

On the Validity of Educational Evaluation and its Construction

Xiaoping Huang¹ & Zhongfeng Hu²

¹ Institute of Education Research, Jiangxi Normal University, Nanchang, China

² The School of Public Administration, South China Normal University, China

Correspondence: Zhongfeng Hu, The School of Public Administration, South China Normal University, Guangzhou Daxuecheng, Guangzhou, China. Tel: 139-2271-9042. E-mail: huzhongfe@163.com

Received: June 8, 2015

Accepted: June 29, 2015

Online Published: July 21, 2015

doi:10.5539/hes.v5n4p99

URL: <http://dx.doi.org/10.5539/hes.v5n4p99>

Abstract

The main problem of the educational evaluation validity is that it just copies the conceptual framework system of validity from educational measurement to its own conceptual system. The validity conceptual system that fits the need of theory and practice of educational evaluation has not been established yet. According to the inherent attributive character of educational evaluation, this paper constructs eight concepts of validity in educational evaluation: objective validity, construct validity, content validity, cross validity, subject validity, process validity, function validity and result validity, and elaborates its basis and connotation.

Keywords: educational evaluation, validity, construct

1. Introduction

Educational evaluation is a process that making value judgment on educational phenomenon based on certain educational value or objective with viable scientific methods in order to provide evidence for continuous optimizing the education and educational policy-making by collecting and analyzing information systematically (Hu, 2006). Generally, we need to examine the extent that the results of educational evaluation achieve established educational value or target in the practice of educational evaluation, research the realization degree that educational evaluation index system reflecting the educational purpose in concrete evaluation, and study the extent that the evaluation tools and methods specially realize the evaluation effectively on educational objects when collecting all kinds of educational evaluation information. How to control all kinds of error variances that influence the validity of evaluation results to guarantee the internal validity of the evaluation when implementing some evaluation activity; what extent did the results of educational evaluation can popularize to the situation, scope, and field with a larger objects when making education policy, thus to guarantee the external validity of educational evaluation and so on. Basically, the realization degree of all the educational evaluation activities must be evaluated and tested by the validity of educational evaluation, and then judge the realization degree of educational value or objective.

1.1 *The Problems Existing in the Validity of Educational Evaluation*

1.1.1 Copying the Conceptual Framework System of Validity from Educational Measurement to the Conceptual System of Educational Evaluation Validity

Educational measurement is the feature description of one or more attributes of various objects in education. Its uniqueness is that measurement is a kind of quantitative description, which is using digital system to depict the attributes of educational object. In the conceptual system of educational measurement, the classic definition of validity is “it refers the degree that people desire when they measure the substance” (Cronbach, 1992). Therefore, when study the validity types of education object, researchers often highly emphasize four types of validity: construct validity, internal validity, statistical conclusion validity and external validity. In 90 general, construct validity refers to the degree that the actually measured construct fits the presupposed construct; Internal validity refers to the degree of deterministic relationship between variables; Statistical conclusion validity refers to the degree that conclusion can be explained and reflect things attributes; External validity refers to the degree that the measured results can be popularized to other situations.

But educational evaluation is making judgment on the value or feature of the education object under certain standard. Value can not only be the evaluation object in educational evaluation, such as study the value of some

educational object, but also can be the standard of educational evaluation, such as the selection and determination of evaluation standard. Thus, from this perspective, educational evaluation has a certain degree of subjectivity, which is an unity of subjective assessment and objective measurement. And educational evaluation activity is more complicated than educational measurement activity. From the perspective of educational evaluation factors, an evaluation activity includes value goal and standard, organization and staff, evaluation method and technology, the evaluation object and evaluation staff's psychological control and so on. From the perspective of educational evaluation object (education object) and the elements of evaluation content, the attention of educational evaluation is more enriched and complicate than educational measurement in content dimension, evaluation method, judgment basis, value judgment and so on.

