

PAPER

Validity of Kit-Build Method for Assessment of Learner-Build Map by Comparing with Manual Methods

Warunya WUNNASRI^{†a)}, Jaruwat PAILAI[†], Nonmembers, Yusuke HAYASHI[†],
and Tsukasa HIRASHIMA[†], Members

SUMMARY This paper describes an investigation into the validity of an automatic assessment method of the learner-build concept map by comparing it with two well-known manual methods. We have previously proposed the Kit-Build (KB) concept map framework where a learner builds a concept map by using only a provided set of components, known as the set “kit”. In this framework, instant and automatic assessment of a learner-build concept map has been realized. We call this assessment method the “Kit-Build method” (KB method). The framework and assessment method have already been practically used in classrooms in various schools. As an investigation of the validity of this method, we have conducted an experiment as a case study to compare the assessment results of the method with the assessment results of two other manual assessment methods. In this experiment, 22 university students attended as subjects and four as raters. It was found that the scores of the KB method had a very strong correlation with the scores of the other manual methods. The results of this experiment are one of evidence to show the automatic assessment of the Kit-Build concept map can attain almost the same level of validity as well-known manual assessment methods.

key words: concept map assessment method, kit-build concept map, validity

1. Introduction

Concept Map was developed in 1972 in Novak and Musonda’s research program [1]. They investigated changes in children’s knowledge of science based on the learning psychology of Ausubel et al. [2]. In Ausubel et al.’s research, they discussed the assimilation of new knowledge into existing knowledge by learners. A concept map represents conceptual understanding via connections between concepts. A concept in a concept map can be a term or symbol that is enclosed in a box, and a link is a line that is connected to two concepts. A linking word is a word on the link that represents the relationship between concepts. To build the concept map, creators have to organize their knowledge following their target. They can limit the scope of their concept map by constructing a concept map for answering the focus question. Then the creators build a concept list from the main idea of the content and order these concepts from general to more specific aiding in hierarchical construction. Proposition of the concept map, or unit of meaning, can be constructed from linking two or more concepts via a proper relationship. The concepts should be

ordered by placing the general concept in the top hierarchy and specific concepts at the bottom [3]. Moreover, concept maps can help learners to significantly reduce their learning cognitive load because concept maps assist in the integration of knowledge and facilitate learners in their independent learning and thinking [4]. Due to these characteristics, the concept map is used to organize and represent knowledge extensively.

Afterward, the concept map is used for evaluating learners’ understanding. Several concept map evaluations were proposed. Many criteria were applied to evaluate concept maps following each specific objective. Novak and Gowin [5] mainly used structural of concept map for scoring. Crosslink, link is connected between two different segments, receives the highest priority in this method for evaluating learner’s creativity. Relational scoring of McClure and Bell pays the attention to meaning of each proposition in concept map for assessing learner’s understanding [6]. This method is more meaningful scoring than the structural scoring of Novak and Gowin. Rates should consider about the objective of evaluation before choosing the scoring method for evaluating concept map. These assessment methods which were used to evaluate learners’ concept map manually are accepted widely, but they entail high costs, such as time and human workload, for scoring each concept map. Hence, an automatic concept map assessment is proposed for decreasing time cost and human workload.

The Kit-Build concept map (KB map) is a framework to realize automatic concept map assessment [7], [8]. In the KB map framework, a learner builds a concept map by using only a provided set of components, referred to as the set “kit”. Instant and automatic assessment of a learner-build concept map, realized in this framework, is referred to as the “Kit-Build method” (KB method). In this framework, the set of components are made by decomposing a concept map that is built by a responsible teacher. This map is called the “teacher-build map”. The responsible teacher is requested to build the teacher-build map as a criterion to assess a learner’s comprehension for a specific topic or teaching. Then, a learner is requested to build a concept map to express his/her comprehension for the same topic or teaching. Because all components of the learner-build map are the same as the teacher-build map, automatic assessment of a learner-build map is realized by comparing the learner-build map with the teacher-build map. KB map and its automatic assessment method have already been

Manuscript received May 31, 2017.

Manuscript revised November 15, 2017.

Manuscript publicized January 11, 2018.

[†]The authors are with Graduate School of Engineering, Hiroshima University, Higashihiroshima-shi, 739–8527 Japan.

a) E-mail: warunya@lel.hiroshima-u.ac.jp

DOI: 10.1587/transinf.2017EDP7177

practically used in classrooms in various schools, for example, in science learning in elementary schools [9], [10], junior high schools [11], and university [12], [13]. Also, it is proper for use in various educational fields such as English, science, and social science [14]–[18]. And KB map had been investigated about the effect to memory retention in the experimental environment [19].

These practices have shown that the KB map is suitable for use in teaching situation where the instructor gives directions followed by instructor's interpretation. However, we have not previously compared the KB method with other well-known manual methods that are also used practically. Although the automatic assessment method has advantages over manual assessment, for example, real time assessment/feedback, load reduction of the rater/teacher, etc., the quality of automatic assessment requires investigation. Normally, although the manual concept map assessment method is used by one teacher on their own way individually, McClure et al. tried to prove quality of several concept map assessment methods [20] from viewpoints of reliability and validity. In McClure's research, validity is evaluated by correlation among assessment results of the same concept map and reliability is evaluated by g -coefficient among several raters. Their reliability and validity investigation procedures are explained clearly so we designed the experiment following their procedures. In our study, validity of KB method was evaluated by correlation with the manual methods reported to be of high quality in McClure's paper. Although reliability of KB method is perfect because it is systematic and automatic method, we used this reliability to evaluate the results of quality of the manual methods conducted in this study. As the two manual methods, (1) structural scoring proposed by Novak [5] and (2) relational scoring proposed by McClure [6] were adopted. We conducted an experiment as a case of study where 22 university students were designated as subjects, and four were designated as raters. The results of the experiment showed that the scores of the KB method had a statistically significant correlation with the scores of the other manual methods [21]. The results suggest that automatic assessment using the KB method can attain almost the same level of validity as well-known manual methods. Although the results of this experiment can be used as an evidence to represent of the validity of KB method, this is a case study. It is necessary to examine more several cases in order to confirm the generality of the validity.

