

USING DELPHI TECHNIQUE IN META-EVALUATING THE MATHEMATICS ASSESSMENT PRACTICES OF MATHEMATICS EDUCATORS

Mario C. Oli

Cagayan State University

Carig, Tuguegarao City, Philippines

E-mail Address: mariooli696@yahoo.com

ABSTRACT

While assessment practices are very important into the students' learning and development, the processes of assessment may follow a logical progression from the selection and development of classroom assessment to the communication of the assessment results. Teachers may have similar assessment practices but they vary on the processes in conducting the assessment. Through Delphi Technique, evaluation experts assessed the extent to which mathematics assessment practices satisfy meta-evaluation criteria of utility, feasibility, propriety, accuracy and accountability employed by Mathematics Educators of pre-professional mathematics teachers in some select State Universities in the Philippines. The assessment practices of the Mathematics Educators were meta-evaluated with high ratings on utility, feasibility, propriety, and accuracy, while only moderate in accountability. Assessments performed in the largest state university have better overall utility, accuracy and accountability with overall feasibility and propriety about the same level across the four state universities. Hence, the respondents should develop assessment strategies for students with different learning abilities, continuously improve their technique in assessing students' learning and have sound judgement not only through the students' quantitative scores but the impact of feedback about their performance for future use.

Keywords: utility, feasibility, propriety, accuracy, accountability

Quality pre-service teacher education is a key factor to achieve quality Philippine education (CMO 30, s. 2004). Efforts have been made to improve the quality of teacher education in the country because of its dependence on the service of teachers who are properly prepared to undertake the different important roles and functions of classroom teachers. Thus, it is imperative that the highest standards should be set in defining the objectives, components, and processes of the pre-service teacher education curriculum.

In pre-service teacher preparation, Feuer, Floden, Chudowsky, and Ahn (2013) believed that quality of instruction greatly contributes to students' learning process. They expounded on the need to have a record from observations of teaching for it measures quality of feedback from mentors and assesses whether they are applying what they have learned during the preparation stage.

According to the Joint Committee on Standards for Educational Evaluation (JCSEE, 2013), practices and processes of assessment serve as bases in determining the progress of students as planned and in effectively planning for students' future learning opportunities. While assessment practices are very important into the students' learning and development, the processes of assessment may follow a logical progression from the selection and development of classroom assessment to the

communication of the assessment results (NCTM, 1995 & JCSEE, 2013). Teachers may have similar assessment practices but they vary on the processes in conducting the assessment. Different individuals have different views of the concepts being mentioned. Through the differences, quality standards may be difficult to establish. Thus, credibility, fairness and utility may be sacrificed depriving the rights of the students of having such. However, NCTM (1995) presents four interrelated phases in the assessment processes which are the planning assessment, gathering evidence, interpreting evidence, and using results. Since these phases are interactive, differences between them could hardly be determined and that they should not be seen as necessarily sequential. The same applies to the assessment practices of teachers in classroom mathematics.

The Concept of Meta-evaluation

The reliability of assessment practices and process as patterned in the national and international standards of assessment and evaluation pose a question on this side. According to Stufflebeam (2001), the works of educators need to be further evaluated to ensure the presence of utility, feasibility, propriety, accuracy and accountability of their output (Stufflebeam, 2012). Giving assessment and the process of administering it take crucial part in the development of the learners. Such dimensions are the components or standard checklists for final and summative meta-evaluations organized according to the Joint Committee on Program Evaluation Standards.