Therefore, on the one hand, we cannot neglect the connection between educational evaluation and educational measurement when we study the validity evaluation of educational evaluation. And we should also see the difference between them, we cannot just copy the concept of validity from educational measurement to the conceptual system of educational evaluation. On the other hand, we should establish the connotation and concept of the educational evaluation according to its studying object and attributes to build its unique conceptual system of validity.

1.1.2 The Validity Conceptual System that Fits the Need of Theory and Practice of Educational Evaluation has not been Established Yet

Establishing the validity conceptual system of educational evaluation is the requirement of developing the theory and practice of educational evaluation. In China, there are seldom researchers did specialized and systematic research on the validity theory of educational evaluation. The representative study was done by Gu Zhiyue (1992), based on the perspective of improving the quality of educational evaluation. He studied in detail and analyzed three validity factors that influence the quality of educational evaluation in primary and secondary schools: external validity, which manifests the separation between evaluation results and objectives; internal validity, which reflects the existed deviation between information and facts; construct validity, which shows the non-isomorphic phenomenon between index system and evaluation object. But these concepts of the validity are still established within the category of educational measurement's conceptual system and did not make greater expansion and construction on the concept of validity and system of educational evaluation. To some extent, little attention we pay on the concept of validity of educational evaluation will influence theory of educational evaluation in our country developing toward the depth development.

However, foreign researchers study on the validity of the theory is very rich especially in USA. In the area of educational measurement and evaluation, the validity theory in modern sense is entirely different from the traditional validity theory, for examples, (1) modern validity theory is defined as the supporting degree of explaining and using test score, but not the attributes of measuring tool itself. (2) Messick (1989) put forward the concept "the unified conception of validity" in the research of the education test validity, and gave the specific guidance on validation. (3) Validation is identified as a process of evaluation. Validation need to use quantitative and qualitative methods to collect enough evidence in supporting of the inference made based on test score. (4) concept of validity and reliability is different in traditional validity theory. In the modern validity theory, reliability is regarded as the evidence of validity, also it is an indispensable and important part of validity. Under the modern validity theory, validity is demanded to interrogate the testing during the whole procedure from test development to explanation and to consequence, because "each link is important source of validity evidence" (Dowing, 2006). Whether educational evaluation is real and effective or not, it needs to collect evidence to verify if educational goal or objective is achieved. Therefore, from the point of actual demand of educational evaluation theory's development and modern validity theory's development, we need construct conceptual system of educational evaluation based on its characteristics and attributes.

2. Construction of Validity Concept of Educational Evaluation

Because of multiple value standards are involved in educational evaluation, it is difficult to judge effectiveness of educational evaluation just based on a type of value standard, so we should expand and construct validity of educational evaluation through multiple dimensions.

This paper takes factors that constitute educational evaluation system into account as logical basis, and discuss the construction of validity conceptual system of educational evaluation by inspecting the effectiveness of element features, such as objective, system, content, standard, method, subject, process, result and so on during the course of evaluating process.

2.1 Objective Validity, Which Realize the Degree of Certain Objective of Educational Evaluation and Educational Idea

Objective validity is the degree of educational evaluation achieved certain educational objection or value. This is the most fundamental value embodiment of educational evaluation, because validity itself is judged by realization degree of “objective or goal”. Thus, objective validity is judged based on realization of “educational objective or goal”. If an educational evaluation results in practice there is evidence to show that to achieve the established goals, implement and complete the educational goal of the original value, it’s can be strongly convinced that validity evaluation project has a certain target.

Fundamentally, objectivity validity is to examine reliability and validity of the overall goal of the educational evaluation, and under the general objective of each evaluation dimensions of sub-targets implementation degree, reflects the fundamental goal of educational evaluation and value demands.