2. Overview of the Concept Map Assessment Method

2.1 Manual Concept Map Assessment Methods

A manual concept map assessment method is used by a human who can understand the meaning of words in the concept map well. The human is often called a "rater". In this study, we focus on the methods that pay attention to the structure of a concept map and the meaning of the proposition of a concept map.

Several concept map assessment methods evaluate the concept map by investigating the structure of the map, such as, the levels of the hierarchy, the characteristics of the branch, etc. In this study, we focus on the structural scoring of Novak and Gowin [5] as a typical structural method. This method gives high scores for each correct level of the hierarchy and each valid crosslink because ordering the concepts into the hierarchy, and connecting the crosslinks, can facilitate the constructor's creative thinking. However, structural scoring, which tends to score the structure more than the meaning, may be the cause of substantial meaning-leakage in a concept map.

Many manual assessment methods which pay more attention to the meaning of a proposition for scoring the concept map, rather than the structure, have been proposed. They focus on language and understanding of the representation. These meaningful methods always have a printed set of criteria as the rubric for assessing knowledge and giving feedback. From investigating various meaning methods, we focused on the relational scoring from McClure and Bell [6], which is referred to as relational scoring in this paper, and is a common concept map assessment method. This method scores the concept map by checking the possible relationship between each proposition, suitability of label between concepts of proposition and compatibility between label and the direction of arrow or hierarchical.

There are many researches which invested on the reliability and validity of concept map assessment. Ruiz-Primo et al. [22], [23] focused on reliability and validity of scores from two concept mapping method which are high directed "fill-in-a-skeleton-map" and low directed "construct-a-map-from-scratch". Fifty-five learner-build maps were scored by three raters and the scores across raters were used to examine the generalizability coefficient (g -coefficient). The correlation between two concept mapping methods and multiple-choice test was used to show the validity. In this study, they concluded both mapping techniques are tapping somewhat similar but not identical aspects of learners' connected understanding. The correlation between score from the multiple-choice test and both concept map techniques confirmed that the mapping techniques were not equivalent. The pattern of correlation coefficient was different across mapping techniques. McClure et al. [20] also investigated on the reliability and validity of concept map assessments, they requested 63 students to construct concept maps by using 20 provided concepts, creating their own linking words. Then, 12 raters scored individual maps by assessing each proposition on the concept map separately. The raters awarded scores of zero to three points for each proposition based on the suitability of the meaning of the proposition. The authors claimed that this relational method has the highest reliability when using the criteria map, (teacher-build map), using the holistic method and the structural method as comparisons (Novak and Gowin structural scoring). The authors confirmed this result by using the g -coefficient value as an estimate of the score reliability of scores assuming a single rater. For the analysis of validity, they investigated from

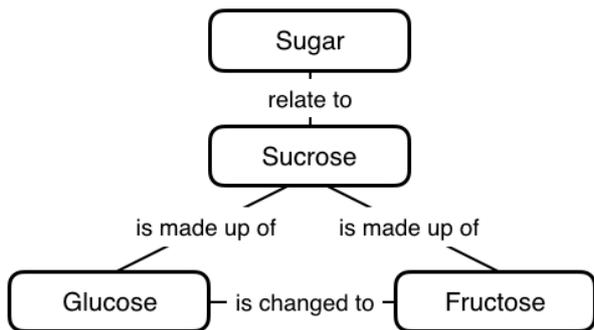


Fig. 1 Teacher-build map.

the correlation of map scores with a measure of similarity between each map and the criteria map. Then, they conclude the relational scoring method with criteria map correlated most closely with maps’ measure similarity. Based on these considerations, we have designed an experiment for testing the reliability of a manual method, similar to the experiment of McClure et al. We selected the structural scoring proposed by Novak and Gowin, and the relational scoring proposed by McClure and Bell, to compare with the KB map proposed in the current study because they are the typical manual methods which had been already proved the reliability and validity of assessment. Then, we investigated validity of KB method from the correlation of KB map scores with scores from the manual methods, similar to the McClure et al.’s experiment.

2.2 Kit-Build Concept Map and Automatic Assessment

The Kit-Build concept map framework is one of the automatic concept map assessment methods that use a teacher-build map to compare with the learner-build map by using exact matching at the propositional level. It is utilized in the form of a learning task or exercise for checking learners’ comprehension of a topic that they have already learned. The task of the KB map is separated into two subtasks. The first is the segmentation task where a teacher is requested to prepare the teacher-build map, which is an expression of an eligible comprehension of the topic for the teacher. An example of the teacher-build map is illustrated in Fig. 1. After submitting the teacher-build map to the server, the teacher-build map is extracted to be the kit that contains a list of concepts and relationships from the teacher-build map. The kit from the teacher-build map in Fig. 1 is shown in Fig. 2. Moreover, this kit is provided to help learners to reduce their cognitive load more than the traditional concept map, where they must create all components themselves. Using the kit, the learners are not requested to create any component. They only have to recognize the provided components and connect them.

The second task is called the structuring task. Learners are given the learning task of reconstructing a concept map by using the kit, creating a map which is referred to as the learner-build map (Fig. 3). After the learner-build maps are

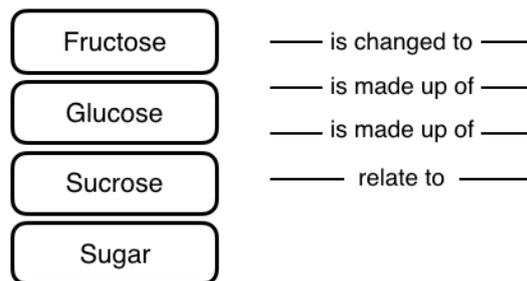


Fig. 2 Kit.

