Evaluation of the Correlation Between Formative Tests and Final Exam Results – Theory of Information Approach

Jerzy Rutkowski

Abstract—Today, computerization of the Assessment Program is a norm, numerous advantages of Computer-Assisted Assessment (CAA), both formative and summative, can be enlisted. Motivating students to systematic work and Self-Regulated Learning seems to be the biggest challenge faced by the teacher. This motivation can be reached by providing students with clear information that shows an evident effectiveness of formative quizzes, undisputable correlation between student engagement in taking formative quizzes and the final exam result. It is proposed to evaluate this effectiveness quantitatively in the field of Information Theory, using the Discrete Memoryless Channel description of relationship between the set of quizzes taken and the set of exam results. The Case Study is presented and if proves high correlation between these two sets.

Keywords—Technology Enhanced Learning; Computer-Assisted Assessment

I. INTRODUCTION

Nowadays universities are in a heart of global revolution in education, due to dynamic progress in Information & Communication Technology (ICT) and common access to Internet. These enable significant enhancement of quality of education, as many new forms of knowledge delivery and assessment may support the face-to-face (f2f) form. There is considerable diversity among course delivery methods used by individual instructors. However, taking into account percentage of content delivered online, the following basic course classifications can be given[1]:

- Traditional: no online technology used, content is delivered in writing or orally.
- Web facilitated: ICT is used to facilitate what is essentially a f2f course--web pages are only used to post the syllabus and assignments; content delivered online ranges between 1% and 29%.
- Blended/Hybrid: online and f2f delivery are blended, substantial proportion of the content is delivered online, typically uses online discussions and has a reduced number of f2f meetings; content delivered online ranges between 30% and 79%.
- Online: most or all of the content (80%–100%) is delivered online, typically no f2f meetings.

It is commonly accepted that blended model is the best solution to constraints and disadvantages of traditional (f2f) and online model. Two major components can be distinguished in Learning Activities of a blended course:

- Learning Content (knowledge delivery)
- Assessment Program (knowledge assessment).

Both have to be tightly correlated and comply with the applied format of blended course, e.g. both have to have a modular structure in case of flip teaching format [2]. Today, computerization of the Assessment Program is a norm, numerous advantages of Computer-Assisted Assessment (CAA), both formative and summative, can be enlisted [3]. Some extensive research has already been done in this field, effective CAA systems have been reported [2-7]. This research shows that, when suitably organized, automated self-assessment can lead to significant enhancements in learning and achievement. However, further extensive studies are necessary to make the automated assessment system fully reliable, such that traditional assessment methods can be effectively replaced, in fact practically eliminated. Obviously, when designing CAA, the teachers have to remember that students can, with difficulty, escape from the effects of poor teaching; they cannot escape the effects of poor assessment [8].

To make full use of formative quizzes, students have to accept Self-Directed Learning (SDL) philosophy [9], also called Self-Regulated Learning (SRL). A model of SRL is presented in Fig. 1 [10,11].

There is considerable research evidence to show that effective Feedback, both Internal and External, leads to learning gains. A synthesis of the research literature led to the following seven feedback principles, which support and develop self-regulation in students learning [10,11]:

1. helps clarify what good performance is (goals, criteria, expected standards);
2. facilitates the development of self-assessment (reflection) in learning;
3. delivers high quality information to students about their learning;
4. encourages teacher and peer dialogue around learning;
5. encourages positive motivational beliefs and self-esteem;
6. provides opportunities to close the gap between current and desired performance;
7. provides information to teachers that can be used to help shape the teaching.

While there would normally be an overlap between the student’s goals and those of the teacher, the degree of overlap may not be high, e.g. if the student wishes only to pass the assignment. Students can only achieve learning goals if they

This work was supported by the Ministry of Science and Higher Education funding for statutory activities (decision no. 8686/E-367/S/2015)
J. Rutkowski is with the Institute of Electronics, Silesian University of Technology, Gliwice, Poland (e-mail: jrutkowski@polsl.pl).
understand those goals. Understanding engineering concepts, links between theory and practice is the main learning goal, as stated by the teacher. Unfortunately, significant mismatches between teachers’ and students’ conceptions of goals and of assessment criteria and standards can be observed. Teacher’s goals can’t be reached when learning by heart is in force. Students, following bad practices acquired at the high school, consider learning by heart as the best way of achieving learning goals and consequently passing the exam.