There are five standards of metaevaluation that involved in this study. These are utility, feasibility, propriety, accuracy and accountability. First, the utility standards aimed to increase the extent to find assessment practices and processes valuable (Sharifi, A. & Hassaskhah, J., 2011) in providing the needs of the stakeholders. It covers the following sub-criteria standards such as evaluator's credibility, attention to stakeholders, negotiated purposes, explicit values, relevant information, meaningful practices and processes, timely and appropriate communicating and reporting of results and concern for consequence and influence. Second, the feasibility standards are intended to increase evaluation effectiveness and efficiency. With this standards, it includes project management, practical procedures, contextual validity and resource use as subcriteria of evaluations. Third, the propriety standards support what is proper, fair . legal, right and just in evaluations. Among others, responsive and inclusive orientation, formal agreements, human rights and respect, clarity and fairness, transparency and disclosure, conflicts of interests and fiscal responsibility are its sub-criteria of evaluations. Fourth, the accuracy standards are intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings especially those that support interpretations and judgements about quality. Its sub-criteria of evaluations are justified conclusions and decisions, valid information, reliable information, explicit program and context descriptions, information management, sound design and analyses, explicit evaluation reasoning, and communication and reporting to avoid misconceptions, biases, distortions, and errors of information. Lastly, the accountability standards encourage adequate documentation of evaluations and the perspective of meta-evaluation focused on the improvement and accountability of evaluation and products. It is concerned on the document evaluation, internal and external meta-evaluation.

Stufflebeam (2012) and Scriven's (1969) have common knowledge about meta-evaluation, that is, "evaluation of evaluation". Moreover, Stufflebeam (1974) earlier explained that it is a procedure for describing an evaluation activity and judging it against a set of ideas concerning what constitutes good evaluation. He further expounded that it is also a the process of delineating, obtaining, and applying descriptive information and judgmental information about an evaluation's utility, feasibility, propriety, and accuracy and its systematic nature, competence, integrity/honesty, respectfulness, and social responsibility to guide the evaluation and publicly report its strengths and weaknesses.

In 2009, Scriven has simplified his definition of meta-evaluation but is now more explicit. He defined it as the consultant's version of peer review, i.e. doing their assessment work and submitting the results directly to the client or other audience. Moreover, comments on the output given by experts do not manifest weakness, rather a recognition that an independent expert's look at one's work usually generates insights for its improvement.

The study was conducted to initiate the process of meta-evaluation on the assessment practices among mathematics educators. Through literature review, there has been no study found similar to the present study. This means that practices and processes of assessment employed by teachers have never been explored based on the standards of meta evaluation. Hence, the purpose of this study.

This study aimed to evaluate the assessment practices of content faculty, student-teaching supervisors and cooperating mentors of pre-professional mathematics teachers in State Universities in the Cagayan Valley Region. Also, it attempted to answer the extent assessment practices satisfy the following meta- evaluation criteria: utility, feasibility, propriety, accuracy and accountability (Stufflebeam, 2012 and JSCEE, 2012; finding the difference in meta-evaluation of assessment practices by the content faculty, student-teaching supervisors and cooperating mentors across meta-evaluation criteris and state universities.

METHODS

This study employed qualitative-descriptive and quantitative-comparative research designs. It was conducted in four different State Universities in Northeastern Philippines offering the course Bachelor in Secondary Education major in mathematics and to select Secondary Schools in the Department of Education affiliated with the State Universities because of the functions of their Math teachers as Cooperating Mentors to these pre-professional math teachers relative to their training and development as future math teachers.

The sampling technique utilized in this study was purposive and quota sampling. Content faculty, who had a class with the pre-professional math teachers in one of the major subjects in mathematics during the first semester of the Academic year 2014-2015, Student-teaching supervisors and Cooperating mentors of the four state universities were the main subjects of this research. The Student-teaching supervisors were the College Instructors designated to do the transactions concerning the deployment and monitoring the performances of the pre-service math teachers in their off- campus experience (practicum) while the Cooperating mentors are

the Secondary math teachers from the Department of Education who were given pre-service math teachers to assist and guide them in the duration of their practicum.

Table 1 shows the number of faculty teaching major subjects in mathematics to the BSE-math major students, the number of student-teaching supervisor in the program and ratio of cooperating mentors to pre-professional math teachers who will be in the Practice Teaching course. It is seen in the table that CSU and ISU had same number of content faculty which is 21.43% while the other two had 28.57% of the total number of content faculty. On the other hand, the average ratio of pre-professional math teachers to the cooperating mentors is one-to-one.