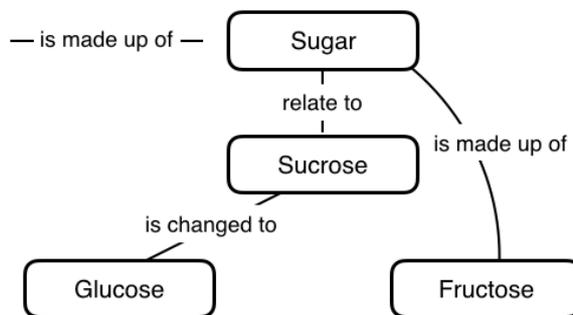


Fig. 3 Learner-build map.

uploaded to the server, the KB map will evaluate learner-build maps by exactly matching each learner’s proposition with the teacher-build map’s proposition. For example, the relationship between the concepts “Sugar” and “Sucrose” is checked. If the relationship is identified as “related to,” the score for this learner-build map will increase by one point. In the case of the concepts “Sucrose” and “Glucose,” if the learner connected them by using the relationship “is changed to,” this does not exist in the teacher-build map. Following the teacher-build map, the relationship of this proposition should be “is made up of,” so this proposition is not awarded any point from the system. This corresponds to the scoring by propositional level exact matching method. This method makes the KB map different from the manual methods which allow learners to create their own linking words, preventing the learner-build map from being straightforwardly compared with the criteria map. The manual methods require time for considering the meaning of each proposition carefully. After checking the connections of the learner-build maps by the propositional level exact matching, the system will generate a score in a percentage format which is calculated via the number of correct links of learner-build map divided by the number of links of teacher-build map. For the example, the learner-build map in Fig. 3 will be given 25 percent score from one correct link “relate to” divided by four links from the teacher-build map. The instructor can also investigate learners’ misunderstanding individually as a difference map and can find the overview of all learners by overlaying all learner-build maps, as the group map, and the group-goal difference map on the analysis screen of the KB. In the difference map, three types of error link are represented as shown in Fig. 4. The lacking

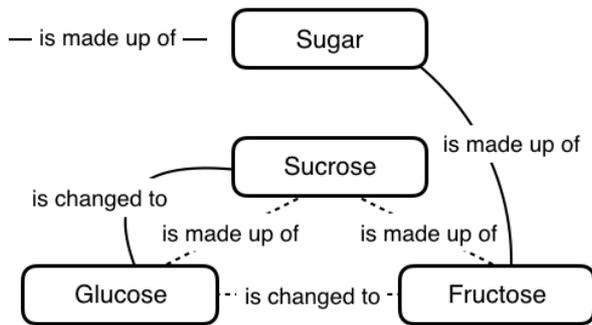


Fig. 4 Difference map.

link, which is represented by a dashed line, is a link that exists in the teacher-build map but does not exist in learner-build map. The excessive link, which is shown as a solid line, is a link that occurs in learner-build map but does not occur in the teacher-build map. Lastly, a solid line that is not connected to any concepts in the learner-build map is the leaving link. The instructor can use these links to find the holistic leaking understanding of all learners. Following the KB map framework's ability, the instructor can use the KB map to check understanding of individuals or groups of learners, and can use the diagnosis result to discuss with learners the meaning of each of the error links. After error link analysis, the instructor can adjust the teacher-build map or teach learners about the content that learners have not understood completely. We have confirmed that teachers can use the feedback from KB map in their class effectively [24].

In the other automatic methods [25]–[28], they allow learners to create their own linking words so they cannot compare a learner-build map with the criteria map straightforwardly. Hence, they require synonym word matching, which is very flexible for evaluation using the meaning of words, but which has not yet reached a sufficient level of accuracy. In contrast, the KB map provides the kit which can be assessed by using the propositional level exact matching and can create informative diagnosis results. Moreover, the KB map can provide the group map and group-goal difference map, which can support the instructor in analyzing comprehension in both an individual learner and as an overview of the whole class. These are the prominent advantages of the KB map when it is utilized in a classroom situation.

3. Research Methodology

In this research, we assume that assessment results of well-known manual methods have reliability and validity. Then, the validity of KB map is investigated by comparing with the results of manual methods. For this investigation, we designed an experimental procedure to compare the KB map and the manual methods in terms of their ability to assess the comprehension of learner on a topic. Usually, the KB map is used in teaching situations, however, it is desirable to ensure that the KB map as can be used in a reading situation also. Hence, the experiment was designed to operate in two

Table 1 Comparison between the attributes of each concept map assessment method.

Assessment Method	Assessment			Provided Items	
	Raters	Level of Analysis	Matching Method	Concepts Words	Linking Words
Structural Scoring	Manual	Structural	Synonym	Provided	Not Provided
Relational Scoring	Manual	Propositional	Synonym	Provided	Not Provided
Kit-Build Concept Map	Automate	Propositional	Exact	Provided	Provided

learning situations. Moreover, to compare the difference between the KB map and the manual method, the important attributes of the concept map assessment method are shown in Table 1.

Two typical scoring methods, which are widely used for assessing concept maps, namely the structural scoring as structural level analysis, and the relational scoring as propositional level analysis, were chosen for comparison. The manual method is inferred from the research of McClure et al. [20], who provided a list of concepts to learners and requested that they construct concept maps by creating linking words themselves. The synonym matching method was used for evaluating the meaning of each proposition. However, the KB map provides both the concepts and the linking words, which are decomposed from the teacher-build map, to learners. Thus, the automatic exact matching method can be used for checking the correctness of each proposition.

