Assessment of knowledge and confidence for E-learning

Devkan Kaleci*, Department of Computer Education and Instructional Technology, Inonu University, 44000 Malatya, Turkey

Ergun Akleman, Department of Visualization, College of Architecture, Texas A&M University, College Station, TX 77843

Abstract

One of the most important goals in E-learning is to guarantee that participants reach the learning objectives. We have observed that having the knowledge of the subject is not sufficient for reaching learning objectives. The participants must also develop understanding that they know the subject, which we have named confidence. In this work, we have demonstrated that it is possible to assess both knowledge and confidence using only two different types of multiple-choice test questions. We have developed 1) a method to design questions to identify both knowledge and confidence and 2) a method to estimate actual knowledge and confidence from answers. We have evaluated our method using Monte-Carlo simulations. Our simulations demonstrated that it is always possible to obtain reliable estimations for knowledge and confidence using approximately 100 multiple choice test questions in a given subject.

Keywords: E-learning, assessment, shape modeling, immersions, 3D-thickenings, shape algebra.

* ADDRESS FOR CORRESPONDENCE: **Devkan Kaleci**, Department of Computer Education and Instructional Technology, Inonu University, 44000 Malatya, Turkey. *E-mail address*: devkankaleci@gmail.com / Tel.: +90-422-377-4456

1. Introduction and motivation

1.1. Introduction

Reliable assessment of knowledge and confidence is very important to evaluate learning objectives in teaching and learning (Ainsworth et al., 2007). Education researchers observed that it is especially crucial to improve students' confidence of their knowledge (Peyre, Peyre, Sullivan, & Towfigh, 2006; Stiggins, 1999). Therefore, education researchers have developed several methods to assess confidence in a classroom environment (Chappuis & Stiggins, 2002; Stiggins, 1992; Stiggins & Chappuis, 2005).

1.2. Motivation

An E-learning environment is different than a classroom environment (Rosenberg, 2001). In an E-learning environment, such as educational games and distance education, we cannot have one-to-one and personal assessment of knowledge and confidence (Zhang, Zhao, Zhou, & Nunamaker, 2004). Therefore, there is a need for automatic assessment of knowledge and confidence for such E-learning systems. In this paper, we present a method to estimate actual knowledge and confidence using a limited set of questions in an E-learning environment.

1.3. Multiple choices

Multiple-choice type test questions are easiest to use in E-learning. However, they do not necessarily correspond to knowledge. One person, without any knowledge of the subject can receive a non-zero exam score by randomly answering questions. Moreover, people who are confident but not knowledgeable can still increase their test scores. It is, therefore, essential to develop assessment methods to differentiate and evaluate knowledge and confidence individually from replies to multiple-choice questions.

1.4. Contributions

1) We show that actual test results are not good indicators of knowledge. 2) By providing additional types of questions, it is possible to estimate both knowledge and confidence.

2. Theoretical framework

2.1. Common knowledge in a learning environment

Any learning environment (Classroom or E-learning) consists of only two types of agents: teachers and students. Edwards and Mercer recently formulated that the goal of teaching is to eventually reach a common knowledge among all agents, teachers and students (Edwards & Mercer, 2013). In other words, the goal of teaching is that 1) the students must not only learn the subject but also become aware that they now know the subject and 2) teachers must know that the students know the subject.

2.2. Ideal environment

Ideally, teachers are the only ones that know the subject initially. Teachers also know that they, themselves, know the subject. On the other hand, initially, students are 1) not supposed to know the subject and 2) supposed to know that they do not know the subject. The goal is to develop methods to reach common knowledge in a fast and efficient way (Edwards & Mercer, 2013). We observe that this ideal environment does not exist. In this work, we present an approach for the development of methods to reach common knowledge even in a non-ideal environment. We observe that the confidence of agents is one of the main issues, which changes an environment from ideal to non-ideal.

2.3. Confidence

Confidence is not a type of knowledge but a type of belief that manifests itself from people's subjective observations of their own knowledge. These subjective observations may not necessarily be correct. Confidence is, therefore, different than knowing-to-know, which is a common knowledge concept of people's awareness of their own knowledge (Aumann, 1976; Rescher, 2005). Knowing-to-know is a type of knowledge and it is always correct.

