RESEARCH ARTICLE

Evaluation of Auto8Grade Grading Software

Aree Limwudhikraijirath

Abstract

The evaluation of Auto8Grade Grading Software (Auto8Grade) discusses the evaluation objectives, research methods, findings and recommendations, in effort to evaluate Auto8Grade to ensure that teachers will have ease-of-use valid grading software that they want. The evaluation was guided by descriptive method or non-experimental research design method. Archival research and survey research accompanying with interviewing were used. The study samples completed a survey in addition to contributing to a qualitative interviewing. The evaluator studied and examined existing books to make sure that Auto8Grade used an accurate process for 8-grade level grading system.

The evaluation revealed that (1) Auto8Grade is accurate grading software. Teachers can trace back from all worksheets that created by Auto8Grade that it used the accurate process as said in the books. (2) Auto8Grade has essential characteristics that teachers want, which are ease-of-use and flexibility that teachers can adjust grade boundaries easily. The need assessment report that the developer studied before the software development said that the most two essential characteristics of grading software that teachers wants were ease-of-use and flexibility that teachers can adjust grade boundaries easily respectively. Teachers satisfied in ease-of-use and flexibility of Auto8Grade. The most three satisfied characteristics are ease-of-use, providing descriptive statistics, and providing graph of score distribution respectively. (3) In order to increase number of teachers who use Auto8Grade, the evaluator recommend that the developer or the stakeholder should set some short time seminar for PR (public relation) this software, and the developer may improve this software so it can be generalized to be used for all grade level other than only 8-grade level grading system.

Keywords : Grading Software, 8-grade level grading system, ease-of-use and flexibility

บทความวิจัย

การประเมินโปรแกรมตัดเกรด Auto8Grade

อารีย์ ลิ้มวุฒิไกรจิรัฐ[1]

บทคัดย่อ

การประเมินโปรแกรมตัดเกรด Auto8Grade นี้ ครอบคลุมถึงจุดประสงค์ของการประเมิน วิธีการวิจัย ข้อค้นพบ และคำแนะนำ อันเป็นส่วนประกอบของความพยายามที่จะประเมินโปรแกรมตัดเกรด Auto8Grade ว่าครูอาจารย์ผู้สอนจะได้ใช้โปรแกรมตัดเกรดที่ใช้งานง่ายอย่างที่ตนต้องการ การประเมินได้ใช้วิธีการแบบพรรณนา ไม่ได้ใช้การทดลองปฏิบัติการ การประเมินมีการค้นคว้าเอกสาร การสำรวจข้อมูล และการสัมภาษณ์ร่วมด้วย กลุ่มตัวอย่างที่ทำการศึกษาได้ตอบแบบสอบถาม และให้สัมภาษณ์ซึ่งจะทำให้ได้ข้อมูลทั้งในเชิงปริมาณและคุณภาพ ผู้ประเมินได้ศึกษาและตรวจสอบจากหนังสือต่าง ๆ เพื่อประกันความมั่นใจว่า Auto8Grade ได้ใช้กระบวนการอันถูกต้องสำหรับการตัดเกรดสำหรับระบบ 8 เกรด

ผลการประเมิน คือ (1) Auto8Grade เป็นโปรแกรมตัดเกรดที่ได้ใช้กระบวนการอันถูกต้องสำหรับการตัดเกรด ครูอาจารย์ผู้สอนสามารถจะตรวจสอบได้จากแผ่นงานทั้งหลายที่ถูกสร้างขึ้น เขาสามารถจะตรวจสอบค่าต่าง ๆ ได้ว่าตรงกับที่ควรจะเป็นตามขั้นตอนต่าง ๆ ในการตัดเกรด เช่นเดียวกับที่อธิบายไว้ในหนังสือ (2) Auto8Grade มีคุณลักษณะสำคัญที่ครูอาจารย์ผู้สอนต้องการ ซึ่งคือความง่ายดายและความยืดหยุ่นในการใช้งาน โดยที่เขาสามารถจะปรับเกณฑ์ขั้นต่ำ เกณฑ์ขั้นสูงของเกรดแต่ละระดับได้อย่างสะดวกรวดเร็ว ก่อนการพัฒนาโปรแกรมตัดเกรดนี้ ผู้พัฒนา Auto8Grade ได้ทำการประเมินความต้องการในเบื้องต้นของผู้ใช้งานแล้วพบว่า 2 สิ่งนั้นคือคุณลักษณะสำคัญที่ผู้ใช้ต้องการ ผลการประเมินในการศึกษาครั้งนี้คือ 3 คุณลักษณะที่ครูอาจารย์ผู้สอนพอใจมากที่สุดสำหรับการใช้งาน Auto8Grade คือ ความง่ายดายในการใช้งาน การให้ค่าสถิติพรรณนาต่าง ๆ และการให้กราฟแสดงการแจกแจงของคะแนน ตามลำดับ (3) เพื่อเพิ่มจำนวนผู้ใช้ Auto8Grade ผู้ประเมินแนะนำให้ผู้พัฒนาหรือผู้ที่เกี่ยวข้องกับการพัฒนา ควรประชาสัมพันธ์ให้ครูอาจารย์ผู้สอนได้รู้จักและใช้โปรแกรมตัดเกรดนี้ และผู้พัฒนาควรจะปรับปรุงให้เป็นโปรแกรมตัดเกรดโดยอัตโนมัติของการตัดเกรดกี่ระดับก็ได้ (Auto n Grade)