3.1 Subjects

Subjects for this study were recruited from university students who possessed a good level of English. The 22 students, who were volunteers from various education fields, were given the role of learners. They were given introductory training in concept maps before participating in the experiment. Four students, who were familiar with the use of the concept map and understood the content of the experiment material well, were assigned as raters. These raters were given an explanation of the procedure of each assessment method, and they were required to study the procedures carefully before scoring the learner-build map. In addition, one graduate student was assigned the role of instructor. The instructor was required to prepare the article and teaching material for the experiment and the instructor was also required to construct the teacher-build map following specific instructions. In this study, the article "Sugar", which uses common explanatory words, was chosen for the learning process so subjects who are from various faculties can understand without bias. This article contained three sections, covering one third of a page, defined as the introduction to sugar, types of sugar and how sugar is produced [29].

3.2 Map Production

Initially, the instructor chose a 1,594 word article, prepared the teaching materials and built the teacher-build map. The

teacher-build map, which was also used as the criteria map for manual methods contained 15 concepts and 16 relationships. In the study, the procedure of this experiment is displayed in Fig. 7. Firstly, learners were requested to read the article in ten minutes (1), and they were then provided with the list of concepts. Next, they were required to create

linking words by themselves for the construction of a concept map in 15 minutes using the Cmap Cloud application (2), as illustrated in Fig. 5 [30]. These learner-build maps were scored by the two manual methods. The learners were then asked to construct a concept map again in 15 minutes by integrating the kit of the KB map (3), which provided both a list of concepts and a list of linking words. The initial representation of the KB map in this experiment is shown in Fig. 6. After the learners had completely connected the propositions and uploaded their map to the server, these learner-build maps were evaluated using the KB map assessment method based on exact matching at the propositional level.

After the reading session concluded, the instructor taught learners based on the same reading article but following the instructor’s interpretation using 16 slides delivered over ten minutes (4). Afterward, learners were required to construct the learner-build maps following the same procedure as in the reading situation, namely, constructing learner-build maps by creating linking words by themselves (5) and integrating the kit to create a learner-build map using the KB map (6). For the reason why we did not request learner to modify their previous concept map from reading situation but request for creating the new one, we aim to evaluate the concept map that learners constructed from their actual understanding in each learning situation. When learners completed all learner-build map construction, they were asked to answer a questionnaire (7).

For this experiment procedure, we designed for comparing a reflection of learners’ understanding between concept map construction, which learners can create linking words freely, and KB map, which learners were provided both of the concept list and linking word list. If the provided components do not support learner too much and they do not disturb learners to express understanding, the results of KB method and the manual methods should be in the

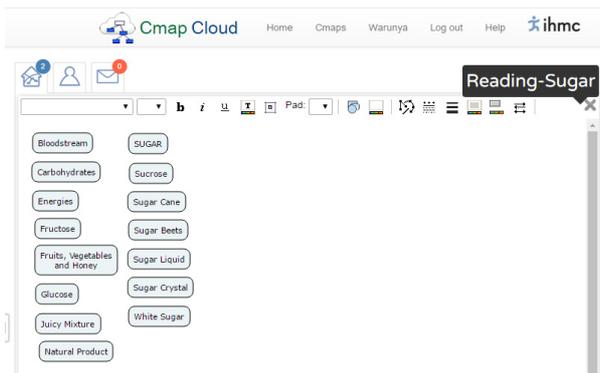


Fig. 5 Cmap Cloud screen.

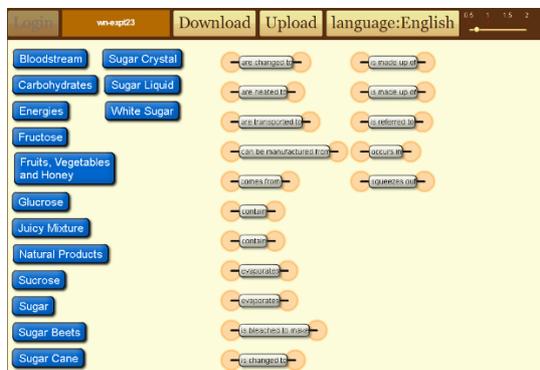


Fig. 6 Kit-Build concept map screen.

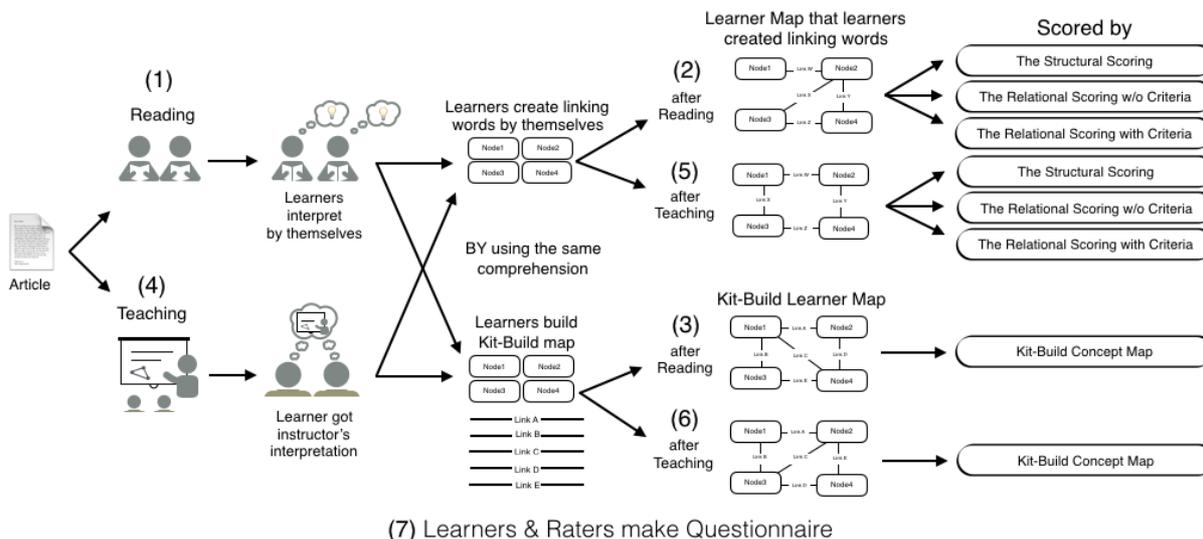


Fig. 7 The experiment procedure.

same way. So all subjects were requested to participate in all situations, which are usual concept map, KB map, reading situation, and teaching situation. For the order effects, we designed reading situation was produced before teaching because subjects should think by themselves before receiving the interpretation of teacher. In the same way, we requested subjects to create their linking words freely before they received the components from KB map because subjects should feel free for creating the linking words. They should not be guided by the provided components of KB map.