Table 1. Number of Math Faculty, Student-teaching Supervisors and Cooperating mentors of Pre-Professional Math Teachers

Educator	SUC	SUC	SUC	SUC	TOTAL
Content	4	3	4	3	14
Student-	1	1	1	1	4
Cooperating	5	13	9	27	72

Table 2 shows the number of cooperating mentors for student teachers in every State University. As the table provides, out of 37 Cooperating mentors of SUC 4, 27 of them were considered, others have served to validate the instruments. However, the largest percentage came from SUC 3 with 9 out of 12 mentors. SUC 1 has the smallest number of Mentors considered (i.e. 5 out of 7) during the time of data gathering. One of its pre-service math teachers was assigned in the High School Laboratory of the University and another mentor has been given assignment by the Division of Quirino outside the School.

Table 2: Number of Cooperating mentors

SU	Total Number	Cooperating Mentors	%
SUC 1	7	5	71.43%
SUC 2	18	13	72.22%
SUC 3	12	9	75.00%
SUC 4	37	27	72.97%
Total	74	54	72.97%

The instrument used in gathering the data for this study was the abridged meta-evaluation checklist. The meta-evaluation checklist consisted of five major standards: utility, feasibility, propriety, accuracy, and accountability. The original instrument underwent factor analysis and its reliability coefficient (0.932) was highly considered.

The study was conducted through a) personal semi-structured interview and b) process of meta-evaluation. In the interview on the assessment practices, the implementation of each practice was initially asked. General idea or concept of assessment from each respondent was then solicited. A video-camera was used to

capture the interview. The interviews were transcribed for later analysis. English translations of interview transcripts in the vernacular/dialect were slightly modified in grammar and in sentence structure to present the respondents' thoughts and ideas in more coherent manner.

The meta-evaluation process was the key process in this study. Four professionals were invited to do the meta-evaluation because of their expertise in evaluation and assessment. Using the Delphi Technique, experts were given the videos and transcripts of the interview for them to evaluate using the abridged meta-evaluation checklist. Discussions were made regarding the assessment process and the statements in the instrument. The transcriptions being evaluated by these professionals were sealed in different envelopes. The sequence and presentation of the transcriptions were made similar. The meta-evaluators were synchronized about the group of respondents to be meta-evaluated.

All quantitative data gathered were entered into the Microsoft Excell and analysed using a statistical software. Descriptive statistics which include frequencies and percent, and standard deviation, were used, wherever appropriate to describe the practices and adherence to the criteria of the standards. Inferential statistics such as repeated measures analysis of variance (RMANOVA) was used to determine significant differences in ratings on the various meta-evaluation criteria and practices (Mauchly's $W = 0.799$, p -value=.076). One-way analysis of variance was utilized to determine significant differences in meta-evaluation criteria and practices when the assessments are grouped by State University. Least significant difference (LSD) was used for post-hoc pair-wise comparisons. Statistical hypotheses were tested at significance level of 5%. Also, responses in the interview were categorized according to the criteria of meta-evaluation.

RESULTS

1. Extent of the assessment practices and processes satisfying the following Stufflebeam's and JCSEE meta-evaluation criteria: utility, feasibility, propriety, accuracy and accountability.

A. Utility standards

The assessors were rated high as regards their competence being reflected in the designation assigned to them per Civil Service Commission (CSC) and CHED memoranda, giving of immediate feedback to the students concerned, issuance of brief, simple and direct reports to concerned individual, and describing the purpose of assessment or evaluation, procedures and results. Generally, the State Universities were rated to have high utility standards regarding their implemented assessment practices and processes.