[1] Ed. D. (Applied Educational Studies/College Interdisciplinary) จาก Oklahoma State University ประเทศสหรัฐอเมริกา

Description of the Evaluand

An evaluand is thing being evaluated (Patton, 2001). This evaluation was conducted for the Department of Business Administration, Faculty of Management Sciences, Prince of Songkla University (PSU). The department head and the academic staff gave a fund from the department's income to develop Auto8Grade for many teachers who need grading software to help them send their grade reports to the registration unit of the university on time.

In PSU, midterm exam will last 1 week, while final exam will last 2 weeks. After the last day of final exam, teachers have to examine students' exam, collect all students' works and exam scores, transform them to grades and submit all subjects' grade in about 7 days. For teachers who their class size are not big, it maybe no problem to do their jobs in 7 days, but for teachers with big class size, it is rather a rush week for them. Having ease-of use grading software should be a solution from many solutions to help teachers to send their grade reports to the registration unit of the university on time. So Auto8Grade was developed.

The developer of Auto8Grade taught subjects that relevant to both Statistics and software programming. Before developing Auto8Grade, the developer did the need assessment and found that the most important characteristic of grading software that teachers want was ease-of use. However, they need flexible software that they can adjust the boundary of each grade level easily as well. An evaluation was requested to assure that Auto8Grade is accurate grading software that serves the need of the teachers who use 8-grade level system. Moreover there is some attempt to make it generalize to be wide-spread grading software.

Logic Model Description for Auto8Grade Grading Software

A logic model is a model that the program staff develops. It can be used for program evaluation. It starts with the long-term vision of how program participants will be better off because of the program (Fitzpatrick, Sanders, & Worthen, 2004). The style of the logic model representative of the Auto8Grade was chosen due to its lucidity in illustrating the CIPP model of evaluation. The focus of the CIPP model is on the program activities .CIPP is representative of Context, Inputs, Process and Product. Context evaluation assesses needs, assets, and problems within a defined environment. Input evaluation assesses competing strategies and the work plans and budgets of the selected approach. Process evaluations monitor, document, and assess program activities. Product evaluation part is divided into impact, effectiveness, sustainability, and transportability evaluations (Stufflebeam,

2002). All four of these elements were captured in the logic model of the Auto8Grade. The Auto8Grade logic model is presented in the figure below.