3.3 Concept Map Scoring by Manual Methods

The concept maps, which were constructed using Cmap Cloud, were scored by three manual methods that contained, (a) the Novak and Gowin structural scoring (the structural scoring), (b) the McClure and Bell relational scoring without the criteria map (the relational scoring without criteria map) and (c) the McClure and Bell relational scoring with criteria map (the relational scoring with criteria map). The raters were required to read the instructions of each assessment method carefully without time restrictions. The score of the manual methods was normalized to a percentage score by using the perfect score for each method. After the scoring was completed, the raters were requested to complete the questionnaire. Procedures for each method were prepared based on the description in [20]. The reliability of the results of the manual methods is discussed in Sect. 4.

3.4 Questionnaires

The questionnaires were assigned to both raters and learners to assess their familiarity with concept mapping and their opinion of the experiment. For the learners' questionnaires, the aim was to assess their background in concept mapping and in the content of the article. A further aim was to understand how their experiences differed when constructing the concept maps by creating their own linking words and when using the KB map.

For the raters, the questionnaire contained two parts. The first part of the questionnaire assessed their familiarity with the concept map and with the content of the article. The questionnaire also asked about their disposition when they were scoring the concept maps. The second part of the questionnaire is constructed based on conclusion of the important characteristics of each concept map scoring method in McClure et al.'s study [20]. We requested raters to rank each scoring method in four aspects covering (i) hardness of decision, (ii) use of memory, (iii) time taken and (iv) reasonableness of the score.

4. Experimental Results and Discussion

4.1 Correspondence of the KB Map and the Manual Method

To confirm the KB map's validity as a framework for

Table 2 Score of learner-build map from each scoring method.

	Reading		Teaching	
	Average	S.D.	Average	S.D.
Structural Scoring	19.91	11.15	49.26	18.34
Relational Scoring w/o Criteria	29.67	15.38	61.93	14.99
Relational Scoring w/ Criteria	38.81	14.94	67.64	12.64
Kit-Build Concept Map	23.50	23.19	55.36	22.94

Table 3 The g-coefficient for each manual method and the study of McClure et al. [16].

	Current Study		McClure's
	Reading	Teaching	
Structural Scoring	0.7520	0.9029	0.23
Relational Scoring w/o Criteria	0.8659	0.8540	0.51
Relational Scoring w/ Criteria	0.8874	0.9133	0.76

assessing learners' comprehension of a topic by comparing with reliable manual methods, we aim to first investigate the reliability of the manual methods. The scores from three manual methods: (a) the structural scoring, (b) the relational scoring without criteria map and (c) the relational scoring with criteria map and Kit-Build map are represented in Table 2. A difference of each scoring method effects S.D. value because the score of each link from the manual methods can be multiple scales (from zero to three points) while KB method scores one correct link as one point. The scores from manual methods were used to evaluate the reliability of each manual method by calculating g-coefficient value. After that, the score of Kit-Build map will be ensured the correlation with the score of reliability manual methods for confirming that Kit-Build method can attain almost the same level of validity as manual assessment methods.

The scores of learner-build maps from each manual method were used to perform generalizability analysis through the GNOVA software [31] which returns the g-coefficient, as used in the reliability investigation by McClure et al. [20]. The g-coefficient is analogous to the reliability coefficient in classical test theory [32]. In this study, we interpret the g-coefficient as an estimate of score reliability assuming a single rater which shows the consistency of each scoring method as shown in Table 3. All values of g-coefficient of the current study are higher than values reported in McClure et al. [20]. Then, the relational scoring with the criteria map resulted in the highest score reliability in both reading and teaching situations, which is consistent with the investigation of McClure et al. which indicated that the relational scoring method is reliable in assessing the concept map. Based on these results, we concluded that the manual assessment conducted in this research is reliable and it is possible to evaluate validity of KB map by comparing with the results of the manual assessment. As for the reason why the g-coefficient obtained in the current study is higher than that obtained by McClure et al. we guess that the current study was conducted with a smaller number of subjects and raters, that is, 12 raters in McClure et al., and four raters

in the current study.

A comparison between the KB map's result and the reliable manual method's result is required for analyzing the validity, which is an overall evaluative judgment, founded on empirical and theoretical rationales, of adequacy and appropriateness of inferences and actions based on test score [33]. In McClure's research, they investigated the validity by the correlation between the concept map scores from each manual method and the similarity measure of each learner-build map with the criteria map. The similarity measure of each learner-build map and criteria map was held as the reasonable scoring method. To investigate the validity of KB method, the Pearson's correlation was computed using the R programming language and the correlation value is shown in Table 4. Following the strength of the correlation from Evans [34], the relational method with criteria map, which achieved the highest reliability score, has a very strong correlation with the KB map in both reading and teaching situations. This is because raters use the criteria map as a frame for their scoring, in a similar way to the teacher-build map used in the KB map. For the remaining methods, the results from the relational scoring without criteria map have a very strong correlation in the reading situation and strong correlation in the teaching situation. This is because the procedure of relational scoring without the criteria map is too wide for meaningful evaluation of the learner-build maps, which are constructed for checking the understanding following a specific teaching situation. The structural scoring has a strong correlation with the KB map in both situations, even though structural scoring scores the concept map by giving precedence to the structure of the concept map, which is a different approach compared to the KB map.