2.2 Construct Validity, Which is the Consistency Degree between the Objective of Evaluation System and Educational Object

Messick (1989) considered the core of the modern validity theory is construct validity, and it does not exist several different validity, such as the validity of content, empirical, construct, generalization, external, result, and so on. But we think it is too general to sum up all the validity types to a construct validity, it is not helpful to distinguish the validity of different using purpose and also not accordant with the truth of educational evaluation. Therefore, in the area of educational evaluation, construct validity is different from the concept in educational measurement. It usually uses the certain given evaluation system and the organized evaluation scale (measuring tool) to represent and reflect the evaluation object, such as the abstract concept: creativity and intellect; the level of evaluation object such as education quality level and the realization degree of all included factors. Construct validity pays close attention to whether the behavioral indicator, element feature and evaluation dimension in the evaluation system reflect the evaluation object, and what it wants to express finally is the consistency degree between the objective of evaluation system and educational object.

Thus, construct validity not only need to focus on the theory or construction system (sometimes it probably do not need a construction system such as we can evaluate a result by if it meets the demand of educational policy), but also focus on whether the evaluation system really reflect all aspects of changes of the evaluation object’s psychological state, behavior orientation and ability level. Meanwhile, a tool of educational evaluation that possesses the good construct validity can accurately separate one group of examinee from the other. For example, a evaluation tool need to detect students’ ability to collect information when they confront the society. When test this evaluation result, if the designed evaluation items can separate the students who have excellent ability of collecting information with the weak ones, then this evaluation tool has construct validity. On the contrary, if the system and dimension of evaluation can not reflect the difference between students in the ability of collecting information, then this evaluation tool does not have construct validity.

2.3 Content Validity, Which Shows the Appropriate Degree of Educational Evaluation Goal or Behavior Sample Examined

In practice, educational evaluation needs to decompose general goal of educational evaluation. The general goal is decomposed to concrete sub goals of education, while the sub goals regulate which contents need to be examined and the corresponding weights. Thus, content validity is the appropriate degree of content or behavior sample which will be examined. In the process of actual evaluation, content validity is tested by evaluation goal as well as concrete and operable content dimension goal. First, it reflects adequate representation of content of general goal, whether the dimensional goals reflect demands of general goal that need to be studied; second, the items need to be evaluated should be the representative sample of defined content; third, the concrete behavior (such as knowledge and skills) that need to be evaluated in each item should keep consistent with evaluation goal of upper level. Meanwhile, each sub goal dimensions should be independent and not overlap, and strives to reflect the demands of general goal. We can use equivalent form method, retest method or experience method to examine content validity.

2.4 Cross Validity, Which Mutually Inspect and Verify the Consistency Degree of Various Educational Evaluation Methods

Actually, multiple evaluation criteria brings some difficulties to evaluate in practice. Sometimes, the valuator fuse different criteria at the operational level for some practical motivations, but often overlooked the difference between “procedural or technical methods” and “logical proof”, and lack self-consciousness and introspection toward to the methodology which hides the method of their own use (Smith & Heshusius, 1986; Tan, 1998). As a

result, expert groups are hard to evaluate the quality of their researches. When it comes to assessing the validity of educational evaluation, we need to fully examine the various evaluation methods and criteria for cross validity, because of different criteria. Triangulative methods can be used to test different data sources or different methods of data collection, it aims to confirm each other among diverse standards. Therefore, this method can be used as a way of testing cross validity.

In general, there are two basic evaluation methods in educational evaluation: quantitative evaluation (also known as quantifiable valuation method) and qualitative evaluation (also known as qualitative assessment method). Owing to the fact that these two methods are based on different paradigms, so they have great difference on research method, research approach, stages of research and so on, while, the types of evaluation validity also have essential difference. As follows in Table 1, it summarizes representative methods under different paradigms, and types of validity assessment which established on them.