To conclude, the following main barriers in making formative quizzes more effective, as observed by the Author and confirmed by other researchers, can be enlisted:

1. Students SDL Readiness (SDLR) [12] is low, freshmen SDLR in particular.
2. If students perceive formative assessment as primarily examining content knowledge, they will tend to do little more than rote learning, especially when they wish only to pass the exam [13].
3. Students tend to ignore activities that do not directly contribute to grades and degree class; even though they could see the benefit of developing competencies, they do not take advantage of it [13].

A comprehensive study on how to overcome these barriers, develop formative assessment strategies and student SDL can be found in [10,13]. Normally, formative online quizzes are obligatory, but unfortunately, significant portion of students treat them as a necessary evil rather than scaffolding to reach higher order thinking skills. Such students are not systematic in learning, solve online quizzes just days or even hours before the deadline, and only the obligatory percentage of tasks, perhaps some of them solve tasks not fully by themselves. Then, regardless of all three barriers mentioned before, motivating students to systematic work and self-solving of majority of tasks seems to be the biggest challenge faced by the teacher. This motivation can be reached by providing students with clear information that shows an evident effectiveness of formative quizzes, undisputable correlation between student engagement in taking formative quizzes and the final exam result. It is proposed to evaluate this effectiveness quantitatively in the field of Information Theory, developed by C.E. Shannon in the late 40’s of the last century. The relationship between activity in formative quizzes and exam results is described by means of Discrete Memoryless Information Channel (DMC). Principles of DMC description and calculation of Mutual Information are presented in Section II. A case study, use of DMC to evaluate effectiveness of formative quizzes for Electric Circuit Analysis course, is presented in Section III, some guidelines and conclusions are given in Section IV. The used Calculated questions are briefly discussed in Section V (Appendix).

II. DISCRETE MEMORYLESS CHANNEL

Fig. 2 presents Source-Channel-Receiver information system [14].

\[ X = \{X_1, \ldots, X_M\} \]  

\[ Y = \{Y_1, \ldots, Y_N\} \]

\[ P_X = \{p(X_1), \ldots, p(X_M)\} \]  

\[ P_Y = \{p(Y_1), \ldots, p(Y_N)\} \]

is the discrete input source of information, in short the Source, set of samples characterized by the probability assignment

is the channel output source, in short the Receiver, characterized by the probability assignment

\[ H(X) \]

\[ H(Y) \]

\[ H(X/Y) \]

\[ H(Y/X) \]

\[ I(X/Y) \]
Channel itself is characterized by $MN$ transition probabilities that relate events of input and output source:

$$p(Y/X_i); i=1,...,M, j=1,...,N$$ (3)

Then, for the given probabilistic model of the Source and the Channel, information loss $H(X/Y)$, misinformation $H(Y/X)$ and mutual information $I(X/Y)$ can be defined. Mutual information between events (sources) $X$ and $Y$ is the information provided about the event $X$ by the occurrence of the event $Y$, or vice versa.

$$I(X/Y) = H(X) - H(X/Y) = H(Y) - H(Y/X)$$ (4)

$$H(X) = -\sum_{i=1}^{M} p(x_i) \log_2 p(x_i)$$ (5a)

$$H(Y) = -\sum_{j=1}^{N} p(y_j) \log_2 p(y_j)$$ (5b)

$$H(Y/X) = -\sum_{i=1}^{M} \sum_{j=1}^{N} p(x_i, y_j) \log_2 p(y_j / x_i)$$ (6)

Exemplary channels: binary channel and 3-input/3-output channel, are presented in Fig. 3.