B. Feasibility standards

The promptness of the Content faculty, Student teaching supervisor and Cooperating mentors in addressing evaluation results to concerned individuals, their implementation of assessment practices that others are carrying out, being realistic in scheduling of assessment or evaluation, making evaluation or assessment procedures a part of routine events and providing information on responsible use of resources to produce results are high. Generally, the total mean feasibility rating for Mathematics Content faculty, Student-teaching Supervisors and Cooperating Mentors or the

effectiveness and efficiency and the assurance that assessment practices and processes are realistic, prudent, diplomatic and frugal is also high.

C. Propriety standards

The Mathematics Content faculty, Student teaching supervisors and Cooperating mentors are high in terms of promoting excellent service in assessment, explaining the assessment procedures to be implemented by the evaluators to concerned individuals, making clear to stakeholders that the evaluation will respect and protect the rights of the concerned individuals, explaining the intended purposes of the evaluation, showing respect to individual difference, keeping concerned individuals informed of the evaluation or assessment result, reporting to concerned individual his or her strengths as provided by the result of evaluation, reporting to concerned individual his or her weaknesses as provided by the result of evaluation, providing a thorough explanation of the assessment process and explaining to concerned individual his/her strengths could be used to overcome his /her weaknesses. Generally, they highly support what is proper, fair, legal, right and just in evaluations.

D. Accuracy standards

The accuracy of Content faculty, Student teaching Supervisors and Cooperating mentors were rated high in terms of reflecting the evaluation procedures and findings, focusing the evaluation on goals and objectives of the program, explaining or documenting how information from each procedure was scored, analyzed, and interpreted, obtaining information from variety of sources, employing a variety of data collection methods, checking systematically the accuracy of scoring, explaining the assessment processes to the concerned individuals to ensure fair and impartial reports and referring to colleagues of the purposes of evaluation or assessment. They are moderately accurate in citing evidence supporting each conclusion, choosing assessment instruments that have shown acceptable levels of reliability for their intended uses, reporting the factors that influenced the reliability, including the characteristics of the examinees, the data collection conditions and the assessor's biases, justifying the means used to obtain information from each source, and using multiple evaluators and checking the consistency of output.

E. Accountability standards

The assessors are moderately accountable in asking their colleagues of the assessment or evaluation design which they found effective, collaborating with fellow evaluators as regards assessment or evaluation procedures, and constructing scoring rubrics with individuals concerned in assessing outputs. However, they are highly accountable in terms of recording all data collected and recording analyzed data and outcomes, analyzing discrepancies between intended purposes and procedures and those which actually took place during the assessment, employing both formative and summative evaluation of assessment, determining from the record which audiences will receive the report on evaluation of assessment, evaluating the instrumentation, data collection, data handling, and analysis against the relevant standards, evaluating evaluator's involvement of and giving of feedbacks to concerned individuals to improve the performance of the students (Mann, 2004) and in maintaining a record of all steps, information, and analyses of evaluation of assessment. Generally, the State Universities have a moderate accountability as regard concerns on the adequacy of

documentation for evaluations and a meta-evaluative perspective focused on improvement of learning and for assessment processes and outputs.

2. Comparison in the meta-evaluation of assessment practices and processes by content faculty, student teaching supervisors and cooperating mentors across the Meta-evaluation criteria and across State Universities.

Meta-evaluation of assessment practices and processes across Meta- evaluation criteria.

All of the State Universities have high standards on the meta- evaluation criteria as indicated by their individual means except for Accountability standards. They were rated high in terms of finding assessment practices and processes valuable in meeting the needs of the intended users, effectiveness and efficiency of implemented assessment practices and processes which ensures realistic, prudent, diplomatic and frugal, supporting what is proper, fair, legal, right and just in assessment practices and processes, and focusing on dependability and truthfulness of assessment representations and propositions and findings which support interpretations and judgments on the quality of assessment practices and processes. However, they were moderate in terms holding themselves accountable of the sufficiency of documentation for evaluation and meta-evaluative perspective focused on learning improvement.

