ASSESSING STUDENT REACTIONS TO INSTRUCTION

Stephen A. Graf

In University-handled Student Reactions to Instruction, attempts have been made to provide data for decisions such as reappointment, tenure, and promotion. If data is to be used for decision-making (whatever the situation), the data display should cue appropriate interpretative responses on the part of anyone viewing the display, especially the decision-makers (Johnston and Pennypacker, 1980).

The most important starting point is to ensure that what is being measured is actually behavior. In addition, the behavior should be accurately described, e.g., label 'student responses to instruction' as such and refrain from calling such responses 'teacher behavior'.

'Student Reaction to Instruction' instruments often have the following format. Questions are asked (e.g.,"How well did this instructor perform in the teaching of this class?") and a limited number of choices are offered as responses, one to be chosen by each student responding. Each category is described and ordered from 'best' to 'worst'. Such category scales are ordinal, in that no quantitative distances exist between one category and another.

I have suggested elsewhere that allowing 'free operant ratio responding' would be both feasible and more appropriate (Feitler and Graf, 1978), but since category scaling is in such widespread use, the present focus is on displaying the data from category scales. I'd like to pinpoint some of the faulty measurement practices typically used, and offer some possibilities for improvement. Table 1 lists some inappropriate practices with a brief rationale and example, and Table 2 lists more appropriate alternatives, also with brief rationales and examples.

Insert Tables 1 & 2 about here

Individuals familiar with six-cycle Standard Celeration Charts (Pennypacker, Koenig, and Lindsley, 1972) will likely be familiar with the frequency, celeration, and bounce measures suggested in Table 2. White and Haring (1976) detail the utility of such an approach in teaching, and Lindsley (1979a, 1979b) has provided some updated facts and findings.

A standard data display that represents real behavior on a background of real time and incorporates descriptive measures of change and variability should be the starting point for the evaluation of student reactions to instruction by decision-makers. Appropriate assessment of measured responses is a necessary prerequisite to interpretative evaluation of those responses in a natural science of behavior.

REFERENCES

Feitler, F.C., & Graf, S.A. The use of ratio production scales to assess quality of teaching performance. ERIC Document Reproduction Service No. ED 163 076, 1978.

Johnston, J.M., & Pennypacker, H.S. Strategies and Tactics in Human Behavioral Research. Hillsdale, New Jersey: Erlbaum & Associates, 1980.

Lindsley, O.R. Learning Picture Facts. Kansas City: Precision Media, 1979(a).

Lindsley, O.R. Standard Celeration Charting Flash Cards. Kansas City: Precision Media, 1979(b).

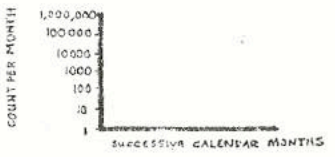
Pennypacker, H.S., Koenig, C., & Lindsley, O.R. Handbook of the Standard Behavior Chart. Kansas City: Precision Media, 1972.

White, O.R., & Haring, N.G. Exceptional Teaching. Columbus, Ohio: Charles Merrill, 1976.

Table 1. Inappropriate Assessment Practices of Student Responses to Instruction

Inappropriate Behavior	Rationale	Example
Assignment of an equal-add-interval numerical weight to each category	Assignment usually after-the-fact; imparts interval measurement to ordinal data; such numbers are not behavior; fictions should not be used where real behavior could be	Outstanding=5, Very good=4, Good=3, Adequate=2, Inadequate=1
Calculation of mean of numerical weights	Tries to represent a distribution of responses with a number derived from arbitrary numbers; a fictitious performance measure	Mean=4 (Use of more than one significant digit, e.g., 4.1, isn't allowable but is commonly found)
Summarization of means across terms to produce a 'mean of means'	Collapsing data (whether means or real responses) across time destroys any 'trend' in the data.	Successive means of 4,4, 3,2,2=mean of means of 3; Successive means of 2,2, 3,4,4=mean of means of 3
Transformation of responses to percentages within categories	Imposes ceiling of 100%; lose information on numbers of students responding	100% of responses=outstanding, class size=5; 100% of responses=outstanding, class size=150
Collecting data one term per year	Ten data points representing each category are optimal for interpreting trends and variability	Data for three terms per year yields 10 data points per category in just over three years; two terms per year would require five yrs; one term per yr. would require 10 yrs.

Table 2. Suggested Practices in Assessment of Student Responses to Instruction

Appropriate Behavior	Rationale	Example
Use of the actual number of responses in each category per term	Collections of student responses in each category legitimately represent a "performance" measure or "frequency" (count per term); real behaviors are located in real time	Winter, 1980: 120 students responded; 35 responded 'outstanding', 40 responded 'very good', 15 responded 'good', 20 responded 'adequate', 10 responded 'inadequate'
Use of a standardized data display with a multiply scale for the frequency of responses per term, and an add scale representing real time (successive calendar months) across the bottom	Standardization facilitates accurate interpretation; a multiply scale facilitates appropriate responses to relative variability; a real time scale helps locate responses in real time	
Drawing the "celeration line" for each category, including the total number of students responding	Celeration is the scientific name for performance across time (count/term/months), and measures trend; celeration is independent of any single performance	From Fall '75 to Spring '80, total number of students responding accelerated by a factor of 1.1 every six months, while no. of students responding 'outstanding' accelerated by a factor of 1.4 every six months, etc.
Drawing the "bounce" around the celeration line for each category, including the total number of students responding	The bounce shows the variability per term of the responses; bounce is independent of celeration and a data display of both allows us to visually separate the bounce from the celeration	From Fall '75 to Spring '80, total no. of students responding bounced x4, while no. of students responding 'outstanding' bounced x3, etc.
Collecting data every term	Ten data points representing each category are optimal for interpreting trends and variability	Data for three terms per year yields the 10 data points in just over three years