An exam can be considered as measurement of students’ knowledge and then, it can be described by Discrete Memoryless Information Channel (DMC). Students’ knowledge, the measured quantity, can be expressed by a number of formative quizzes taken and, after discretization, it consists the set of samples, the input Source. Set of exam results consists the output source (Receiver). Exam can be considered as measurement of students’ knowledge. It can be assumed that this knowledge is designated by formative quizzes taken, number of tasks solved by the student. For the binary channel ($M=N=2$), both sources can be discretized as follows:

- $X_1=X_D$ Diligent students, $D$ students that solved at least 50% of tasks,
- $X_2=X_M$ Minimalists, $M$ students that solved less than 50% of tasks,
- $Y_1=Y_P$ P students that Passed exam,
- $Y_2=Y_F$ F students that Failed exam.

It is assumed that numbers $D, M, P, F$ are known and they designate probability assignments $P_X$ and $P_Y$, e.g. $p(X_D)=D/E; E=D+M=P+F$. Conditional probabilities that relate students’ diligence and exam results (3):

$$p(Y_i/X_3), i=D, M; j=P, F$$

are also known, e.g. $p(Y_P/X_M)=P_M/M$ is the probability of Passing the exam by the Minimalist, where $P_M$ is the number of Minimalists that Passed. Then, relationship between formative assessments and summative assessment (exam) can be modeled by means of binary information channel, as depicted in Fig. 3, and mutual information can be calculated. This information may be interpreted as the information provided about the measured data (students’ knowledge) by the occurrence of measurements (exam), as the measure of formative assessment effectiveness.

More complex, 3-input/3-output channel ($M=N=3$) can be considered, Diligent students can be split into:

- $X_1=X_{DH}$ $D_{Hi}$ students that solved more than 67% of tasks,
- $X_2=X_{DL}$ $D_L$ students that solved between 50% and 67% of tasks,
- $X_3=X_M$ $M$ Minimalists.

Pass can be split into:

- $Y_1=Y_{PH}$ $P_H$ students that passed with mark 4 out of 5 or higher,
- $Y_2=Y_{PL}$ $P_L$ students that passed with mark 3.5 or 3,
- $Y_3=Y_F$ $F$ students that Failed exam.

The conditional probabilities have to be split as well, e.g. $p(Y_{PH}/X_{DL})=P_{HL}/D_L$ is the probability of Passing High the exam by Diligent Low student, where $P_{HL}$ is the number of Diligent Low that Passed High. Then, the relationship between formative assessments and summative assessment (exam) can be modeled by means of 3-input/3-output information channel, as depicted in Fig. 3.

III. CASE STUDY: ELECTRIC CIRCUIT ANALYSIS COURSE

The described methodology of evaluation of relationship between number of formative quizzes taken and exam results has been verified in the academic year 2014/2015. An ensemble of $E=50$ (all enrolled) students that entered the Electric Circuit Analysis exam constituted the test group. Formative Moodle [15] quizzes consisted of nine quizzes with fifteen Calculated questions (see Section V) per each quiz, which gives the total of 135 questions. The quiz was considered Passed if at least eight questions were answered. To be admitted to the exam, passing of five quizzes was obligatory and this gives the minimum limit of 40 questions = 30% of all questions. Taking into account student diligence in
solving quizzes, the ensemble of E=50 students has been split into:

- \( D=22 \) Diligent students: students that solved at least 50% of questions (67 out of 135),
- \( M=28 \) Minimalists: students that solved less than 50% of questions (from 40 to 66).

The exam quiz consists of ten questions: eight calculated questions, marked 0 or 1 and two multiple-choice, marked 1, 0 (no answer) or −0.5 (wrong answer). The calculated questions are drawn from formative quizzes, eventually with slight modifications. The pass threshold at 3.5 points has been experienced as the most adequate [6]. The following numbers have been collected:

- \( P=23 \) number of students that Passed exam,
- \( F=27 \) number of students that Failed exam,
- \( P_H=19 \) number of Diligent students that Passed exam,
- \( F_D=3 \) number of Diligent students that Failed exam,
- \( P_L=4 \) number of Minimalists that Passed exam,
- \( F_M=24 \) number of Minimalists that Passed exam.