Comparison in the meta-evaluation of assessment practices and processes by content faculty, student teaching supervisors and cooperating mentors across State Universities.

a. Utility standards

Only the Content Faculty, Student-teaching Supervisors and the Cooperating Mentors of the State University 4 have very high ratings in terms of their competence as reflected in their designation assigned them by authorities. However, the Content faculty and Student-teaching Supervisors of the four State Universities (SU) and Cooperating Mentors were generally high in terms of finding the assessment practices and processes valuable in meeting the needs of the intended users

b. Feasibility standards

The Content faculty, Student teaching Supervisors and Cooperating mentors of the Pre-service Math teachers were high as regards their promptness in addressing evaluation results to concerned individuals, implementing assessment practices that others are carrying out, scheduling assessment or evaluation realistically, making evaluation or assessment procedures a part of routine events, and providing information on responsible use of resources to produce result. Hence, they are high in terms of effectiveness and efficiency of the implemented assessment practices and processes which ensures realistic, prudent, diplomatic and frugal evaluation. Generally, The Feasibility standard of the State Universities is generally high.

c. Propriety standards

The State Universities are high in terms of promoting excellent service in assessment, explaining the assessment procedures to be implemented by the evaluators to concerned individuals, making clear to stakeholders that the evaluation will respect and protect the rights of the concerned individuals, explaining the intended purposes of the evaluation, shows respect to individual differences, keeping concerned individuals informed of the evaluation or assessment result, reporting to concerned individual his or her strengths as provided by the result of evaluation, reporting to concerned individual his or her weaknesses as provided by the result of evaluation, providing a thorough explanation of the assessment process, and explaining to concerned individual his/her strengths could be used to overcome his /her weaknesses. Generally, their support to what is proper, fair, legal, right and just assessment practices and processes is high.

d. Accuracy standards

The State Universities are high in terms of accurately reflecting accurately the evaluation procedures and findings, focusing the evaluation on goals and objectives of the program, explaining or documenting how information from each procedure was scored, analysed, and interpreted, in documenting reliability of an instrument, reports the factors that influenced the reliability, including the characteristics of the examinees, the data collection conditions, and the evaluator's biases, obtaining information from a variety of sources, employing a variety of data collection methods (if appropriate), checking systematically the accuracy of scoring, and explaining the assessment processes to the concerned individuals to ensure fair and impartial reports.

However, they are moderate in citing the evidence that supports each conclusion, choosing assessment instruments that in the past have shown acceptable levels of reliability for their intended uses, justifying in the documentation the means used to obtain information from each source, and using multiple evaluators and check the consistency of their work.

In general, the State Universities were rated high in terms of their dependability and truthfulness of assessment representations, propositions, and findings especially those that support interpretations and judgments about the quality of assessment practices and processes is high.

e. Accountability standards

The State Universities were rated high in terms of recording fully all data collected, recording analysed the data and outcomes, analysing discrepancies between intended purposes and procedures and those which actually took place during the evaluation, employing both formative and summative evaluation of assessment, determining from the record which audiences will receive the report on evaluation of assessment, evaluating the instrumentation, data collection, data handling, and analysis against the relevant standard, evaluating the evaluator's involvement of and giving of feedbacks to concerned individuals against the relevant standards, and maintaining a record of all steps, information, and analyses of evaluation of assessment.

However, they are moderately accountable in terms of referring to colleagues of the purposes of evaluation or assessment, asking colleagues of the assessment or evaluation design which they found effective, collaborating with fellow evaluators as regards assessment or evaluation procedures, and constructing scoring rubrics with individuals concerned in assessing outputs. Generally, the State Universities are moderately accountable in the assessment practices and processes they have implemented.