2.4. Issues with confidence

We observed that there are two types of issues with confidence: 1) High knowledge and low confidence and 2) Low knowledge and high confidence. The people with high knowledge and low confidence cannot be sure their answers are correct even when they know the answers. On the other hand, the people with low knowledge and high confidence can believe that they know the subject and resist learning. Both of the cases cause irrational actions that can make it hard to evaluate achievement of common knowledge.
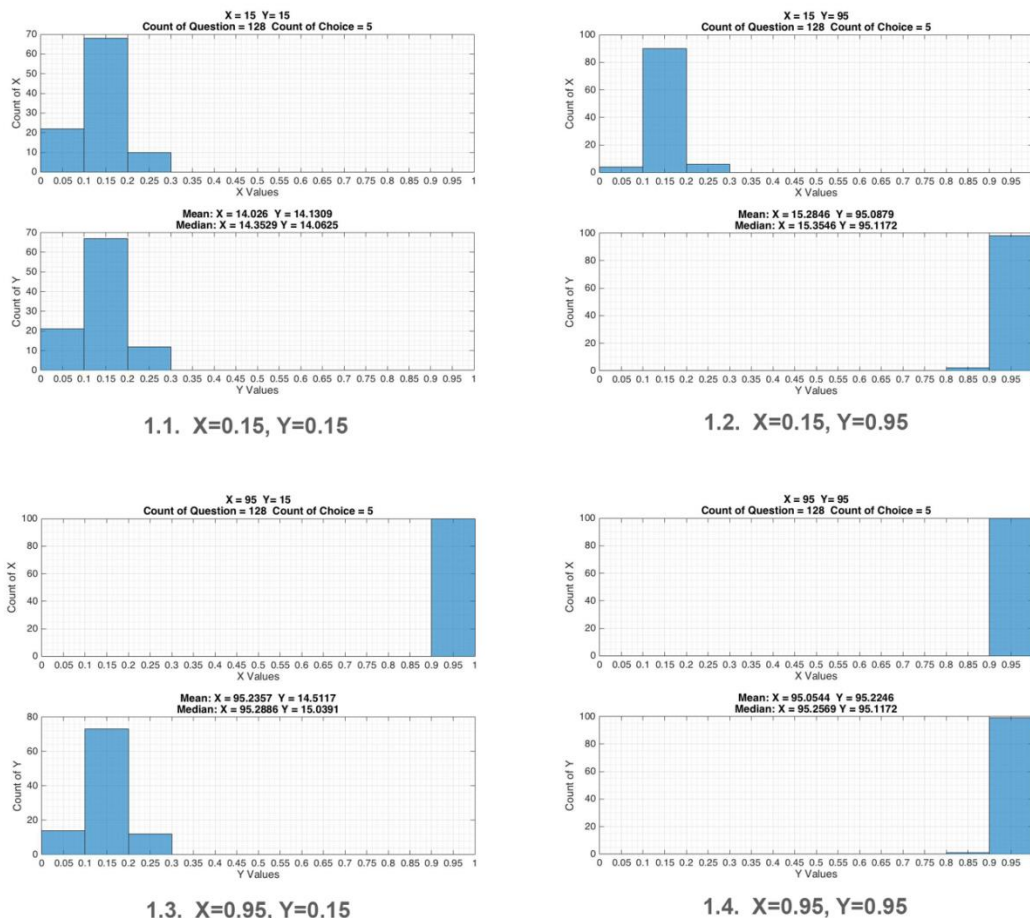

Figure 1. Examples of results from our estimation of knowledge and confidence. *X* denotes knowledge as a percentage of all knowledge of the given topic and *Y* denotes confidence as a percentage of all knowledge of the given topic. These graphs show our estimation of knowledge and confidence based on 100 random Monte-Carlo simulations using 128 test questions. Our method can identify actual knowledge and confidence with less than 10% of a standard deviation

2.5. Modeling confidence

To model confidence, we use a model proposed by Akleman et al. (2015) to represent narrative events. In this model, actions and states of agents are represented in three—1) Physical, 2) Expression, and 3) Observation—layers, as shown in Figure 2.

2.6. Physical layer

This is the lowest layer that provides the precise description of real events and real states such as interactions among agents and emotions of agents. This is the layer where all agents physically exist and interact. We assume that in any given time, a physical layer has a well-defined state, which is given as the collection of all the states of all agents in the physical layer. In this particular case, physical layer is simple. It contains only real knowledge and real confidence as a simple ratio. *X* and *Y* in the Figure 1 correspond to real knowledge and confidence. No agent can really know its own knowledge and confidence.

