COMMENTS ON PROFESSOR WILEY'S PAPER ENTITLED "DESIGN AND ANALYSIS OF EVALUATION STUDIES"

Theodore Husek

From the Proceedings of the

SYMPOSIUM ON PROBLEMS IN THE EVALUATION OF INSTRUCTION

University of California, Los Angeles
December, 1967

M. C. Wittrock, Chairman

Sponsored by the Center for the
Study of Evaluation

The research and development reported herein was performed pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education under the provisions of the Cooperative Research Program.

CSE Report No. 30, May, 1969
University of California, Los Angeles

COMMENTS ON PROFESSOR WILEY'S PAPER ENTITLED "DESIGN AND ANALYSIS OF EVALUATION STUDIES"

Theodore Husek

I find myself much more interested and stimulated by the latter sections of David Wiley's paper than by the introduction, the definitions, and the refinements of terminology. I know the language framework is necessary, but right now I find, so far as my own work is concerned, I am not as interested as perhaps I should be in the definitional problem.

I see the major task of the methodology in evaluation as being the development of new ways of helping the content specialist construct and evaluate educational products. As part of this task we need to do a better job of data collection and data analysis.

In this context I think the paper brought out some extremely important issues. We should be interested in the distributions of scores on tests as well as the mean. At the same time I think we need to use the traditional item more, and also reexamine the nature of the items which we use in evaluation studies. We have to examine new indices, whether or not they are obtrusive or unobtrusive.

Wiley's point about paying attention to the unit of study is also important. We seldom pay as much attention as we should to whether we are studying students, classes, teachers, or school systems. Many times we really are not interested in the individual

student, and in these cases I feel that item sampling may provide an immense break-through in data collection procedures. As a footnote I would like to say that I think it unfortunate that this particular term "item sampling" got started--I do not think it really represents what is happening; there is more involved than sampling just items.

If we do not have to ask every student every question in our study, then it may be possible to begin to obtain data on the multitude of measures that we all seem to think are important. As a simple example, in the classroom situation we can use tests which in part serve to help us grade the students, in part help us to judge the course, and also give us a little data about anything we might be interested in.

With respect to item sampling, I feel that there are at least two important questions for which we do not have answers. The first of them is the context effect. One way to use item sampling is to give each student one item and to give different items to different students, but I have as yet no idea of the physical effect this has on the student. If you give him one item out of context, will he respond differently than if the item were in the context of similar items, or, for that matter, in the context of different items? Ken Sirotnik and I are now performing a study to examine this issue.

My other question about item sampling concerns its optimal use. Given a set of subjects under certain circumstances and items with certain characteristics and various test conditions, what is the optimal number of items to give to how many students? Currently Dr. Sirotnik and I are also planning a computer simulation study to examine this messy issue.

The item sampling research I have been pursuing has reminded me of another dimension which must be considered in evaluation studies. In one of two empirical tryouts of item sampling procedures we obtained an item matrix sample that produced a negative variance for the population from which we were sampling. We finally decided that there was no mistake in the formula and discovered the negative variance would be produced by item matrix samples with negative coefficient outputs.

This led us to some serious thinking about the nature of the collection of items from which we had samples. We were led, for one thing, to see the need for a special kind of homogeneity in the population from which you are taking your items--not necessarily a homogeneity in the coefficient alpha sense--but some other kind. The main conclusion we reached was that we had to pay more attention to the purposes of the test than we had thought necessary, and this is another dimension of extreme importance in evaluation.

Not only is the content of the test important, not only is the unit of study, but also the nature of the test is important. Do we

want an achievement test with maximum variance? Do we want a test to measure change? Should the test be course-vocabulary free?

The question here is one of defining criteria for the various purposes. I do not think that we need a new statistics for any of the points I have made up to now; I do not think we need a new test theory. I do think we have to be a lot clearer about what we are trying to do.

As my last point I would like to bring up something I do not know how to handle at all. I will use Dr. Popham as an example, largely because we have talked about his particular issue. He is trying to train product researchers--people who will be near-technicians and who will hopefully produce better instructional programs. It is one thing to say that this can be done by just using the rules of the game that we already know, but it would be silly not to try to learn how better products are built. Given that the need to produce a product includes the possibility of performing an interminable series of experiments which examine each variable, how will we be able to collect and use data from the ongoing developmental process to help understand product development and, most of all, improve it?

OE 6000 (REV. 9-66)

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION

ERIC REPORT RESUME

(TOP)

ERIC ACCESSION NO.

CLEARINGHOUSE ACCESSION NUMBER | RESUME DATE 5-1-69 | P.A. | T.A.

IS DOCUMENT COPYRIGHTED? YES ☐ NO ☐

ERIC REPRODUCTION RELEASE? YES ☒ NO ☐

001

TITLE

100 Comments on Professor Wiley's Paper Entitled "The Design and Analysis
101 of Evaluation Studies: Comments and Suggestion."
102
103

PERSONAL AUTHOR(S)

200 Theodore Husek

INSTITUTION (SOURCE) | SOURCE CODE

300 UCLA - CSE

310 REPORT/SERIES NO. CSE Report No. 30

OTHER SOURCE | SOURCE CODE

320 Symposium on Problems in the Evaluation of Instruction, Dec. 1967

330 OTHER REPORT NO.

OTHER SOURCE | SOURCE CODE

340

350 OTHER REPORT NO.

400 PUB'L. DATE 5-69- | CONTRACT/GRANT NUMBER OEC 4-6-061646-1909

PAGINATION, ETC.

500
501

RETRIEVAL TERMS

600
601
602
603
604
605
606

IDENTIFIERS

607

ABSTRACT

800 Husek suggests here that the development of new ways of helping the
801 content specialist construct and evaluate educational products is the
802 major task of methodology in evaluation. He emphasizes the distribution
803 of scores on tests as well as the man, more use of the traditional item,
804 reexamination of the items used in evalution studies, and an examination
805 of new indices. The paper also discusses methods and problems of
806 item sampling.
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822