Table 1. Comparison of two methods of evaluation and types of validity assessment

| Method Class | Examples of representative evaluation method | Type of validity assessment |
|--------------------------------|--|---|
| quantitative evaluation method | Mathematical methods: mathematical statistics method, for example: multivariate statistics (include cluster analysis, judgment analysis, correlation coefficient method, factor analysis, Bayes method) regression analysis, correlation coefficient method, fuzzy mathematics: such as fuzzy comprehensive judgment, AHP (Analytic Hierarchy Process), model of fuzzy distance, gray system theory (gray comprehensive evaluation) | Focus on Statistical Result Validity, Validate Validity, Construct Validity |
| Qualitative evaluation | subjective-based evaluation method: peer review, expert evaluation method, Delphi method, survey study, case analysis method, speculative analysis method; the evaluation method based on the situation and content: ethnography, phenomenology, interviewing method, content analysis, narrative investigation, field visit, archives literature research, philosophy. | Descriptive Validity, Interpretive Validity, Theoretical Validity, generalization validity, Evaluative validity |

Note. The classification of concept of validity in Table 1 is proposed by Maxwell (2002) when he defined “the validity of qualitative research”. Descriptive validity refers to accuracy (It means the consistency among researchers and the consistency towards to physical and behavioral in the event, such as frequency). Interpretation validity points at the meaning of the behavior or actions from participants’ perspective. Theoretical Validity is designed for the constructs of researchers; it includes independent construct and the causal relationship between constructs. Generalization validity attempts to extend the conclusion to the person in the community, organizations or institutions. Evaluative validity manifests the appropriateness of evaluation acts or events on the perspective of values.

It is important to note that, how to examine the validity of the methods of evaluation when integrating these two methods together. The general rule is that when we discuss the result of quantitative evaluation, we should consider the appropriateness that the result of research extend to the samples which from the general according to the laws of mathematical probability. When we discuss the qualitative evaluation result, we need to use different standards of validity evaluation criteria and validity type depending upon qualitative research goals, research methodology and the policy as well as research procedures. As it shown in Table 1, whether we use quantitative education methods or qualitative evaluation methods, it requires us look for validity criterion evidence in many ways, we have to take into account factors like the easy availability and possibility of measurement index.

2.5 The Validity of Evaluation Subject, Which Manifests the Accuracy and Fairness When Various Evaluation Subjects Assess the Evaluation of Object and Evaluation Results

There often exists several subjects and multiple stakeholders in modern educational evaluating practice. Consequently, different evaluation subjects with different evaluation objectives that the using of evaluation conception, evaluation criteria, evaluation method and appraisal tool may be significantly diversified, or even on contrary, which makes evaluation lose its validity. Essentially, evaluative validity focus on the purpose. When different evaluation subjects hold different evaluation purposes, values, positions, standards, the difference will

be existed according to the range of evaluative validity explanation, purpose, value and the significance of evaluate itself. Thus, when we appraise an educational activity, ensure that the immediate subject of valuation in accordance with validity assessment with the same objective is of great importance to us. Of course, there is some difference among the members with the same subject. When it comes to the evaluation of educational objectives, it is possible that the reliability of the valuator becomes lower because the command of evaluation standard of inconsistencies between individuals, which affects the overall validity of the evaluation.

Consequently, it is necessary for us to investigate the validity of subject of educational evaluation when an educational evaluation activity is finished. We divide the validity of the subject into two aspects, namely:

On the one hand, when it exists multiple subjects, then, there will be more than one validity of evaluation subjects that can also complement each other. It is convenient for the education determiner to choose and appraise it among various validities properly, and listen to all useful opinions. Thus, it is beneficial for the education determiner to get much more information about educational evaluation and open to good advice so that they can do their job better.

On the other hand, multiple evaluation subjects will bring some difficulties to the activity of validity assessment. It also makes the result and interpretation of the educational evaluation with fuzziness and uncertainty, as a result of that, it leads to different educational evaluation. Particularly, when it refers to assess the interest relationship, it will affect the educational evaluation activity's justice and equity. Furthermore, it is going to influence the efficiency of evaluation result. Therefore, when we pay attention to the diversity of the validity of evaluation subject plays a role promoting educational evaluation result, distinction, comparison, and identification must be made among various validity of the education subject according to the directions which educational evaluation used for, and then select the subject which its validity is much higher than others.