The Logic Model

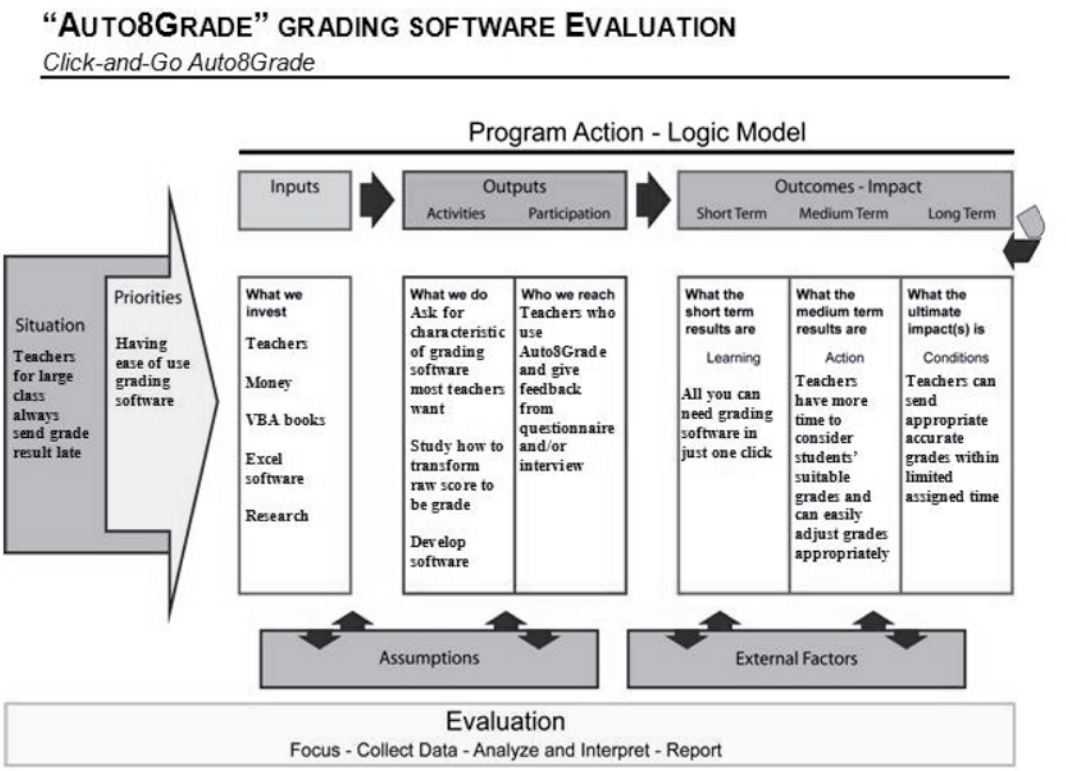

Figure 1 : Logic Model for Auto8Grade grading software

From "Evaluation of Auto8Grade Grading Software," by Aree Limwudhikraijirath, 2010, *Proceedings of the International Conference on e-Commerce, e-Administration, e-Society, e-Education, and e-Technology, 1*, p.1338. Used with permission of the author.

Evaluand Program Goals and Objectives

The evaluator identified the program goals and objectives from the developer's proposal and need assessment report. The completed Auto8Grade must have many crucial characteristics that most teachers need, which composed of :

- Ease-of-use, it can show each student's grade in 8-grade level system in just one click after all input data ware already entered.
- Flexibility, teachers can adjust the boundary for each grade easily and Auto8Grade will change each student's grade due to the

new criteria.

- Descriptive statistics such as mean, standard deviation, maximum and minimum of the students' scores.
- Graph that show the distribution of the scores.
- Standard scores other than raw scores.
- Transparency, everyone can trace back the result that Auto8Grade use the accurate grading process.

Description of Stakeholders

Agents are those persons involved in producing, using, and implementing the program (Fitzpatrick, et al., 2004). Agents involved in the Auto8Grade included :

- Auto8Grade developer
- Teachers
- Head of the Department of Business Administration, PSU and her academic staff Beneficiaries are those persons who benefit from the use of the program (Fitzpatrick, et al., 2004). Beneficiaries involved in the Auto8Grade included :
- Teachers who use 8-grade level system
- Possibly other teachers who want to use Auto8Grade in any grade level system Victims are persons negatively affected by the program (Fitzpatrick, et al., 2004). Victims involved in the Auto8Grade included :
- Teachers who were not able to use Auto8Grade

Evaluation Questions

This evaluation was requested to ensure that Auto8Grade is accurate grading software that has essential characteristics that many teachers want. Moreover, the evaluation included questions about how can Auto8Grade can be used more widely. The following questions were posed as evaluation questions in order to produce the most comprehensive and useful evaluation for Auto8Grade:

- Is Auto8Grade accurate grading software?
- Is Auto8Grade has essential characteristics that teachers want?
- How can Auto8Grade be used more widely?

Description of How Evaluation Questions Were Generated

Evaluations are conducted to answer questions concerning program adoption, continuation, or program improvement (Fitzpatrick, et al., 2004). An evaluation strategy adapted for educational development by Baume and Baume (1995) from Nevo (1986) cited in Macdonald & Wisdom (2002, pp. 9-10) is in the following list :

1. Decide what is or are to be evaluated, and when.

2. Identify stakeholders in the project.

3. Identify stakeholders' questions and concerns.

4. Identify the criteria for judging answers to stakeholders' questions.

5. Devise and pilot the evaluation method and instruments.

6. Carry out the evaluation.

7. Report to the stakeholders.

8. Change project practice as necessary.

9. Review evaluation methods from time to time.

Evaluators' main responsibilities are to gather and interpret information that can help key individuals and groups improve efforts, make enlightened decisions, and provide credible information to the public (Fitzpatrick, et al., 2004). This evaluation was a consumer-oriented evaluation approach. Fitzpatrick, et al., 2004 state that "developers of products have come to realize, however, that using the checklists and criteria of the consumer advocate while the product is being created is the best way to prepare for subsequent public scrutiny" (p. 101). The most widely used methods of collecting data in consumer-oriented evaluations are stringent evaluation criteria and checklists, which provide consumers with defensible results on a wide variety of product factors (Fitzpatrick, et al., 2004). Since the beginning of the evaluation process, the evaluator integrated several sources to come up with effective evaluation questions.