The results above suggest that the KB map can assess learners' comprehension of a topic as well as the manual concept map assessment methods. If the manual methods give a relatively high score to a learner, the KB map also has a high possibility of giving a relatively high score to the learner. In addition, learners who get a relatively low score

from the manual methods, also have a high possibility of getting a relatively low score from the KB map. As indicated by the high correlation value, the KB map has validity, and is comparable to the manual methods, in identifying learners' comprehension for a topic and evaluating the concept map reasonably.

4.2 Results of Questionnaire

Two sets of questionnaires, one was for learners and the other was for raters, were used in this study. The questionnaires for learners were answered just after they completed all of their tasks. The results are presented in Table 5. From the learners' questionnaire analysis, learners who did not have existing knowledge about the learning material before obtained a good understanding of the content after reading. In addition, the learners could accept the instructor's interpretation clearly after they received an explanation in the teaching situation. When learners constructed their learner-build map by creating their own linking words, most of them concluded that they could represent their understanding adequately; similarly, users of the kit KB map were able to express their understanding appropriately. This summary suggests that the KB map is appropriate to use in supporting learners to express their understanding, and that it produces similar results to using the concept map where the linking words are created freely.

The raters' questionnaire included two types of questions. One type is general questions as shown in Table 5 and the other type is questions to compare the three manual methods as summarized in Fig. 8. As shown in Table 6, all raters identified their familiarity with using the concept map and their understanding of the learning material as strong confident.

The results of comparison of the three manual methods by raters are illustrated in Fig. 8. The values are calculated by averaging the given score from raters in each manual method. Raters were requested to give score from one to three points by ranking each scoring method following four criterions: hardness of division, use of memory, time taken, and reasonableness of score. In Fig. 8, the structural scoring was the hardest assessment method, because the rater had to decide on the suitability of each hierarchy and crosslink. McClure et al. [20] concluded that it made high complex cognitive when the raters tried to compare the quality of many maps. Conversely, it was easiest to use the relational scoring with criteria map since the criteria map could be used as a guide for scoring. For the cost of scoring,

Table 4 The correlations in scores between each manual method and the KB method.

	KB in Reading	KB in Teaching
Structural Scoring	0.7360	0.7360
Relational Scoring w/o Criteria	0.8532	0.7371
Relational Scoring w/ Criteria	0.8671	0.8165

Note: Calculated Pearson product correlations are statistically significant as indicated by p-value < 0.01

Table 5 A part of the learners' questionnaire.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Learners know about concept map before	9%	14%	9%	55%	14%
Learners know about material before	18%	27%	9%	41%	5%
Learners can represent their understanding by using Cmap Cloud	0%	5%	18%	73%	5%
Learners can represent their understanding by using KB map	0%	5%	0%	36%	59%

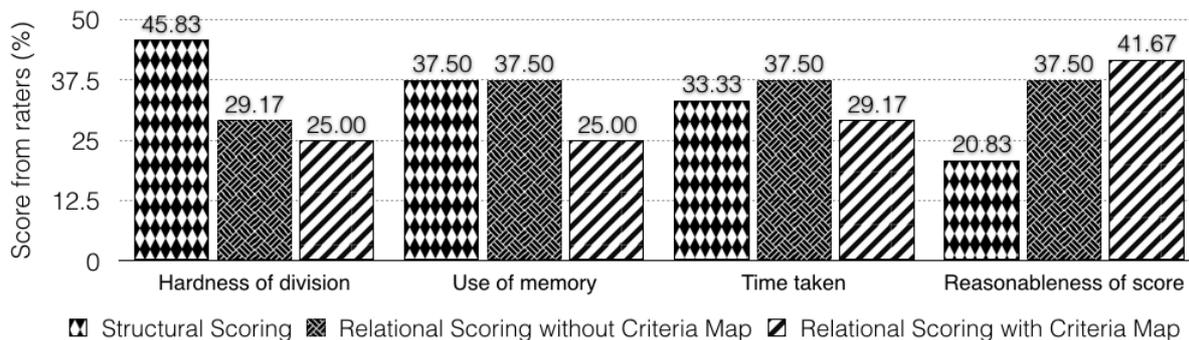


Fig. 8 A part of the score from the raters' questionnaire.

Table 6 A part of the raters' questionnaire.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I am familiar to use concept mapping	0%	0%	0%	75%	25%
I understand the article of the experiment	0%	0%	0%	25%	75%
For the Novak & Gowin structural scoring,					
- It is hard to evaluate the hierarchy of concept map	0%	25%	0%	50%	25%
- It is hard to evaluate the cross-link	25%	25%	25%	0%	25%
For the McClure & Bell relational scoring without master map					
- It is hard to score meaning of each proposition when I do not have a guide	0%	75%	25%	0%	0%
For the McClure & Bell relational scoring with master map					
- It is hard to score meaning of each proposition when I have a guide	25%	50%	0%	25%	0%
- I hesitated to score when the master map is provided	25%	50%	0%	25%	0%

the raters noted that the structural scoring and the relational scoring without criteria map used their memory load and time more than the relational scoring with criteria map. This was because of the difficulty in thinking about the learner-build map structure and recalling how previous learner-build maps were scored. For this challenge, the criteria map can help the scoring of the learner-build map by using the relational scoring with the criteria map. In the final question, the raters were requested them to rank the most reasonable method in their opinion. The relational scoring with criteria map achieved the highest rating. This ranking corresponds with the comparison between six concept map assessments by McClure et al. [20]. Hence, the strong correspondence between the KB map and the relational scoring with criteria map confirms that the propositional level with exact matching of the KB map has validity for assessing the efficiency of learning and the KB map can be used as an alternative automatic method for assessing the concept map.