2.6 Procedure Validity: Effectiveness of the Process Which all Kinds of Evaluation Information and Data Collected

Educational evaluation is performed by valutors, it is a process that the valuator gather information and data by using a variety of assessment methods and tools. Actually, evaluation activity is an ongoing and cumulative process. So, the point of the evaluation process's validity is the dynamic and developmental process of the object being evaluated, there is more of an emphasis on the experience which happens as they grow under the stated objectives, and the totally historical and realistic achievements during the whole educational process. For examples, the performance evaluation carried out during the process that collect student's excellent works by using portfolio assessment method, and employ value-added assessment approach to school's developing evaluation. From another point of view, the process of educational evaluation is a vital assessment step and operating principle during the whole activity of educational evaluation, any one of the links and key evaluative feature, such as treat the evaluation index differently or procedure errors when estimate the same index can produce bias into the whole validity of education. For instance, what kind of method should we use when we collect some data or information, using a questionnaire or a structural interview? Quantitative evaluation or qualitative evaluation should be adopted? Once we make a decision, then, the strategy and key point of evaluation would be different. Thus, it is necessary for us to use a number of validity indexes during the evaluation.

How to look the procedure validity, it requires a combination of observations. When we collect the evaluation information of the object being evaluated, more attention should be paid to the internal validity of data. If we want to establish an index, the scientific state, completeness and independence of index should be taken into full account. While, we must consider the discrimination validity of index, and whether each measurement reflects the content validity of the overall target. Exploring the predictive effect of the existing measuring tools, we have to investigate real-time validity of the tool. And when we use measuring tools, like a certain scale, we should keep a watchful eye on the construct validity of them. Well, when it comes to make formative evaluation of students' academic ability level, we need to use cognitive diagnostic evaluation tool, and pay more attention to the statistical conclusion validity of models and methods. Therefore, for the validity evaluation of assessment process, it makes the whole synthesis validity assessment under the level that all kinds of synthetic validities are relatively high. It is necessary to evaluate the methods, tools which used during evaluation, recheck it (known as meta-evaluation) and ensure the procedure validity.

2.7 Functional Validity, Which Realize the Degree of Educational Evaluation Objective's Existing Function

Owing to the fact that educational evaluation has the function of chosen, identification, selection, analyze its strengths and weaknesses along with the ability to sort, so we have to observe the functional validity of educational evaluation. Functional evaluation means the realization degree of educational evaluation objective's

existing function, it shows the utility of educational evaluation. The main content of the functional validity is shown as it should reflect the incremental index of educational evaluation and give a reflection of practicable, reasonable and easy navigable. Hence, the focus of functional validity is “result indicator” validity which connects with the function of educational evaluation directly. For instance, the college entrance examination’s assessment function is selection, the functional check index, like the college enrollment rate and the major university admit rate is realized as a concrete functional validity representation. The index shall be the most important criteria for validity evaluation. In fact, the functional validity of educational evaluation shows the realization degree of educational objective. Actually, evaluator select what kind of evaluating function determined by the educational objective that he holds, it also decides the functional validity of evaluation.

2.8 Consequential Validity, Which Achieve the Degree of Educational Evaluation Result

Functional validity, objective validity and consequential validity which presented in the end are in the same strain. It explains how large extent the evaluation result achieve the desired objectives and evaluation targets, and makes sense for valuator’s intended use as well as the value of the target of education, meanwhile, it can also better reflect the evaluate requirements of the object being evaluated, and get good results in practice. In summary, four aspects affect the consequential validity are investigated separately as follows:

Firstly, consequential validity should reflect the purpose of evaluation, that is, it should meet the requirement of purpose.

Secondly, evaluation information feedback must reflect the various needs of the object being evaluated. Especially, let the object being evaluated acknowledge the evaluation result, so that it can achieve the original thought and target of educational evaluation, and can better play the role of supervisory.

Thirdly, the result of evaluation should be able to get applied and verified in practice with varying degree of successful utility.

Fourthly, the target group and scope of anticipated that the result of educational evaluation can be extended. Therefore, the content of consequential validity includes four factors: fulfill the evaluation objectives, get recognition for the requirement of the object being evaluated, achieve the expected educational results and reach its promotion.