Methodology : Data Collection Procedures

The population for the study included all teachers. Even though the main target population were only teachers who used 8-grade level system, all teachers other than the teachers who used 8-grade level also included because the evaluator need to know how can this software can be used more widely. Only 4 participants were interviewed. The main information got from survey method using 2-pages questionnaires. There were 120 purposive samples. A purposive sample is one that is selected for some particular reason. It is not a random sample (McBurney, 1994). Only 116 questionnaires returned, so the response rate was 96.97 percent.

My research methodology used a triangulation mixed methods design because "by combining multiple observers, theories, methods, and data sources, researchers can hope to overcome the intrinsic bias that comes from single-methods, single-observer, and single-theory studies" (Denzin, 1989 cited in Patton, 2001, p. 555). I also compared both of the qualitative and quantitative data sources for similar results and themes. The research procedures are described herein.

My quantitative procedure included a survey. The survey for user satisfaction was administered in November 2004. The collected-data questionnaires were returned about January 2005. The research assistants (students

in Business Statistics class) helped the evaluator in collecting the questionnaires. Before distributing the questionnaires, the evaluator held a meeting with the research assistants to make sure that all assistants understand in the same concept about all question items, so they can answer any questions that the samples ask. Teachers who were my samples were informed of the purpose of conducting the survey and their participant rights. There were 116 participants agreed to participate and were given a survey, which included an informed consent document. The survey included questions about their satisfaction in Auto8Grade. There were six characteristics asked about that software, which were ease-of-use, comfortable-to-use, flexibility, can be trace back, providing descriptive statistics and providing graph. My qualitative research methods included both informal conversational interviews and existing documents.

There are three alternatives for interviewing which are (1) the informal conversational interview (2) the general interview guide approach, and (3) the standardized open-ended interview (Patton, 2001). I used the informal conversation interview because of its strength, as stated in Patton (2001), "The strength of the informal conversational method resides in the opportunities it offers for flexibility, spontaneity, and responsiveness to individual differences and situational changes" (p. 343). There were four teachers as participants. The questions used in the interview covered the same areas as the survey, but were open-ended in order to gather more information from the participants. The interview for each participant took approximately 15 minutes.

There were three kinds of existing document that the evaluator must investigate to make sure that Auto8Grade is accurate grading software.

(1) Books that describe how to transform raw score to standard T score, and grading system.

(2) Books that teach how to develop application software.

(3) Books that teach how to develop application software using Excel Visual Basic for Application.

Data Analysis

All participants' interviews were recorded later after interview and their responses were analyzed. The survey responses were investigated and analyzed. All of the data was interpreted. Tabulation of results can be found in the Findings section and their interpretation in Conclusion section of this report.

This is a formal written evaluation of the Auto8Grade. The client may chose to present the written findings to teachers who interested in using Auto8Grade, which will make this software be used more widely and the client's organization more well-known.

Table 1 : Evaluation of the Auto8Grade grading software

Evaluation Questions	Methodology	Data Collection
Is Auto8Grade accurate grading software?	Archival – Qualitative	Evaluators investigated existing documents and analyzed findings.
Is Auto8Grade has essential characteristics that teachers want?	Survey – Quantitative Interview – Qualitative	Survey and interview were administered to teachers. The evaluator analyzed findings.
How can Auto8Grade be used more widely?	Survey – Quantitative Interview – Qualitative	Survey and interview were administered to teachers. The evaluator analyzed findings.

Table 2 : Methods used for evaluation data collection

Method	Overall purpose	Advantages	Challenges
Survey	To obtain measurable data within time constraint.	Easily obtained large amounts of data. Very inexpensive Less time consuming and easy to analyze	Systematic (nonsampling error), which is error resulting from some imperfect aspect of the research design that causes response error or from a mistake in the execution of the research; error that comes from such sources as sample bias, mistakes in recording responses, and nonresponses from persons not contacted or refusing to participate.
Interview	To gather qualitative information that cannot directly observe.	Allowed evaluators to gather additional information that could not be captured in the survey. Allows interaction amongst the participants.	Interviewer error, which is administrative error caused by failure of an interviewer to perform tasks correctly.

Table 2 : Methods used for evaluation data collection (Cont.)