5. Conclusion and Future Work

This study investigates the validity of the KB map assessment in terms of its ability to identify the efficiency of learning reasonably when it is compared with the well-known reliable manual methods. The objective of the experiment wants to show that the provided components of KB map do not affect learners when they tried to express understanding and the KB method can identify level of learner's understanding via concept map almost same level with the manual methods. An experiment was designed as a case

study to compare the KB map assessment with three manual concept map assessment methods in reading and teaching situations. Selected manual methods contained structural scoring, (which investigates the composition of the concept map straightforwardly), relational scoring without the criteria map, and relational scoring with the criteria map. The relational scoring gives precedence to the meaning of propositions and is reasonable for evaluating understanding from a concept map, but requires expert checking and significant time input for scoring. These manual methods provide flexible and meaningful concept map assessment, and their reliability is widely accepted. However, they are inconvenient due to the limited class time that instructors have to complete a unit of instruction. In this study, the KB map was compared with the manual methods to test the assumption that the KB map has validity for identifying the efficiency of learning. From this study, the results show a strong and significant correlation between the KB map and the manual methods in both the teaching and reading situations. The KB map has the highest correlation with the relational scoring with criteria map, achieving the most reliability score (g-coefficient) in both learning situations. Moreover, the learner-build map scores of the KB map were similar to the manual methods. Based on these results, it is one of evidence, which can suggest that the validity of the KB map assessment is comparable to the manual assessment methods.

For the future work, the number of subjects will be considered to expand. A larger number of subjects can strongly confirm the conclusion of this study. We also intend to

investigate the method for confirming the general validity of KB method and the effect of order when learners make the traditional concept map before using KB map. For the conclusion as KB map is appropriate to use in supporting learners to express their understanding, we will focus on the understanding sharing. The aim will be to use the abilities of the KB map, connected with collaborative learning techniques, to support the understanding sharing. Learners should build shared understanding with their collaborators in the form of agreement and disagreement.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Number 26280127 and 15H02931.

References

- [1] J.D. Novak and D. Musonda, "A twelve-year longitudinal study of science concept learning," *American Educational Research Journal*, vol.28, no.1, pp.117–153, 1991.
- [2] D.P. Ausubel, J.D. Novak, and H. Hanesian, *Educational psychology: a cognitive view*, Rinehart and Winston, New York, 1978.
- [3] J.D. Novak and A.J. Cañas, Technical report IHMC CmapTools, Institute for Human and Machine Cognition, Florida, 2008.
- [4] M.L.M. Hu and M.H. Wu, "The effect of concept mapping on students' cognitive load," *World Transactions on Engineering and Technology Education*, vol.10, no.2, pp.134–137, 2012.
- [5] J.D. Novak and D.B. Gowin, *Learning how to learn*, Cambridge University Press, New York, 1984.
- [6] J.R. McClure and P.E. Bell, *Effects of an environmental education related STS approach instruction on cognitive structures of pre-service science teachers*, State University, Pennsylvania, 1990.
- [7] T. Hirashima, K. Yamasaki, H. Fukuda, and H. Funaoi, "Kit-Build concept map for automatic diagnosis," *Proc. 15th of Artificial Intelligence in Education 2011, Auckland, New Zealand*, pp.466–468, 2011.
- [8] T. Hirashima, K. Yamasaki, H. Fukuda, and H. Funaoi, "Framework of Kit-Build concept map for automatic diagnosis and its preliminary use," *Research and Practice in Technology Enhanced Learning*, vol.10, no.1, pp.1–21, 2015.
- [9] K. Sugihara, T. Osada, S. Nakata, H. Funaoi, and T. Hirashima, "Experimental evaluation of Kit-Build concept map for science classes in an elementary school," *Proc. Int. Conf. on Computers in Education 2012, Singapore*, pp.17–24, 2012.
- [10] K. Yoshida, K. Sugihara, Y. Nino, M. Shida, and T. Hirashima, "Practical use of Kit-Build concept map system for formative assessment of learners' comprehension in a lecture," *Proc. Int. Conf. on Computers in Education 2013, Bali, Indonesia*, pp.906–915, 2013.
- [11] T. Nomura, Y. Hayashi, Y., T. Suzuki, T. and T. Hirashima, "Knowledge propagation in practical use of Kit-Build concept map system in classroom group work for knowledge sharing," *Proc. Int. Conf. on Computers in Education Workshop 2014, Nara, Japan*, pp.463–472, 2014.
- [12] Y. Hayashi and T. Hirashima, "Kit-Build Concept Mapping for Being Aware of the Gap of Exchanged Information in Collaborative Reading of the Literature," *Proc. Int. Conf. on Human Interface and the Management of Information, Greece*, pp.31–41, 2014.
- [13] Y. Hayashi and T. Hirashima, "Analysis of the Relationship Between Metacognitive Ability and Learning Activity with Kit-Build Concept Map," *Proc. Int. Conf. on Human Interface and the Management of Information, United States*, pp.304–312, 2015.
- [14] M. Alkhateeb, Y. Hayashi, T. Rajab, and T. Hirashima, "Comparison between Kit-Build and Scratch-Build concept mapping methods in supporting EFL reading comprehension," *The Journal of Information and Systems in Education*, vol.14, no.1, pp.13–27, 2015.
- [15] M. Alkhateeb, Y. Hayashi, and T. Hirashima, "The effects of using Kit-Build method to Support Reading Comprehension of EFL," *Proc. Int. Conf. on Human Interface and the Management of Information, Greece*, pp.3–11, 2014.
- [16] Y. Hayashi, M. Ui, M. Kayashima, and T. Hirashima, "An Experimental Use of Kit-Build Concept Map with Approval Link in a Lecture of the Humanities," *Japan Journal of Educational Technology*, vol.38, pp.149–152, 2016.
- [17] T. Kitamura, A. Yamanaka, K. Maeda, Y. Hayashi, and T. Hirashima, "Learning Task Generation from a Series of Propositions of a Learning Topic: Kit-Building Task of Concept Map and Multiple Choice Task of Fill-in-the-blank Questions," *Proc. 24th Int. Conf. on Computers in Education, India*, pp.91–93, 2016.
- [18] M. Alkhateeb, Y. Hayashi, T. Rajab, and T. Hirashima, "Experimental Use of Kit-Build Concept Map System to Support Reading Comprehension of EFL in Comparing with Selective Underlining Strategy," *International Journal of Advanced Computer Science and Applications*, vol.7, no.4, pp.80–87, 2016.
- [19] H. Funoi, K. Ishida, and T. Hirashima, "Comparison of kit-build and scratch-build concept mapping methods on memory retention," *Proc. 19th Int. Conf. on Computers in Education, Thailand*, pp.539–546, 2011.
- [20] J.R. McClure, B. Sonak, and H.K. Suen, "Concept map assessment of classroom learning: reliability, validity, and logistical practicality," *Journal of Research in Science Teaching*, vol.36, no.4, pp.475–492, 1999.
- [21] W. Wunnasri, J. Pailai, Y. Hayashi, and T. Hirashima, "Reliability investigation of automatic assessment of learner-build concept map with Kit-Build method by comparing with manual methods," *Proc. 18th Int. Conf. on Artificial Intelligence in Education, Hubei, China*, pp.418–429, 2017.
- [22] M.A. Ruiz-Primo, S.E. Schultz, M. Li, and R.J. Shavelson, "Comparison of the reliability and validity of scores from two concept-mapping techniques," *Journal of Research in Science Teaching*, vol.38, no.2, pp.260–278, 2011.
- [23] M.A. Ruiz-Primo, "Examining concept maps as an assessment tool," *Proc. 1st Int. Conf. on Concept Mapping, Pamplona, Spain*, pp.555–563, 2004.
- [24] J. Pailai, W. Wunnasri, K. Yoshida, Y. Hayashi, and T. Hirashima, "The practical use of Kit-Build concept map on formative assessment," *Research and Practice in Technology Enhanced Learning*, vol.20, no.12, pp.1–23, 2017.
- [25] B.E. Cline, C.C. Brewster, and R.D. Fell, "A rule-based system for automatically evaluating student concept maps," *Expert systems with applications*, vol.37, no.3, pp.2282–2291, 2010.
- [26] D.B. Luckie, S.H. Harrison, and D. Ebert-May, "Introduction to C-Tools: concept mapping tools for online learning," *Proc. 1st Int. Conf. on Concept Mapping, Pamplona, Spain*, pp.261–264, 2004.
- [27] S.H. Harrison, J.L. Wallace, D. Ebert-May, and D.B. Luckie, "C-TOOLS automated grading for online concept maps works well with a little help from WordNet in concept maps: theory, methodology, technology," *Proc. 1st Int. Conf. on Concept Mapping, Pamplona, Spain*, pp.211–214, 2004.
- [28] H. Kornilakis, M. Grigoriadou, K.A. Papanikolaou, and E. Gouli, "Using WordNet to support interactive concept map construction," *Proc. 4th IEEE Int. Conf. on Advanced Learning Technologies, Joensuu, Finland*, pp.600–604, 2004.
- [29] R. Klaus, "Sugar," <http://www.english-online.at/biology/sugar/sugar-carbohydrate-that-gives-us-energy.htm>, accessed March 27, 2017.
- [30] The Institute for Human & Machine Cognition, "Cmap Cloud & CmapTools in the cloud," <http://cmap.ihmc.us/cmap-cloud>, accessed March 11, 2017.
- [31] R.L. Brennan, *Elements of generalizability theory*, American College Testing Program, Arizona, 1983.