The corresponding Binary Memoryless Channel (BMC) is depicted in Fig. 4.

\[
\begin{align*}
 p(X_d)=0.44 & & p(Y_d/X_d)=0.86=19/22 & & p(Y_d)=0.46 \\
p(Y_d/X_M)=0.14=4/28 & & p(Y_d/X_M)=0.14=3/22 \\
p(X_M)=0.56 & & p(Y_d/X_M)=0.86=24/28 & & p(Y_d)=0.54
\end{align*}
\]

Fig. 4. Case Study: Binary Memoryless Channel

Input (quizzes) entropy, output (exam) entropy, misinformation and mutual information are as follows, where \( \log \) is the binary logarithm, i.e., the logarithm to the base 2 (logarithmus duālis):

\[
\begin{align*}
 H(X) &= -0.44\log 0.44 - 0.56\log 0.56 = 0.99 \text{ bit} \\
 H(Y) &= -0.46\log 0.46 - 0.54\log 0.54 = 0.99 \text{ bit} \\
 H(Y/X) &= -0.44(0.86\log 0.86 + 0.14\log 0.14) - 0.56(0.14\log 0.14 + 0.86\log 0.86) = 0.58 \text{ bit} \\
 I(X/Y) &= H(Y) - H(Y/X) = 0.41 \text{ bit}
\end{align*}
\]

The maximum mutual information that can be obtained in the binary channel is information of the noiseless channel: \( p(Y_i/X_i)=p(Y_i/X_i)=0 \), i.e., when all Minimalists Fail, all Diligent students Pass:

\[
I(X/Y)_{\text{max}} = H(X) = 0.99 \text{ bit}
\]

If this information is considered as the primary reference one, then, the normalized mutual information can be calculated:

\[
I_c=I(X/Y)/I(X/Y)_{\text{max}} = 0.41/0.99 = 0.41 \text{ bit/bit}
\]

Such reference channel is too idealistic and practically unreachable. Practice shows, that some Diligent students Fail the examination, due to examination stress rather than insufficient knowledge. However, for the reference channel, all Minimalists should Fail. Then, channel with unchanged probabilities \( p(Y_i/X_d) \), and changed probabilities \( p(Y_i/X_M) \); \( j=P, F: p(Y_i/X_M)=0, p(Y_i/X_M)=1 \) could be considered as the reference one. Output entropy, misinformation, mutual information of this reference channel are:

\[
\begin{align*}
 H(X)_{\text{ref}} &= -0.38\log 0.38 - 0.62\log 0.62 = 0.96 \text{ bit} \\
 H(Y/X)_{\text{ref}} &= -0.44(0.86\log 0.86 + 0.14\log 0.14) = 0.26 \text{ bit} \\
 I(X/Y)_{\text{ref}} &= H(Y) - H(Y/X)_{\text{ref}} = 0.7 \text{ bit}
\end{align*}
\]

and then, the normalized mutual information:

\[
I_c=I(X/Y)/I(X/Y)_{\text{ref}} = 0.41/0.7 = 0.59; I_c\%=100I_c = 59%.
\]

To obtain more complex, 3-input/3-output channel \((M=N=3)\), set of \( D=22 \) Diligent students has been split into two sets, as described in Section II, and the following numbers have been collected:

- \( D_l=17 \) number of students that solved more than 67% of tasks (90 out of 135),
- \( D_1=5 \) number of students that solved between 50% and 67% of tasks (67 to 89).

Set of 23 students that Passed has been split into two sets and the following numbers have been collected:

- \( P_l=10 \) number of students that obtained 6.0 points or more and consequently Passed with the High mark (4.0, 4.5 or 5)
- \( P_1=13 \) number of students that more than 3.0 but less than 6.0 points and consequently Passed with the Low mark (3.0 or 3.5).