Method	Overall purpose	Advantages	Challenges
Archival	To constitute part of the repertoire of field research and evaluation	Documents prove valuable not only because of what can be learned directly from them but also as stimulus for paths of inquiry that can be pursued only through direct observation and interviewing.	Learning to use, study, and understand documents and files is part of the repertoire of skills needed for qualitative inquiry.

(Table adapted from Kelsey, 2005, Patton, 2001 and Zikmund, 2000)

Quantitative Findings

This section contains quantitative findings. Written summaries of the data follow each of the tables.

Table 3 : Teacher demographics N=116

Questions	Frequency (*f*)	Percentage (%)
1. Gender :		
Male	48	41.38
Female	68	58.62
2. Age :		
21-25	9	7.76
26-30	19	16.38
31-35	13	11.21
36-40	9	7.76
41-45	26	22.41
46-50	25	21.55
51-55	10	8.62
56-60	5	4.31
3. Education :		
Below Bachelor	1	0.86
Bachelor	84	72.41
Master	29	25.00
Doctorate	2	1.72

Table 3 : Teacher demographics (Cont.)

Questions	Frequency (*f*)	Percentage (%)
4. Major :		
Business Administration	24	20.69
Education and Educational Administration	34	29.31
Engineering, Science and Technology	33	28.45
Languages	14	12.07
Public Administration	3	2.59
Sociology	8	6.90
5. Public or private institute :		
Public	96	82.76
Private	20	17.24
6. Level of education that your institute provided :		
Kindergarten	19	9.69
Grade 1-6	63	32.14
Grade 7-9	37	18.88
Grade 10-12	41	20.92
Vocational education	11	5.61
University	23	11.73
Other (Work in office 1, Education evaluator 1)	2	1.02
(Can reply more than one answer)***		

The evaluator administered the survey questionnaire for user satisfaction in the Auto8Grade in November 2004. The collected-data questionnaires were returned about January 2005. There were a total 116 participants who responded to the survey. From the survey findings, there were a little bit male less than female (41.38% and 58.62% respectively). Most participants are senior (about 44% were 41 to 50 years old). Moreover, most of them got bachelor and master degree (72.41% and 25% respectively). They were majoring in different fields such as Education and Educational Administration, Engineering, Science and Technology, and Business Administration (29.31%, 28.45%, and 20.69% respectively). The majority of them (82.76%) worked in public educational institute. Moreover, most of them worked in the institutes that provide teaching in grade 1-6, grade 10-12, and grade 7-9 (32.14%, 20.92%, and 18.88% respectively).

Table 4 : User satisfaction after using Auto8Grade grading software

	Degree of Satisfaction					
	Very unsatisfactory	Unsatisfactory but usable	Not very satisfactory	Satisfactory	Quite satisfactory	Total *f* (%)
Ease-of-use. Provide students' grade in just one click.	-	1 (0.86%)	6 (5.17%)	70 (60.34%)	39 (33.62%)	**116 (100%)**
Comfortable-to-use. Provide students' grade that can be used immediately.	-	3 (2.59%)	17 (14.66%)	61 (52.59%)	35 (30.17%)	**116 (100%)**
Flexibility. Teachers can adjust grade boundaries easily.	1 (0.86%)	4 (3.45%)	25 (21.55%)	61 (52.59%)	25 (21.55%)	**116 (100%)**
Everyone can trace back. Show detail about transforming raw scores to be standard scores.	-	2 (1.72%)	21 (18.10%)	58 (50.00%)	35 (30.17%)	**116 (100%)**
Provide descriptive statistics.	-	2 (1.72%)	15 (12.93%)	63 (54.31%)	36 (31.03%)	**116 (100%)**
Provide graph that shows score distribution.	-	4 (3.45%)	16 (13.79%)	56 (48.28%)	40 (34.48%)	**116 (100%)**

The finding shows that a majority of them (60.34%, 54.31%, 52.59%, 52.59%, 50.00%, and 48.28%) satisfy all characteristics of Auto8Grade (ease-of-use, provide statistics, comfortable, flexibility, trace back, and provide graph respectively). There was a considerable remark that one participant was very unsatisfied in flexibility. It can be used as a request that the developer may improve the software to be more flexible. However, the crucial characteristic of this software that users satisfy still be ease-of-use, the same result as appeared in the need assessment before the software development.