In conclusion, the validity of educational evaluation is a concept with rich implication. Validity evaluation index system is such a multidimensional and multi-perspective system that we must construct it from the integrated point of view. The validity concept of educational evaluation has richer connotation than the concept which refers to the validity of educational measurement.

3. Conclusion

Sum up the above analysis and discussion, the validity of educational evaluation can be summarized as follows:

Objective validity is the realization degree of some certain objective of educational evaluation and educational idea. In certain extent, objective validity is the fundamental purpose of the educational evaluation so it is the first place in the framework of validity.

Construct validity is the consistency degree between the objective of evaluation system and educational object. It is deferent from the concept of educational measurement, it mainly focuses on evaluation dimensions and system can truly reflect and meet educational object.

Content validity is the appropriate degree of the educational evaluation goal or behavior sample. Generally speaking, content validity is close correlated to dimension of evaluation system, so the evaluation goal should faithfully reflect trait or talent ability of evaluation objectivity.

Cross validity is the degree of consistency which various educational evaluation methods for inspection and verification. In brief, different methods can validate and test each other, it focuses on the methods’ of effectiveness.

The validity of evaluation subject is the accuracy and fairness that various evaluation subjects toward to the evaluation of the object and evaluation results. Accuracy and fairness possibly affect the benefit of evaluation object, so the validity of evaluation subject must be paid attention to the evaluation process.

Procedure validity is the effectiveness of the process which all kinds of evaluation information and data about educational evaluation collected. In fact, every steps and procedures involving evaluation methods must be surveyed effectively when evaluators are collecting data and information.

Functional validity is the realization degree of educational evaluation objective’s existing function. Function,

sometimes refers to utility are the two basic attributes of evaluation, so the focus of functional validity is “result indicator” validity which connect with the function of educational evaluation directly.

Consequential validity is concrete manifestation of the value realization of educational evaluation and the value realization of subject and object. It consists of four dimensions: objective validity, requirements of validity, the result validity and external validity.

Above all, we mainly discuss eight concepts of the validity of educational evaluation from the properties and characteristic of educational evaluation, and try to expand and construct its content and research method. In fact, due to the richness, complexity and diversity of educational evaluation activity, it decides the method of educational evaluation also has the characteristics of diversity. Just as Patton (1986) mentioned in his *Utilization-focused Evaluation* “To ensure the feasibility evaluation of utility, rationality and accuracy, it must have the sensitivity of the situation, the method of flexibility, diversification of valuator, a good political accomplishment and innovation.” For this purpose, we need to find out the foundation of the criterion assessment which is practical, workable and considerable.

Acknowledgements

This paper is sponsored by Fund Project: “Guangdong Province Science and Technology Innovation Talented structure and Evaluation Standard Research”, a key project of Guangdong Province soft science research in 2012 (Project Number: 2012B070300050).

References

- Chen, X. M. (2000). *Qualitative research methods and social sciences* (1st ed.). Beijing: Educational science publishing press.
- Downing, S. (2006). Twelve steps for effective test development. In S. Downing, & T. Haladyma (Eds.), *Handbook of test development* (pp. 3-25). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gu, Z. Y. (1992). *Analysis of The Quality of The Evaluation of Primary and Secondary Education in China*, 6, 19-23.
- Hu, Z. F., & Mo, L. (2007). Reconstruction of Validity Theory in Psychological and Educational Measurement. *Journal of South China Normal University (Social Science Edition)*, 12, 82-90.
- Hu, Z. F. (2006). *Educational measurement and evaluation* (2nd ed.). Guangzhou: Guangdong higher education press.
- James, T. C., David, M., & Lara, R. L. (2001). *Hawthorne with Evaluation and performance measurement projects*. Beijing: Educational science publishing press.
- Messick, S. (1989). Validity. In R. L. L. Inn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education/Macmillan publishing Co.
- Patton, M. Q. (1986). *Utilization-focused evaluation* (2nd ed.). Sage, Beverly Hills, California.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).