Therefore, we used two M_S, which is much smaller than M_T. Beside, in order to implement the discussion process, we simply made the knowledge transfer (distillation) in bidirectional ways. Overall method imitated that "students use their knowledge (inter)actively to make better results."

Result and Discussion

The performance of the passive learning framework and the interactive learning framework were presented in Table 1. When we compared the result between the passive and the interactive, the interactive learning performed better in most of the datasets. These results supported our hypothesis that active learning enhances performance, as observed in experiment 1. The other minor findings are as follows: the perfor-

Table 1: The performance of training frameworks on the text classification. x denotes training data, M_T denotes the teacher model, which had larger model capacity than the student model (M_S). The arrows ($\rightarrow$ and $\leftrightarrow$) describe the flow of knowledge transfer. Passive and Active can be symbolized as $[x \rightarrow M_T] \rightarrow M_S$ and $[x \rightarrow M_S] \leftrightarrow [x \rightarrow M_S]$, respectively.

Classifier	**Methods**	IMDB	Yelp	AGNews	Yahoo	DBpedia
CNN	$x \rightarrow M_T$	76.61±.73	56.36±.17	88.85±.38	65.45±.30	98.04±.11
	$x \rightarrow M_S$	78.70±.13	56.31±.25	89.54±.12	67.92±.25	98.01±.03
	Passive	78.89±.37	56.60±.08	89.68±.28	66.01±.36	97.85±.12
	Active	**79.04±.28**	**56.79±.15**	**90.21±.13**	**68.69±.10**	**98.14±.03**
LSTM	$x \rightarrow M_T$	77.05±.13	58.94±.19	89.38±.34	72.23±.20.	98.43±.05
	$x \rightarrow M_S$	77.10±.25	58.26±.24	89.45±.47	71.63±.75	98.26±.06
	Passive	77.55±.82	58.90±.20	89.74±.06	72.93±.78	98.33±.02
	Active	**77.58±.16**	**59.00±.14**	**90.53±.23**	**74.44±.55**	**98.67±.06**

mance of the passive learning framework were on par with the conventional learning framework ($x \rightarrow M_S$), and even were better on several datasets. This means that the knowledge captured by the teacher model was successfully transferred to the student model. Besides, in some datasets the teacher model was overfitted to the training data, so their performance on test data was worse than the student model.

To sum up, we tried to imitate interaction in the learning framework of humans on that of machines and used the framework to solve a core NLP task. The result from both humans and machines implies that the more active the learners (models) were, the better outcomes was obtained. Furthermore, the result showed that the idea of the human active learning can be applied to machines, showing better performance than passive learning. In other words, through simulation of machines, we empirically explored the applicability and extensibility of frameworks of human active learning to machines.

General Discussion

Passive learning, such as listening to a lecture has been the predominant instruction ever since universities were founded (Freeman et al., 2014; Lim et al., 2019, 2020). Despite the limitations of passive learning, lecture-centered classes are still widespread. This is because students are more used to lectures getting a feeling of having gained knowledge by just listening. Instructors also prefer lectures because they think that lectures are the most effective way to control the classroom environment and deliver a lot of knowledge. However, passive learning is not associated with an actual increase in learners' knowledge, and also has a great limitation that learners cannot develop their own thinking.

Instead, active learning emphasizes the need for learners to construct their own understanding and through this, they are expected to improve their learning outcomes. However, active learning has been challenged by the traditional and lecture-centered methods, because there has been a lack of empirical and systematical analysis on how these learning methods impact students' performance. Accordingly, the present study was performed to reconfirm the effectiveness of active learning, based on the frameworks of active learning: the DOLA and ICAP framework (Chi & Wylie, 2014; Menekse et al., 2013).

In this study, we conducted two experiments in order to investigate the effect of active learning. Active learning is expected to enhance students' performance better than passive learning because actively participating in class allows the students to activate relevant knowledge, thereby allowing the student to assimilate novel information to fill in the knowledge gaps, whereas passive learning only allows to store novel information infrequently (Menekse et al., 2013). In particular, active modes, as defined, also predict learning outcomes. Interactive modes are superior to constructive activities, and active modes in order: I $\geq$ C $\geq$ A $\geq$ P. With this expectation, in experiment 1, we compared five conditions: lecture (P), self-study (A), lecture and review (C), and two other discussions (I). As a result, the two discussion groups scored higher than the lecture and review group, which was better than the self-study group. The lecture group had the lowest score, as expected. There were slightly different aspects depending on the types of item, but for the most important transfer type items, the results came out in this order. These findings also correspond with the frameworks introduced in that learning performance would be greater in active learning than in passive learning.

In experiment 2, we compared performance of the active learning with the passive learning in machines. In detail, interactive mode, which was the most effective learning method, and passive mode were compared, based on the results of experiment 1. Like in experiment 1, machines also increased their performance when they performed as active learning; the two student models which exchanged their opinions (i.e. distilled) were the most effective.

In the end, the results of these two experiments are summarized as follows: First, active learning is very effective in actual class, and it works better as learners become more active. Although it was an exploratory and simplified study, we showed that the frameworks of human active learning worked for machines as well. Moreover, adopting more sophisticated techniques in implementing knowledge transfer can be used to overcome several limitations of presented frameworks, increasing the model performances.

These are the limitations: First, it is worth pointing out that the students in this study discussed on their own without any intervention. Thus, future study can consider various learning activities before having a discussion (Lim et al., 2019), or interventions from the instructor during the discussion.

Second, as previously explained, group classification in human experiments is not discrete. The lecture group can be said to be passive by limiting physical manipulation, but the boundaries between the self-study group and the review group is not clear. Nevertheless, there was a difference in relative overt active activity, which led to the higher test scores.