Table 5 : Descriptive Statistics

	Descriptive Statistics					
	N	Minimum	Maximum	Mean	Standard Deviation	Meaning
Number of years he/she taught	116	1	37	16.18	10.20	more experience in teaching
Ease-of-use. Provide students' grade in just one click.	116	2	5	4.26	0.60	satisfactory
Comfortable-to-use. Provide students' grade that can be used immediately.	116	2	5	4.10	0.74	satisfactory
Flexibility. Teachers can adjust grade boundaries easily.	116	1	5	3.91	0.80	satisfactory
Everyone can trace back. Show detail about transforming raw scores to be standard scores.	116	2	5	4.09	0.74	satisfactory
Provide descriptive statistics.	116	2	5	4.15	0.70	satisfactory
Provide graph that shows score distribution.	116	2	5	4.14	0.78	satisfactory

The teaching experience of participants ranges from 1 to 37 years, with average of 16.18 years and standard deviation 10.20. The participants satisfy in ease-of-use, provide statistics, provide graph, comfortable, can trace back and flexibility with average of 4.26, 4.15, 4.14, 4.10, 4.09, and 3.91 respectively.

The significant thing that should not be overlooked is goodness of measures. Sekaran (2003) stated that "it is important to make sure that the instrument that we develop to measure a particular concept is indeed accurately measuring the variable, and that in fact, we are actually measuring the concept that we set out to measure" (p. 202). It is also stated in Wiersma & Jurs (1990) that "validity is the extent to which a test measures what it is intended to measure" (p. 183) and "reliability of measurement is

consistency – consistency in measuring whatever the instrument is measuring" (p. 155).

From the above items, Cronbach's alpha is 0.842, so it is reliable. Cangelosi (1990) stated about reliability that "A test is reliable to the same degree that it can be depended on to yield consistent, noncontradictory results" (p. 29). The concept of this user satisfaction survey to find the result that Auto8Grade is an ease-of-use software that teachers need as said in the need assessment or not. The content validity of this measure can be accepted easily. Content validity ensures that the measure includes a sufficient and representative set of items that tap the concept (Sekaran, 2003).

Following is the findings extracted from open-ended question about their benefit using Auto8Grade. The same meaning contents are already grouped together.

1. Comfortable and fast.
2. Comfortable with large class, and fast.
3. Comfortable and more precision in calculation.
4. Decrease time in grading.
5. Comfortable, decrease error, and have recorded data to be used well.
6. Comfortable, and can be trace back.
7. Comfortable, and help me submitted grade on time.
8. It is ease-of-use, and modern.
9. Fast, accurate, and more precision.
10. Help teachers in grading more easily.
11. Decrease time in grading.
12. Can keep data in secret.
13. It is ease-of-use, can be understood easily not complicated.
14. Can adjust data in a short time.
15. Can be used in internal quality control.
16. Can adjust criteria, comfortable and fast.
17. I gain benefit in evaluation and assessment, scoring, and grading.
18. Comfortable, and there is standard in using it.
19. Help in scoring students' test.
20. It's a good program for teachers.
21. I used it to calculate descriptive statistics, and used it in final examination.
22. I used it in teaching improvement; I can know strength and weak point of my students.
23. Use the same standard in grading.
24. It is ease-of-use; users need not to be experts in computer programming.
25. It is very useful.
26. Comfortable, and high efficient.
27. It provide graph for comparing.
28. A little bit complicate.
29. It shows accurate descriptive statistics, ease-of-use, and appropriate to be

used.

30. Ease-of use, and more accurate than I calculate it myself.

31. Comfortable, and easy for analyzing.

Following is the findings extracted from open-ended question about any other comments or suggestions about Auto8Grade. The same meaning comments are already grouped together.

1. It should be used for large class students.

2. It will be better if the developer use numeric grade (1, 2, 3…) instead of alpha grade (A, B, C…).

3. It should be grade 1-4.

4. I need more descriptive statistics.

5. It should be instant to be submitted to the register unit immediately.

6. I would like to applause for the idea that teacher can adjust criteria for grading.

7. It's flexible and allows teachers to use their judgment.

8. The developer should distribute this software to all schools.

9. The developer should develop more new computer program.

10. I would like to encourage the developer to do good thing for social and country development.

11. It should be developed to be used with old version of Microsoft Office such as Office 97.

12. Should allow grading from raw score, because we use criterion-reference in secondary school.

13. Should be more good- looking. Should tell more detail in using the software.

14. Should have wizard for beginner.

15. Complicate method.

16. Should provide grade for elementary or secondary school that is 0, 1, 2, 3, 4.

17. Should train all teachers in schools.

18. Should have description in diskette.

19. Should have more detail.

20. Should release this program to all institutes.

Qualitative Findings for Evaluation Question 1 : Is Auto8Grade accurate grading software?

There were 4 participants in interview. One was male and three were female. From interview, only 1 of 4 said that she would compare with the old software she used that Auto8Grade give the same result from the old one. The other three interviewers did not interested in trace back the formula for transforming raw scores to be standard scores and the final grades, however they liked that software and they would use this software to help them in grading students' achievement in the future.