2.7. Expression layer

This layer gives us a description of how the agents express their internal states. The expression layer also provides a guidance of how the events and states could or should be visualised. In our case, expressions are actual answers to a set of questions given by teachers. Expressions are random functions of physical states. For instance, agents with a careful personality, a high knowledge and a high confidence will most likely answer correctly to any given question. On the other hand, sloppy agents even with a high knowledge and a high confidence will answer some questions wrong even when they actually know the answer. More interestingly, knowledgeable sloppy agents with low confidence will check their answers several times because of their low confidence and they will answer most questions correct. In other words, right or wrong answers are just expressions produced by complex interactions resulting from physical layers.

2.8. Observation layer

This is the layer where each agent's observations are kept. Each agent's observation can be different since every agent can observe a different subset of the expression layer. Moreover, one agent can interpret the meaning of an expression different from another agent. Narration only exists in this layer. Narration comes from the limited knowledge of agents and trigger further events through feedback to the physical layer. Teachers' assessment of student performances is type of observations. Those observations can also be imperfect. Moreover, they can be the function of teachers' physical layers. For instance, a sloppy teacher can simply make mistakes in evaluating a student's performance. Some teachers may evaluate their students differently based on positive and negative feeling towards each student. Note that teachers themselves may not even observe those feelings.
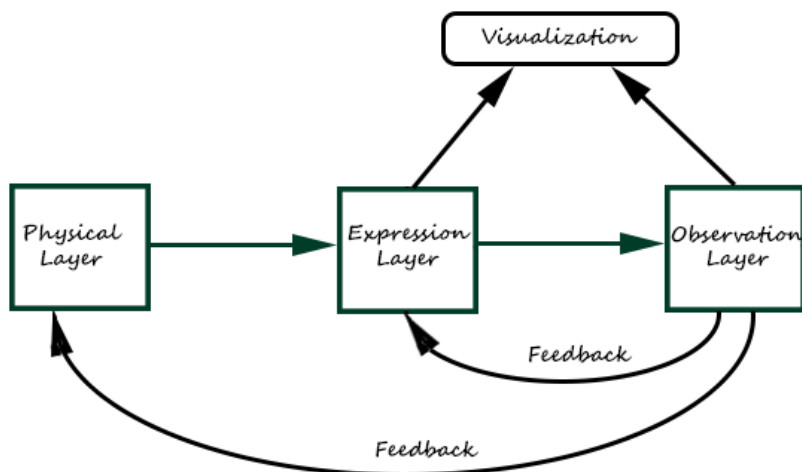
Figure 2. We view every agent as an entity that consists of three layers

2.9. Self-observation or self-awareness

We consider self-observation or self-awareness, observations of expressions of others' internal states and observations of expressions of relational states. Note that self-awareness may affect the expression. Therefore, there is a need for a feedback from the observation layer to the expression layer in Figure 2. The effect can be two ways. Agents who are aware of their internal states can hide it or show it based on their personality. For instance, sloppy agents who are aware of their sloppiness can fight against it by checking their questions carefully. In other words, sloppiness in the physical layer can manifest itself as carefulness in some cases when the agent is aware of this problem.

2.10. Simplifications for simulations

This framework helps us to simulate very complex interactions by including irrational behaviour. In this work, we made a series of simplifications: 1) we assume that teachers behave rationally. This is a reasonable assumption for games or distance education in which the evaluations are done automatically. 2) We assume that the personality traits of students play no role. In other words, we ignore cases such as sloppiness. We assume that the only traits playing a role in getting grades are knowledge and confidence. Based on these simplifications, we quantitatively demonstrated that confidence is an important variable in reaching learning objectives.

2.11. Observations

The mean and median of knowledge and confidence can be estimated reliably for large groups even asking a small set of questions. This is helpful to estimate initial knowledge and confidence of the group. It is possible to estimate knowledge and confidence of individuals using a high number of questions. This is useful to identify people who learned the subject and to give them appropriate grade.

3. Methodology

3.1. Question type classifications

We identified what kind of multiple-choice questions are needed to differentiate knowledge and confidence. We, then, classified the multiple-choice questions into two categories: A and B types as follows.

A Type Questions: In this type of question, the true answer is 'always' given as one of the choices and the answer choice 'none of the above' is always the wrong answer. Two examples of A type of questions are given below:

Question: What is the capital of France?