Last but not least, human and machine studies are not exactly symmetrical: the tasks are different (open-ended question answering vs. text classification) as well as the discussion or knowledge transfer process in humans and machines are not the same. However, it is meaningful in that we applied the interactive mode of human active learning to machines, and tried to verify the effectiveness of the human learning framework through machines.

In the past, constructivists asserted that human learning should emphasize two key principles: learning must be situated and the learner must be active. That is, learning provide opportunities for the learner to acquire knowledge and skills in meaningful, often natural contexts, and students should actively participate in class. The biggest advantage of such active learning is that students develop the power to think for themselves, which leads to higher learning performance.

The effects are also shown in machines. The models learned the same data but the final model weights would be different, since they are updated by gradient descent started from different initial weights. Thus, the models' different points of view, which means different model predictions on the same data, and the process of exchanging the predictions using the weights improve the overall performances.

Now, it is important to apply and utilize the knowledge one has rather than just get a lot of knowledge. In this respect, this study is important in that it provides practical implications on how to change the current knowledge-oriented learning methods. Next, experiment 2 tells us that the knowledge difference is also important in the machines (Jo & Cinarel, 2019), showing that the machine experiment can be used to verify the human experiments. Reversely, the results of human experiment can also be adopted to design a better framework.

Acknowledgments

This work was partly supported by the Institute for Institute of Information Communications Technology Planning Evaluation (2015-0-00310-SW.StarLab/25%, 2017-0-01772-VTT/25%, 2018-0-00622-RMI/25%, 2019-0-01371-BabyMind/25%) grant funded by the Korean government.
We also thank Ji Won Yang and Seung-Hee Park for her extremely helpful comments on an earlier draft.

References

Bligh, D. A. (1998). *What's the use of lectures?* Intellect books.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330.

Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom. 1991 ashe-eric higher education reports.* ERIC.

Chang, M.-W., Ratinov, L.-A., Roth, D., & Srikumar, V. (2008). Importance of semantic representation: Dataless classification. In *Aaai* (Vol. 2, pp. 830–835).

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in cognitive science*, *1*(1), 73–105.

Chi, M. T., Roy, M., & Hausmann, R. G. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive science*, *32*(2), 301–341.

Chi, M. T., & Wylie, R. (2014). The icap framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, *49*(4), 219–243.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, *66*(1), 64–74.

Heller, R. S., Beil, C., Dam, K., & Haerum, B. (2010). Student and faculty perceptions of engagement in engineering. *Journal of Engineering Education*, *99*(3), 253–261.

Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Hochreiter, S., & Schmidhuber, J. (1997, 12). Long short-term memory. *Neural computation*, *9*, 1735-80. doi: 10.1162/neco.1997.9.8.1735

Hrepic, Z., Zollman, D. A., & Rebello, N. S. (2007). Comparing students' and experts' understanding of the content of a lecture. *Journal of Science Education and Technology*, *16*(3), 213–224.

Igo, L. B., Bruning, R., & McCrudden, M. T. (2005). Exploring differences in students' copy-and-paste decision making and processing: A mixed-methods study. *Journal of Educational Psychology*, *97*(1), 103.

Jo, H., & Cinarel, C. (2019, November). Delta-training: Simple semi-supervised text classification using pretrained word embeddings. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (emnlp-ijcnlp)* (pp. 3458–3463). Hong Kong, China: Association for Computational Linguistics. Retrieved from `https://www.aclweb.org/anthology/D19-1347` doi: 10.18653/v1/D19-1347

Johnson, D., Johnson, R. T., & Smith, K. (1991). Active learning: Cooperation in thecollege classroom. *Interaction Book Company*.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college what evidence is there that it works? *Change: the magazine of higher learning*, *30*(4), 26–35.

Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1746–1751).

King, A. (1993). From sage on the stage to guide on the side. *College teaching*, *41*(1), 30–35.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lambert, N. M., & McCombs, B. L. (1998). *How students learn: Reforming schools through learner-centered education.* American Psychological Association.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., … others (2015). Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, *6*(2), 167–195.

Lim, J., Jo, H., Zhang, B.-T., & Park, J. (2020). Human-like active learning: Machines simulating the human learning process. *arXiv preprint arXiv:2011.03733*.

Lim, J., Ko, H., Yang, J. W., Kim, S., Lee, S., Chun, M.-S., … Park, J. (2019). Active learning through discussion: Icap framework for education in health professions. *BMC medical education*, *19*(1), 1–8.

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1* (pp. 142–150).

Mayer, R. E., & Alexander, P. A. (2016). *Handbook of research on learning and instruction*. Taylor & Francis.

Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education*, *102*(3), 346–374.

Muldner, K., Lam, R., & Chi, M. T. (2014). Comparing learning from observing and from human tutoring. *Journal of Educational Psychology*, *106*(1), 69.

Poh, M.-Z., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE transactions on Biomedical engineering*, *57*(5), 1243–1252.

Prince, M. (2004). Does active learning work? a review of the research. *Journal of engineering education*, *93*(3), 223–231.

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T.-Y., & Lee, Y.-H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the united states. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *44*(10), 1436–1460.

Settles, B. (2010). *Active learning literature survey. university of wisconsin.* Madison.

Settles, B. (2011). From theories to queries: Active learning in practice. In *Active learning and experimental design workshop in conjunction with aistats 2010* (pp. 1–18).

Wieman, C., & Perkins, K. (2005). Transforming physics education. *Physics today*, *58*(11), 36.

Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. In *Advances in neural information processing systems* (pp. 649–657).