From existing documents, there were contents that correspond with the formula used in Auto8Grade. The following is the explanation

about z-score.

The simplest of the standard scores, and the one on which others are based, is the z-score. This score expresses test performance simply and directly as the number of standard deviation units a raw score is above or below the mean (Gronlund & Linn, 1990, p. 352).

Another well-known standard score is T-score.

The term T-score was originally given to a type of normalized score based on a group of unselected twelve-year-old children. However, it has come to refer to any set of normally distributed standard scores that has a mean of 50 and a standard deviation of 10. T-scores (linear conversion) can be obtained by multiplying the z-score by 10 and adding the product to 50. Thus, T score = 50 + 10 (z). This formula will provide T-scores only when the original distribution of raw scores is normal because with this type of conversion (linear), the distribution of standard scores retains the same shape as the original raw score distribution (Gronlund & Linn, 1990, p. 352).

The standard score used in Auto8Grade is normalized T-score.

Normalized T-scores are calculated by (1) converting the distribution of raw scores into percentile ranks, (2) looking up the z-score each percentile rank would have in a normal distribution and assigning these z-scores to the corresponding raw scores, and (3) converting the z-scores to T-scores, using the formula presented earlier. The procedure of going from raw score to percentile to the corresponding z-score in a normal distribution is called an area conversion and yields normalized z-scores that are then transformed directly into normalized T-scores. This results in a normally distributed set of standard scores, regardless of the shape of the original distribution of raw scores. Normalizing is used by test publishers to remove minor irregularities in the raw score distributions (Gronlund & Linn, 1990, p. 354).

Qualitative Findings for Evaluation Question 2 : Is Auto8Grade Grading Software has essential characteristics that teachers want?

All participants in interview preferred using Auto8Grade. They satisfied in ease-of use that Auto8Grade provided students' grade in just one click (even though they must enter short password "psu" telling on the popup window. The developer received fund from the Department of Business Administration and distributed this software with no charge, so it is appropriate to show something to tell the source of money for this software development). The other favorite characteristics were software capability in providing descriptive statistics and graph of score distribution. Most participants did not interest in adjusting grade boundaries

which will provide new grades immediately.

Even though all participants said that Auto8Grade can be used easily, one of them said that for her subject, she still wanted to use raw score. She preferred criterion-referenced than norm-referenced measurement. Her opinion was concurred with some suggestions extracted from the survey, which said that they still use raw score and criterion-referenced measurement.

Anyone who is interested in using Auto8Grade can download it from http://vclass.mgt.psu.ac.th/~laree/Auto8Grade.xls

The user manual (in Thai) can be downloaded from http://vclass.mgt.psu.ac.th/~laree/Auto8GradeManual.pdf

However, there is a short description telling how to use Auto8Grade at the first view of it. Anyone who uses Microsoft Excel for collecting data of their students can use Auto8Grade easily by clicking at the button "Auto8Grade Click here". The Auto8Grade first view is presented in the figure below.

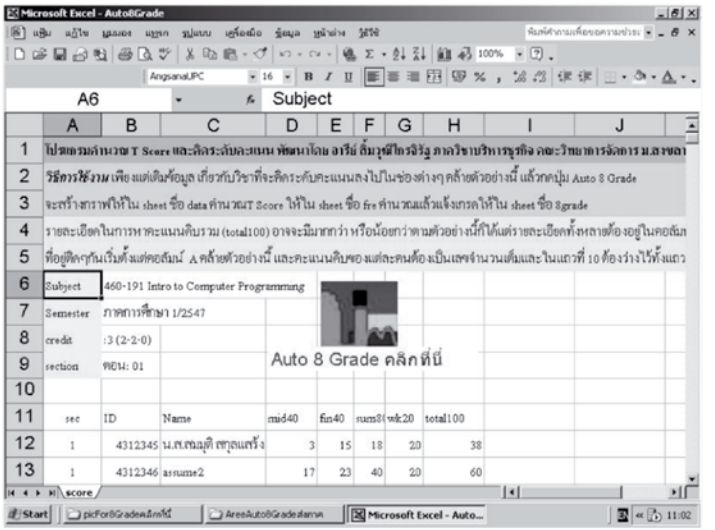

Figure 2 : The first view of Auto8Grade grading software

Qualitative Findings for Evaluation Question 3: How Auto8Grade Grading Software can be used more widely?