1. Paris
2. Bordeaux
3. Lille
4. None of the above

Question: What is the sum of the angles of a hexagon?

1. 360°
2. 540°
3. 720°
4. None of the above

B Type Questions: In this type of question, the true answer is never given in one of the choices and the answer choice 'none of the above' is always correct answer. Two examples of B type of questions are given below

Question: What is the capital of France?

1. Marseille
2. Lyon
3. Nice
4. None of the above

Question: What is the sum of the angles of a hexagon?

1. 180°
2. 360°
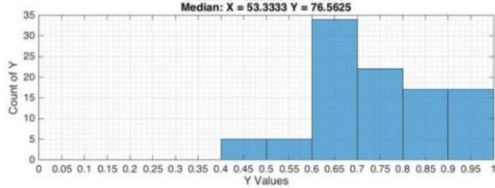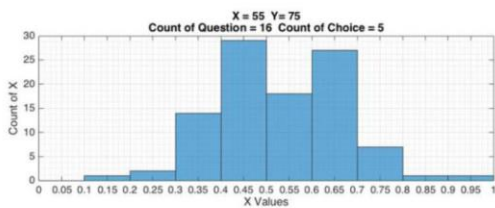3. 540°
4. None of the above

All of these questions and their answers can randomly be created in any distance education course or game. Moreover, by using large question and answer banks, we can significantly reduce the probability of getting the same questions or the same answers.
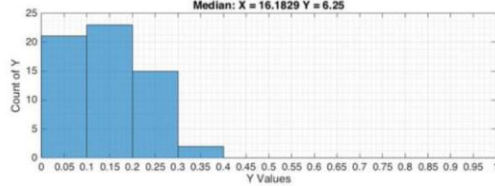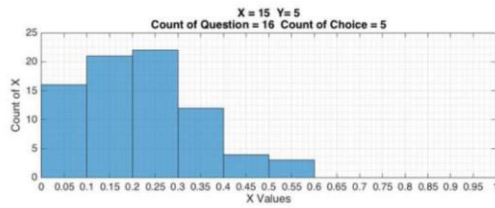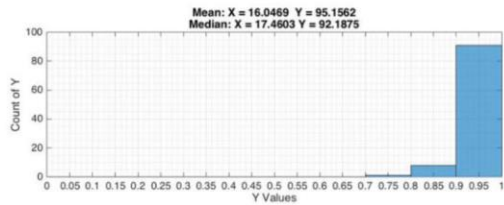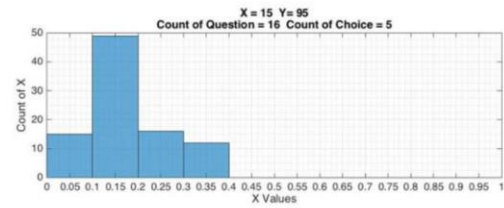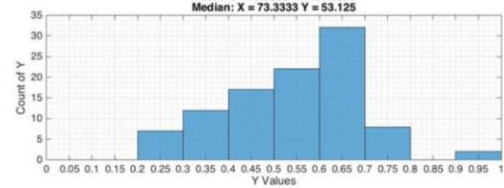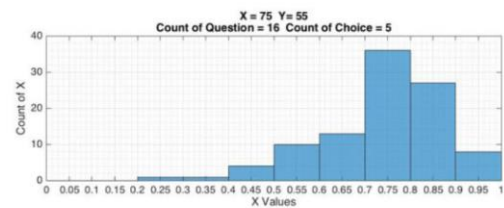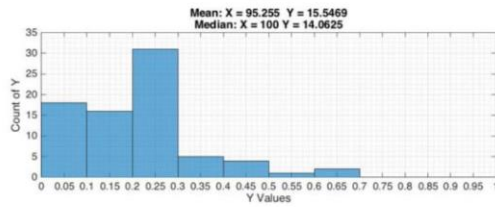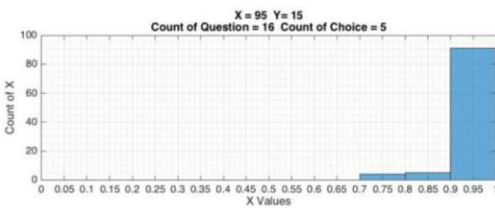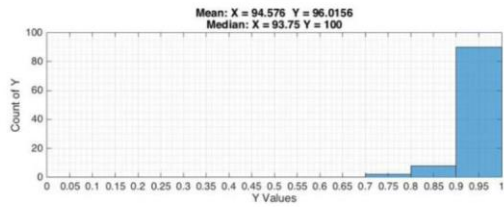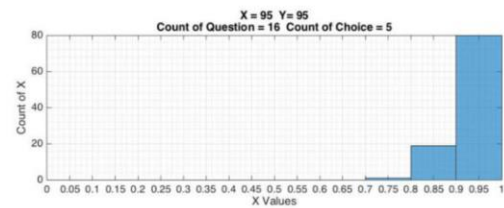
Figure 3. This figure shows that our method can provide an accurate estimation of the mean and median of knowledge and confidence by asking as few as 16 questions. However, for such small sets of questions, standard deviation increases, especially in the cases of low knowledge and low confidence as shown in tables. Therefore, for individuals of low knowledge and confidence, identification of actual knowledge and confidence cannot be reliable for such small sets of questions