3. Differences on the Meta-evaluation of assessment practices across State Universities

The utility ratings among evaluators from different SUs are significantly different in terms of their competence, giving of feedbacks and giving of results. However, they do not differ significantly in describing the assessment's purpose, procedures and results. Evaluators from different SUs vary significantly in terms of the indicators in the Utility standard. The assessment competence of the evaluators from SU 4 is significantly better as compared with the evaluators of SU 2 and SU 3. In giving feedbacks of results of assessments to students concerned and issuing of brief, simple and direct reports to concerned individuals, SU 4 is significantly rated better than SU 1. However, there is no significant difference among the evaluators of four State Universities as regards describing the purpose of assessment or evaluation, procedures and results. Overall, the utility of assessors in SU4 are significantly rated better compared to their counterparts in SU 1 and SU 2.

On the average, the evaluators from SU 4 are rated significantly better in feasibility compared to evaluators from SU 1 and SU 2 in their implementation of assessment practices that others are using in their campus. As regards the other indicators of the feasibility standard, the evaluators from the four state universities are not significantly different. Hence, they do not significantly differ in their promptness in addressing evaluation results to concerned individuals, in providing information on responsible use of resources to produce results, in their realistic assessment scheduling and in routine assessment activity. Overall, the evaluators from the four state universities do not differ significantly in their feasibility.

Mean differences in ratings in all the indicators of the propriety standard are not significantly different across the four state universities, except in respecting and protecting the rights of human subjects and reporting of assessment results. The mean propriety rating of the evaluators in SU 4 in terms making clear to stakeholders that the evaluation will respect and protect their rights as humans is significantly better than those of the evaluators from SU 2 and SU 3. Similarly, evaluators from SU 4 were rated better than SU 1 and SU 2 in reporting to concerned individuals their strengths and weaknesses as provided by the result of evaluation. But, overall, the propriety of assessment is not significantly different across the four state universities.

The accuracy ratings of the assessment practices of the evaluators in the four SUs are different significantly in terms of explaining how information from each procedure was scored analysed and interpreted, in choosing assessment instruments that have shown acceptable levels of reliability, in justifying the means used in obtaining information from each source, in checking systematically the accuracy of scoring, and in explaining the assessment processes to the concerned individuals to

ensure fair and impartial report. In these accuracy indicators, evaluators from SU 4 were rated significantly better.

The evaluators of SU 4 were found to have significantly better ratings along overall accountability. Likewise, SU 4 evaluators were rated significantly higher in the construction of scoring rubrics with individuals concerned in assessing outputs is different from the practices of evaluators, in employing both formative and summative evaluation of assessment, in determining from the records which audiences will receive the report on evaluation of assessment, in giving feedback to concerned individuals against relevant standards and in maintaining the record of all steps, information, and analyses of evaluation of assessment. On the other hand, evaluators from SU 1 were rated the lowest in terms of asking colleagues of the assessment or evaluation design which they found effective, in collaborating with fellow evaluators as regards assessment or evaluation procedures, and in recording analysed data. Evaluators from SU 2 were, however, found with the lowest ratings in terms of fully recording data collected.

The Mathematics content faculty of the different State Universities are different significantly in implementing factual standardized tests, students' use of manipulatives, students' application of mathematics, scheduled major tests, theoretical problem solving exploration and write up of projects as assessment practices in mathematics classroom. Mathematics content faculty of SU 1 and SU 3 implemented more frequently factual standardized tests. On the other hand, mathematics content faculty of SU 3 and SU 4 implemented more frequently students' use of manipulatives, students' mathematics applications, long exams, theoretical problem solving explorations and write-up of projects.

CONCLUSIONS

In view of the findings of the study, it can be concluded that Mathematics student teacher assessments are highly effective and efficient, with the assurance that the practices are realistic, prudent, diplomatic and frugal. The State University assessors highly support what are proper, fair, legal, right and just in their evaluations of Math student teachers. The assessments of Mathematics content faculty in the State Universities are moderate in their concerns for adequacy of evaluation documentation and in their focus on improving both learning and of the assessment process.

The assessments performed in the largest State University in the region have better overall utility, accuracy and accountability. Overall feasibility and propriety are of about the same level in the assessments across the four state universities. However, the assessors differ in the extent of their implementation of assessment practices.