- [32] N.M. Webb and R.J. Shavelson, Generalizability theory: overview, Encyclopedia of Statistics in Behavioral Science, Wiley, Chichester, 2005.
- [33] S. Messick, "The once and future issues of validity: Assessing the meaning and consequences of measurement," ETS Research Report Series, vol.1986, no.2, 1986.
- [34] J.D. Evans, Straightforward statistics for the behavioral sciences, Brooks/Cole, California, 1996.



Warunya Wunnasri received the B.S. degree in Computer Science from Thammasat University and M.S. degree in Engineering from Sirindhorn International Institute of Technology, Thammasat University in 2011 and 2013, respectively. She is currently a Ph.D. student in the Graduate School of Engineering, Hiroshima University.



Jaruwat Pailai received the B.S. and M.S. degrees in Computer Science from Thammasat University in 2011 and 2013, respectively. He is currently a Ph.D. student in the Graduate School of Engineering, Hiroshima University.



Yusuke Hayashi is an associate professor of Department of Information Engineering, Graduate School of Engineering, Hiroshima University since 2012. He received Ph.D. from Graduate School of Engineering Science, Osaka University, Japan, in 2003. He was research associate of the school of Knowledge Science, Japan Advanced Institute of Science and Technology (JAIST) from 2003 to 2005, an assistant professor of the Department of Knowledge Systems, the Institute of Scientific and Industrial Research (ISIR), Osaka University from 2005 to 2010 and an associate professor of the Information Technology Center, Nagoya University from 2010 to 2012. He has been engaged in research on Knowledge modeling, Ontological engineering, Learning engineering.



Tsukasa Hirashima received his B.E., M.E. and Ph.D. from Osaka University in 1986, 1988, and 1991 respectively. He worked at The Institute of Scientific and Industrial Research, Osaka University as a research associate and lecturer from 1991 to 1997. During 1997–2003, he worked in Graduate School of Information Engineering at Kyushu Institute of Technology as an associate professor. He has been a professor of Graduate School, Department of Information Engineering, Hiroshima University since 2004.

Learning Engineering is his major research field. He has received international awards as the Outstanding Paper Award of EDMEDIA95, the Best Paper Award of ICCE2001 & 2002, Honorable Mention Award of AIED2009, and APSCE Distinguished Researcher Award in 2009.