3.2. Reactions of different personalities to A and B type questions

We can now analyse reactions of different personalities to type A and type B Questions.

- **Knowledgeable to type A:** We strongly believe that all knowledgeable people (i.e., $X = 1$) can correctly answer type A questions regardless of their confidence level. In other words, our conjecture is that having a true answer among the choices helps to develop confidence among knowledgeable people and knowledgeable people will always get a perfect score, i.e., 1, from these types of questions.
- **Confident to type B:** We strongly believe that all confident people (i.e., $Y = 1$) can correctly answer type B questions regardless of their knowledge level. In other words, our conjecture is that having no true answer among the top choices helps to make confident people choose 'none of the above', which is always correct answer in B type of questions. Therefore, confident people will always get a perfect score, i.e., 1, from these types of questions.
- **Non-Confident to type B:** We believe that for the people with no confidence (i.e., $Y = 0$), all answers will look equally probable. Therefore, our conjecture is that non-confident people will randomly answer to type B questions. In other words, these people will get 1/N on average from these types of questions, where N is the number of choices.
- **Non-Knowledgeable and Confident to type A:** We believe that for confident people with no knowledge (i.e., $X = 0$ and $Y = 1$), 'none of the above' option in type A questions will always look correct. Therefore, our conjecture is that these people will always make mistake in type A questions and will get a zero.
- **Non-Knowledgeable and Non-Confident to any questions:** We believe that for people with no confidence and no knowledge (i.e., $X = 0$ and $Y = 0$), all answers will look equally probable. Therefore, our conjecture is that these people will randomly answer to all questions. In other words, these people will get 1/N on average from these types of questions, where N is the number of choices.

3.3. General equation for X and Y

Using the above discussion, we can derive two bilinear equations that can provide estimation of grades for A and B type questions for an infinite number of questions as follows:

$$A = \frac{1}{N}(1-X)(1-Y) + 0(1-X)Y + 1XY + 1X(1-Y)$$

$$B = \frac{1}{N}(1-X)(1-Y) + 1(1-X)Y + 1XY + \frac{1}{N}X(1-Y)$$

where N is the number of choices, A and B are grades given as a number between 0 and 1, where 1 means a perfect score. The numbers **1, 0** and **1/N** are control points of the corresponding bilinear shape that comes from our earlier analysis. We can simplify these equations further as

$$A = \frac{1 + (N-1)X + Y + XY}{N}$$

$$B = \frac{1 + (N-1)Y}{N}$$

Note that B reduces to a linear equation and A is a simple bilinear equation. In these equations, A and B are the only observable variables by the evaluators/teachers. X and Y are really parameters hidden in the physical layer that are not known. Therefore, the question is if we observe A and B, can we estimate hidden X and Y values.

Solving X and Y for a given A and B: Estimation of knowledge and confidence from exam results is simply the estimation of *X* and *Y* from a given set of (*A*; *B*). Since these two equations are simple, there is always a unique solution to this problem. The only issue is that the Eqs. (1) and (2) correspond to an idealised case that can only be obtained by asking an infinite number of questions. In practice, we cannot ask even 1,000 questions. If the number of questions is not high, *A* and *B* values depend on random factors and two students with exactly the same knowledge and confidence can get significantly different *A* and *B* test results. It is, therefore, important to evaluate the effect of the number of questions in our estimations.