Numbers that designate conditional probabilities (describe Channel) have been collected in Table I, the Channel is presented in Fig. 5.

| Table I. Case Study: Numbers that Designate 3-input/3-output Channel |
|------------------|------------------|------------------|------------------|
| \( p(Y_i/X_i) \) | \( j=PH \) | \( j=PL \) | \( j=F \) |
| \( i=D_H \) | \( D_H=17 \) | \( 10/17 \) | \( 6/17 \) | \( 1/17 \) |
| \( i=D_L \) | \( D_L=5 \) | \( 0 \) | \( 3/5 \) | \( 2/5 \) |
| \( i=M \) | \( M=28 \) | \( 0 \) | \( 4/28 \) | \( 24/28 \) |

\[
\begin{align*}
 p(X_M)=0.34 & & 0.59=10/17 & & p(Y_M)=0.2 \\
p(X_M)=0.1 & & 0.6 & & p(Y_M)=0.26 \\
p(X_M)=0.56 & & 0.86=24/28 & & p(Y_M)=0.54
\end{align*}
\]

Fig. 5. Case study: 3-input/3-output channel
When observing this channel, the following main conclusions can be drawn:
1. Only students that solved more than 67% of tasks (Diligent High) Passed the exam with High mark.
2. Only 14% of students that solved less than 50% of tasks (Minimalists) Passed the exam (all with Low mark).
3. Only 6% of students that solved more than 67% of tasks (Diligent High) didn‘t passed the exam.
4. Solving between 50% and 67% of tasks (Diligent Low) doesn‘t guarantee Pass: 60% Pass versus 40% Fail.

Input (quizzes) entropy, output (exam) entropy, misinformation and mutual information are as follows:

\[ H(X) = -(0.34 \log_{10} 0.34 + 0.1 \log_{10} 0.1 + 0.56 \log_{10} 0.56) = 1.33 \text{ bit} \]

\[ H(Y) = -(0.26 \log_{10} 0.26 + 0.26 \log_{10} 0.26 + 0.54 \log_{10} 0.54) = 1.45 \text{ bit} \]

\[ H(Y/X) = 0.82 \text{ bit} \]

\[ I(X/Y) = H(Y) - H(Y/X) = 0.63 \text{ bit} \]

We may assume, same as for binary channel, that all Minimalists should Fail and consequently change probabilities

\[ p(Y_j/X_M) = \begin{cases} \Phi, & j = \Phi, PL, F; \\ p(Y_\Phi/X_M) = p(Y_{\Phi}|X_M) = 0, p(Y_{PL}/X_M) = 1, \end{cases} \]

to obtain the reference channel. Then,

\[ H(Y)_{ref} = -(0.2 \log_{10} 0.2 + 0.18 \log_{10} 0.18 + 0.62 \log_{10} 0.62) = 1.34 \text{ bit} \]

\[ H(Y/X)_{ref} = 0.49 \text{ bit} \]

\[ I(X/Y)_{ref} = H(Y) - H(Y/X)_{ref} = 0.85 \text{ bit} \]

and the normalized mutual information

\[ I_a = I(X/Y)/I(X/Y)_{ref} = 0.63/0.85=0.74; I_a\% = 100I_a/74 = 74\% !! \]

This high normalized mutual information proves very high correlation between the set of quizzes taken and the set of exam results, proves high effectiveness of the designed ECA formative assessments.

After failing the first attempt, the students still have had a chance to take quizzes. In majority, they took the opportunity and only four students (8%) didn‘t get the credit after three resits (four attempts). The obtained results have been confirmed, in the same exam session, by the ensemble of around 40 students, of the same course but different field of study.

IV. FINAL CONCLUSIONS AND GUIDELINES

A new method that enables quantitative evaluation of effectiveness of formative assessments has been presented. Theory of Information based approach is proposed, the concept of Discrete Memoryless Channel is utilized. Relationship between number of quizzes taken and the final exam results is described by means of mutual information. It has to be clearly stated, that the in-depth knowledge in the field of Theory of Information is not necessary to use this approach and understand the results.