It is highly recommended that dissemination sessions should be conducted to familiarize assessors in teacher education institutions of the meta-evaluation standards as it leads to global standards. The meta-evaluation checklist may be used in assessing the evaluation practices in student teaching in other subject areas and that further studies may be done on the applicability of the meta-evaluation checklists in other areas of assessment of student teaching in other majors other than mathematics, or in other general areas of assessment, not just student teaching.

REFERENCES

- Commission on Higher Education Memorandum Order No. 30, series 2004. "Revised policies and standards for undergraduate teacher education curriculum".
- De Lange, J. (1999). Framework for classroom assessment in mathematics. Fredenthal Institute and National Center for Improving Student Learning and Achievement in Mathematics and Science. Retrieved August 12, 2014 from www.fisme.science.uu.nl/catch/.../framework/de_lange_framework.doc
- Doran, R., Chan, F., & Tamir, P. (2002). Science Educator's Guide to Assessment. National Science Teaching Association. Arlington, Virginia. United Book Press
- Feuer, M., Floden, R., Chudowsky, N. & Ahn, J. (2013). Evaluation of teacher preparation programs: purposes, methods, and policy options. Washington, DC: National Academy of Education.
- Gold, B., Keith, S., & Marion, W. (1999). Assessment practices in undergraduate mathematics. MAA notes # 49. The Mathematical Association of America. Retrieved May 24, 2014 from <http://www.maa.org/sites/default/files/pdf/ebooks/pdf/NTE49.pdf>
- Huo, F. (2010). Integrating new assessment strategies into mathematics classrooms: an exploratory study in Singapore primary and secondary schools. National Institute of Education in Singapore. Research brief No. 10-003. Retrieved April 24, 2014 from www.nie.edu.sg.
- Joint Committee on Standards for Educational Evaluation (2013). Classroom assessment standards: Sound Assessment Practices for PK-12 Teacher. Draft # 5. Retrieved April 24, 2014 from http://www.teach.purdue.edu/pcc/DOCS/Minutes/12-15_Handouts/2013-
- Keeley, P. & Tobey, C. (2011). Mathematics formative assessment: 75 practical strategies for linking assessment, instruction, and learning. Virginia: National Council of Teachers of Mathematics.
- Mann, Gorge. (2004). "An Evaluation Approach Towards Feedback Betterment in an Initial Teacher Training in EFL". Retrieved on July 09, 2019 at "<https://www.asian-efl-journal.com/.../an-evaluation-approach-towards-feedback-betterment-in-an-initial-teacher-training-in-efl/>
- National Council of Teachers of Mathematics (NCTM) (1995). Assessment standards for school mathematics. Virginia: National Council of Teachers of Mathematics, Inc.
- Scriven, M. (1969). An introduction to meta-evaluation. Educational Products Report, 2, 36-38
- Sharifi, A. & Hassaskhah, J., 2011. "The Role of Portfolio Assessment and Reflection on Process Writing". The ASIAN EFL Journal. Retrieved on July 10, 2019 at www.asian-efl-journal.com/PDF/March-2011-as.pdf

- Stassen, M., Doherty, K., & Poe, M. (2001). Handbook on program-based review and assessment: Tools and techniques for program Improvement. Office of Academic Planning & Planning. University of Massachusetts Amherst.
- Stenmark, J. (1991). Mathematics assessment: myths, models, good questions, and practical suggestions. Virginia, USA. National Council of Teachers of Mathematics. NCTM, Inc.
- Stufflebeam, D. (1974). Meta-evaluation. Occasional Paper Series #3.
- Stufflebeam, D. (2001). The Meta-evaluation imperative. American Journal of Evaluation, Vol. 22, No. 2. American Evaluation Association. ISSN: 1098-2140
- Stufflebeam, D. (2012). Program evaluations meta-evaluation Checklist (Based on the program evaluation standards. USAID). Retrieved June 27, 2014 from Usaid.gov/pdf_docs/pnady.pdf.