4. Implementation

4.1. Monte-Carlo simulation

To identify the effect of a limited number of questions, we have developed a Monte-Carlo simulation to estimate grades of *A* and *B* for individuals with an *X* amount of knowledge and a *Y* amount of confidence. We observe that such an individual can know the answer to a given problem with *X* probability and can become sure of the answer with *Y* probability. Therefore, for every question, we can randomly create a profile and based on this profile, we can make the individuals answer the questions.

4.2. Monte Carlo implementation

If it turns out that a particular individual knows the answer and is confident that the answer is correct for a *B* type of question, that particular individual answers the question correctly. On the other hand, if it turns out that another individual does not know the answer and is not confident that the answer is correct for a question, that individual randomly selects an answer, which can be correct with 1/N probability. It is possible to ask a set of questions to a given artificial agent and compute *A* and *B* values using this observation. In the Monte-Carlo implementation, we ask an equal number of *A* and *B* type questions. By giving the same exam to a number of artificial agents, we can compute the distribution of *A* and *B* values, which corresponds to classical grades.

4.3. Estimating X and Y

For given *A* and *B* values using the perfect situation described in Eqs. (1) and (2), we can estimate *X* and *Y* values. Note that when we make this computation, we will obtain a statistical distribution of *X* and *Y* values (see Figure 3). These distributions appear symmetric for mid-range real *X* and *Y* values. However, when *X* and *Y* are close to boundaries, i.e., 0 and 1, the distributions are skewed, as shown in Figure 3.

5. Analysis of results

5.1. Estimating group knowledge and confidence

If the number of questions is above 16, regardless of the number of questions, we always get a very good estimation of the mean and median of *X* and *Y*. Therefore, a few numbers of questions can be sufficient to estimate the overall knowledge and confidence of a class, as shown in Figure 3.

5.2. Reaching learning objectives

Estimation of group knowledge and confidence is especially useful for evaluating a course whether it has reached its learning objectives. Since a few numbers of questions are sufficient to estimate overall knowledge and confidence of a class, simple pre- and post-class surveys can be sufficient to

estimate the overall increase in knowledge and confidence of the whole class. Ideally, we want to see a distribution like in Figure 3 at the beginning of the class and in Figure 3 at the end of the class.

5.3. Individual knowledge and confidence

Our results demonstrate that to obtain very accurate estimations of actual knowledge and confidence of individuals, we need to ask at least 100 questions. As shown in Figure 1, we still make some grading errors with students of low knowledge. That is acceptable since such low knowledge does not warrant passing the course and small mistakes in actual grades do not matter. On the other hand, for highly knowledgeable individuals regardless of their confidence, we can obtain reasonably accurate results for even 32 questions, as shown in Figure 4.
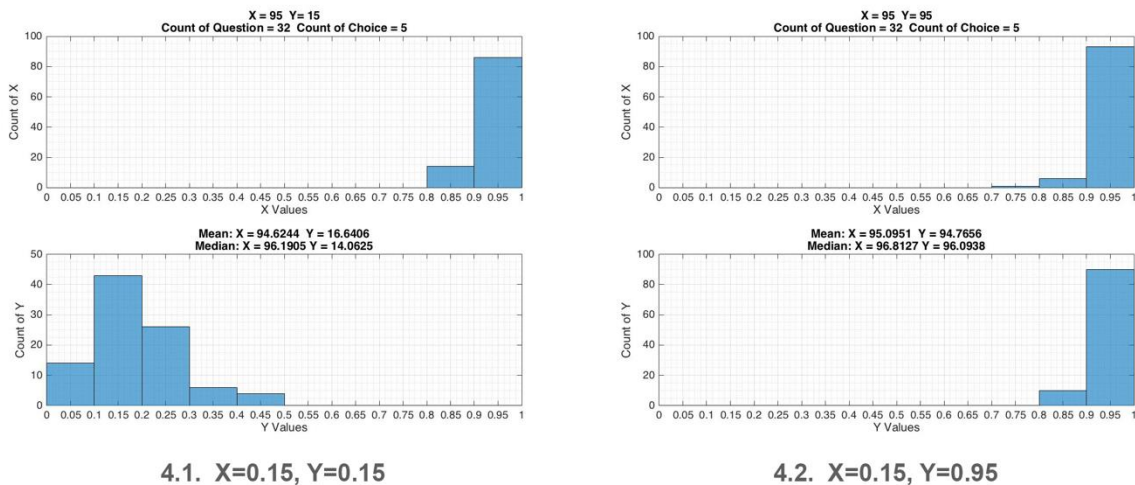

4.1. X=0.15, Y=0.15   4.2. X=0.15, Y=0.95

Figure 4. This figure shows that the method can provide an accurate estimation of individual knowledge and confidence for highly knowledgeable individuals by asking as few as 32 questions

6. Discussion

6.1. Individual grading

These results question current methods of grading based on test results. If the number of *B* type questions is high, highly knowledgeable but non-confident people can get grades much lower than their actual knowledge levels. If we use *X* values for grading, grades can be much better representations of actual knowledge levels of individuals.