Taking into account the case study results, the following main conclusions can be repeated:

1. Only students that have solved majority of formative questions (Diligent High students), practically more than 2/3 (67%), Pass the exam with High mark. Minority of students that have solved less than 50% of questions (Minimalists) Pass the exam, probability of Passing with a High mark is practically zero.
2. Very small percentage of Diligent High students Fail the exam, and even if they Fail by the first attempt they Pass High by the second attempt.
3. Solving between 50% and 67% of formative questions (Diligent Low attitude) doesn‘t guarantee Pass, gives only ≈50% chance of Pass.

These conclusions give very valuable feedback for both teachers and students:

- prove high quality of the designed ECA formative assessments, their compliance with the teaching goals and learning outcomes,
- constitute an evidence of very high correlation between formative quizzes and exam results and therefore guidelines for students how to overlap their Goals and Tactics with those of the teacher.

Obviously, students may ignore these evident guidelines and stay with their bad habits, may resist to break barriers in making formative quizzes more effective, as enlisted in Section 2. Increasing of the obligatory limit from 30% to 67% of tasks seems to be the best solution, leading to better Pass/Fail ratio. Unfortunately, it has been observed that majority of Minimalists have solved the obligatory 30% of tasks during the last week before the deadline and probably with illegal help of their peers. Then, increasing the obligatory limit may lead to more cheating rather than motivate Minimalists to become Diligent students.

Minimalists think that they may Pass the exam somehow, avoiding systematic and solid work, and persuading them that such work is the only way to Pass is a big challenge for educators. Presenting the reliable data, obtained be means of DMC approach, that expose high effectiveness of formative assessments, correlation with exam results, seems to be a good way to reach this goal.

The most recent development of the Computer-Assisted Assessment System has confirmed the correctness of this last conclusion. In winter semester, the first part of the Electric Circuit Analysis (DC Analysis) is lectured and students have to pass summative test (not the exam). During the first lecture of the winter semester 2015/2016, the case study findings have been communicated to students. Then, repeated many times during consecutive lectures. Nevertheless, most of the students were resistant to systematic learning, didn‘t take quizzes as long as they did not become obligatory. Solving formative quizzes was not obligatory before the first test, scheduled just before Christmas, and vast majority (≈90%) of students Failed, only students that solved more than 70% of formative questions (≈10%) Passed. The resit has been scheduled four weeks later and passing formative quizzes was necessary to be admitted. Nearly all students have fulfilled this obligation (passed formative quizzes) and 95% of them passed the resit, some 50% with High mark! It looks that bad habits, acquired at high school, still prevail during the first semester of study and a shock treatment applied during the first test was the only way to wake them up, to convince them that ignorance of systematic work and formative quizzes leads to failure in getting the credit.
V. APPENDIX – CALCULATED QUESTIONS

Calculated Moodle questions offer a way to create individual numerical questions by the use of wildcards \((x, y, \ldots)\) that are substituted with random values when the quiz is taken. When a student takes the test, Moodle will randomly select values for \((x, y, \ldots)\) and grade the response using the result of the Correct Answer Formula designed by the teacher. The main purpose of the calculated question is to create multiple versions of a question with different numerical values. The test will very rarely appear the same way twice as 100 wildcards are drawn and each wildcard variable: \(x, y, \ldots\), may take many discrete values. When Moodle delivers a Calculated question to the student, the wildcards are replaced with randomly selected values. However, these values are not completely random rather, they are randomly selected from a predefined dataset of possible values. This allows teacher some control over the possible values chosen for example, in order to make sure the numbers are realistic. The question is marked “0” = incorrect answer or “1” = correct answer. It is possible to allow a margin within which all responses are accepted as correct. The "Tolerance" field is used for this. Calculated questions can use more than simple arithmetic operators in Correct Answer Formula, some forty functions are available. The details can be found in “Calculated question type – MoodleDocs” [16].

REFERENCES