From open-ended questions and from interview, the evaluator recommends the following :

- If ease-of-use grading software for 8-grade level system still be a most important characteristic, the first strategy to promote this software is distributing it to other institutes.
- The developer may improve the software until it can tell both letter and numeric

grade. Moreover, it can be used for any grade level other than only 8-grade level. However, it will be an attempt to generalize this software, and there will be many dialog boxes in each step to ask users to enter their preferences which will make it more complicate to use.

- The developer may use Open Source software for the next version. The less cost, the more users. However, the appropriate cost considered should be overall cost, not only cost for software itself. This should be an interesting topic for further research.

Evaluative Judgments, Conclusions and Recommendations for Objective 1 : Is Auto8Grade accurate grading software?

At the present time, and according to my research, I conclude Auto8Grade is accurate grading software. The evaluation findings documented that Auto8Grade created the same grade result of the software some teachers used in the past, and it showed detail of calculation that teachers can trace back to make sure that it uses the correct formula and grading procedure.

Teachers who use 8-grade level system and teach in large class should try to use Auto8Grade as a tool for grading. Teachers who want to grade students in less than 8 levels still can use this software with a little bit adjustment. Teachers who preferred criterion-referenced than norm-referenced measurement still can use Auto8Grade to keep eyes on students' learning by looking at their descriptive statistics and graph.

Evaluative Judgments, Conclusions and Recommendations for Objective 2: Is Auto8Grade Grading Software has essential characteristics that teachers want?

The evaluation process revealed that Auto8Grade has essential characteristics that teachers want. The following characteristics are sorted from the most characteristic teachers want to the least one : ease-of-use, provide statistics, provide graph, comfortable, can trace back, and flexibility.

If the developer uses ease-of-use as the most crucial characteristic of grading software, no changes need for this software. However, to serve the teachers' need in numeric grade instead of letter grade, the developer should develop another version that reports the numeric grade.

Evaluative Judgments, Conclusions and Recommendations for Objective 3 : How Auto8Grade Grading Software can be used more widely?

The majority of the participants prefer ease-of-use characteristic of Auto8Grade. The developer should distribute this software to other institutions to make it well-known software for grading.

I recommend the following:

- The developer should create many new versions as derivatives of Auto8Grade and distribute them to other institutes with free of charge.

- The Department of Business Administration, Faculty of Management Sciences, Prince of Songkla University, as a client, should promote this software through many kinds of channel, such as on the web site of the faculty, or by setting some seminars for PR (public relation) this software.

REFERENCES

Cangelosi, J.S. (1990). **Designing tests for evaluating student achievement.** White Plains, NY : Longman.

Fitzpatrick, J.L., Sanders, J.R., & Worthen, B.R. (2004). *Program evaluation: Alternative approaches and practical guidelines*. (3rd ed.). Boston, **Massachusetts: Pearson Education,** Inc.

Gronlund, N.E., & Linn, R.L. (1990). **Measurement and evaluation in teaching.** (6th ed.). Belmont, CA : Brooks/Cole Publishing Company.

Kelsey, K. (2005). **Evaluating distance education programs using best practices.** In K. Dooley, J. Lindner, & L. Dooley (Eds.), *Advanced methods in distance education : Applications and practices for educators, trainers, and learners*. Hershey, PA : Idea Group.

Limwudhikraijirath, A. (2010). Evaluation of Auto8Grade Grading Software. In T. Hung, & C. Li (Eds.), **Proceedings of the International Conference on e-Commerce, e-Administration, e-Society, e-Education, and e-Technology, Macau,** *1*, 1336-1358.

Macdonald, R., & Wisdom, J. (2002). **Academic and educational development : research, evaluation and changing practice in higher education.** Srerling, VA : Stylus Publishing, Inc.

McBurney, D.H. (1994). **Research methods.** (3rd ed.). Belmont, CA : Brooks/Cole Publishing Company.

Patton, M.Q. (2001). **Qualitative research and evaluation methods.** (3rd ed.). Thousand Oaks, CA : Sage Publications, Inc.

Sekaran, U. (2003). **Research methods for business: A skill building approach.** (4th ed.). Third Avenue, NY : John Wiley & Sons, Inc.

Stufflebeam, D.L. (2002). **CIPP Evaluation model checklist : A tool for applying the Fifth Installment of the CIPP Model to assess long-term enterprises.** Retrieved July 1, 2007, from http://www.wmich.edu/evalctr/checklists/cippchecklist.htm

Wiersma, W., & Jurs, S.G. (1990). **Educational measurement and testing.** (2nd ed.). Needham Heights, MA : Allyn and Bacon.

Zikmund, W.G. (2000). **Business research methods.** (6th ed.).Mason, OH : Thomson Learning.