6.2. Increasing confidence

One of the crucial issues in teaching is increasing confidence as mentioned earlier (Chappuis & Stiggins, 2002; Stiggins, 1992; Stiggins & Chappuis, 2005). Since our method can provide an estimation of confidence, this information can be used to improve confidence for highly knowledgeable individuals. Another important use of our method is that it can be used to reduce overconfidence as well, which is a significant obstacle for learning new ideas.

6.3. Conclusion and future work

In this work, we demonstrated that assessment of knowledge and confidence is possible by only asking multiple-choice questions using Monte-Carlo simulation. Our results show that we can reliably

predict actual knowledge and confidence using approximately 100 test questions. Assessment of reaching learning objectives for the whole group of students can be done by asking as few as 16 questions.

6.4. Reduction of number of questions

In this work, we assume all questions are equally difficult and student knows only a percentage of the domain. This assumption is valid for subjects that do not build upon previous knowledge. We are planning to extend our simulations by using questions with varying difficulty levels that rely on each other. Using such questions and feedback mechanisms, we expect that it is possible to identify knowledge and confidence levels with less number of questions.

6.5. Ratio of A and B type questions

In the simulation, we assume that the number of A and B questions are equal. It can be possible to reduce the total number of questions by using an unequal number of A and B types.

6.6. Other question types

Although our analysis based only on multiple choice-type questions, the method can easily be extended to include other type of questions such as multiple selections. We expect those questions can quickly identify high-knowledge individuals; however, they do not help to differentiate between different knowledge levels. We expect that it is possible to analyse the responses to predict actual knowledge and confidence in such cases also.

6.7. Other answer types

Our current model cannot provide precision for low knowledge. For instance, both 0:1 and 0:001 are very small but one of them 100 times more in terms of dynamic range. We are planning to study the effect of meaningless answers to better identify significant problems in knowledge. This can be especially important to diagnose the problems that can prevent to learn.

References

Ainsworth, L., Almeida, L., Davies, A., DuFour, R., Gregg, L. & Guskey, T., et al. (2007). *Ahead of the curve: the power of assessment to transform teaching and learning*. Bloomington, IN: Solution Tree.

Akleman, E., Franchi, S., Kaleci, D., Mandell, L., Yamauchi, T. & Akleman, D. (2015). A theoretical framework to represent narrative structures for visual storytelling. *Paper presented at the proceedings of bridges 2015: mathematics, Music, art, architecture, culture*.

Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, 1236–1239.

Chappuis, S. & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership, 60*(1), 40–44.

Edwards, D. & Mercer, N. (2013). *Common knowledge: the development of understanding in the classroom*. Routledge.

Peyre, S. E., Peyre, C. G., Sullivan, M. E. & Towfigh, S. (2006). A surgical skills elective can improve student confidence prior to internship. *Journal of Surgical Research, 133*(1), 11–15.

Rescher, N. (2005). *Epistemic logic: a survey of the logic of knowledge*. University of Pittsburgh Pre.

Rosenberg, M. J. (2001). *E-learning: strategies for delivering knowledge in the digital age* (Vol. 9). New York, NY: McGraw-Hill.

Stiggins, R. J. (1992). High quality classroom assessment: what does it really mean?. *Educational Measurement: Issues and Practice, 11*(2), 35–39.

Kaleci, D. & Akleman, E. (2018). Assessment of knowledge and confidence for E-learning. *World Journal on Educational Technology: Current Issues. 11*(1), 104–115.

Stiggins, R. J. (1999). Assessment, student confidence, and school success. *The Phi Delta Kappan, 81*(3), 191–198.

Stiggins, R. & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory into practice, 44*(1), 11–18.

Zhang, D., Zhao, J. L., Zhou, L. & Nunamaker, J. F. (2004). Can e-learning replace classroom learning?. *Communications of the ACM, 47*(5), 75–79.