

1998

A Comparison of Likert Response Formats in Student Affairs: The ABC's of Assessment

Oscar T McKnight, *Ashland University*

Robin W Gagnow, *Ashland University*

Sue Heimann, *Ashland University*

A Comparison of Likert Response Formats

in Student Affairs:

The ABC's of Assessment

Oscar McKnight, Director of Counseling Center, Ashland University

*Robin W. Gagnow, Director of Residence Life and Leadership
Development, Ashland University*

B. Sue Heimann, Dean of Students, Ashland University

The evaluation of higher education is not a new concept, but the intensity of outside agents demanding verification of outcomes is increasing. As student affairs (SA) professionals we need to increase our accountability procedures by incorporating on-going assessment of our programs, services, and facilities needs. At the university level, Cress (1996) reviewed pressures and efforts of legislators, business, and the public to report outcomes of higher education. Cornesky et al. (1991) addressed total quality management within an educational system, and Upcraft and Schuh (1996) wrote an assessment guide specifically for student affairs (SA) practitioners. They define assessment as "any effort to gather, analyze, and interpret evidence which describes institutional, departmental, divisional or agency effectiveness" (p. 18).

There are many standard and packaged assessment tools a SA practitioner can employ with national scope or norming (Upcraft and Schuh, 1996). However, though these packaged models are pragmatic and valid, they are without restriction. Some general limitations that impact SA include: a single evaluation period; the national norming group; the areas are categorical; and high start-up costs. Contrast this with the realities of SA departments needing multiple assessment periods, recognizing population differences, and attempting to fit SA programming into one category or another within a single department appears flawed. As for cost, most SA departments have limits in real dollar expenditures. Hence, the weight of examining a SA question is proportional to cost

and dependent on response. Therefore, it is not unusual for SA departments to create their own survey instrument.

ASSESSMENT

Historically, most evaluations or assessments used some form rank-order scaling to evaluate attitude (Thurstone and Chane, 1929). Likert (1932) formalized this categorical rank-ordered process. Other researchers, Edwards (1957) offered techniques for constructing attitude scales; Oppenheim (1966) focus on questionnaire designs; Kerlinger (1973) established and reviewed the foundations of ail behavioral research; and Newman (1976) offered suggestions for combining theory and practice when detailing the basic procedures in conducting survey research. Thus, though briefly reviewed, the theoretical foundations for establishing a SA evaluation instrument are in place.

One option available to a SA professional engaged in assessment is to state questions of interest and ask students to give that area of concern a grade (i.e., A, B, C, D, or F). A grading system or approach to SA assessment offers promise to researchers. Questions are brief, answers are familiar, and the ABC format is a qualitative measure with quantitative properties (McNeil et al., 1975; or, Lee and Fielding, 1991).

THE CURRENT STUDY

Student Affairs provides an array of services within any institution of higher education. Typically, one associates SA with arranging, planning, supporting and evaluating a broad spectrum of student life areas. The goal of SA is to enhance student learning and development through holistic collaboration. As such, SA targets identified need areas and intervenes with educational programming or support services. How well SA intervenes in terms of programming and support services is a critical question. It requires carefully and well planned assessment of individual departments to determine strengths and

weaknesses. The results can be used in long term planning and assist in strategic planning for the SA program.

Patton (1990) reviewed the distinctions and choices involved when deciding between the qualitative, quantitative, or combination method in assessment. Guidelines suggested that it is the researcher's responsibility to validate procedures to answer questions of interest; and accordingly, the specific research design must address posed queries. Chang (1994) examined the reliability and validity of 4-point and 6-point scales. Chang found that criterion related validity was not affected by the number of scale points; however the researcher noted scaling may depend on the empirical setting. Garg (1996) investigated the effects of positively and negatively worded Likert statements and found bias to follow the direction of question wording. Hence, the manner of questioning and correcting for bias is a design consideration in evaluation. Likewise, Kerlinger (1973) noted that valid interpretations and generalization cannot be made with less than 80-90 percent survey return rate.

In the current study, researchers investigated two common Likert response formats used in SA. The first (A) was a 6 point Likert scale; and the second (B) was a qualitative grade assignment system. The overall goal was to consider the impact that scaling has on the turn-around time, the return rate, the response sets, and the interpretation. The specific research questions of interest were as follows:

- (a) Does the manner of Likert scaling influence turn-around time?
- (b) What is the difference between return rates given the structure of Likert scaling?
- (c) Will the method of Likert scaling impact response sets?
- (d) Can researchers combine a qualitative-quantitative format to yield valid interpretations with statistical properties?

METHOD

Participants

All subjects were volunteer full-time residential students in a private mid-west university. The interval of interest was equivalent to one graduation class (i.e., 4 years). When averaged, the proportional class rank was as follows: 38% Freshmen; 29% Sophomore; 21% Junior and 12% Senior; with a respective female - male ratio of 60:40.

Procedure

Data collection occurred during the *Fall semester* (i.e., October and November) throughout a four year study interval. Each student received one of two types of surveys (see Note Section in Table 2.). Distribution of surveys was the responsibility of Resident Assistants (RA) assigned to various halls and floors. Once gathered, valid surveys entered a data pool for random selection. Determining the exact number of surveys for analysis followed a 95% confidence level for representativeness (Krejcie and Morgan, 1970). For both instruments, a correction for response sets ensued.

Instruments

The assessment tools under investigation examined the university mission, individual departments within SA, and specific residential questions. As such, instrument construction used an expert panel of judges to determine relevant questions, response-format, and content validity. In the first two years participants responded to a single lead question asking: how would you rate the following services at this university X using a 6-point Likert scale. However, in the second two years, participants responded to the lead question asking: what grade would you give university X in the following areas and requiring a student to fill-in a letter grade.

RESULTS

This study posed four non-directional questions. The first questioned if the manner of scaling influences the rate of turn-around time. Descriptive findings suggest that the ABC scale design does reduce the number of days required for completion. A second concern centered return rates and the impact of Likert-response formats. Further descriptive findings indicate that return rate undergoes a noticeable increase when employing an ABC response-format. When examining the impact of creating a response-set by survey format, the third question of interest; the ABC response format produced a lower proportional response-set than the typical Likert scale employed. Table 1. presents an evaluation summary of comparative results addressing the first three research questions.

Table 1.

Comparative Evaluation of Two Survey Designs in SA

Year	Scale Design	Survey Total	Turn-Around	Return Rate	Response Set	Pool Size
1	(A)	950	17	38%	45%	199
2	(A)	1050	19	42%	51%	217
3	(B)	1120	07	97%*	17%	902
4	(B)	1290	06	81%*	25%	783

Note. Year = rank-ordered according to design application; Scale Design (A) = Likert format: 1 - 6 where, one was significantly less than expected... three met expectations... and 6 significantly exceeded expectations; Scale Design (B) = grading system: where, a student assigned an academic letter grade (i.e., A, B, C, D, or F) to SA

services; Survey Total = Total number of distributed surveys; Turn-around = (measured in days) point between distribution and return; Return Rate = ratio between number of surveys distributed and number returned; Response Set = number of surveys eliminated from analysis due to a negative, average, or positive response set (i.e., an assessment filled-in with only one rating value); Pool Size = number of surveys available for a valid analysis following a correction for response sets;

* = Significant return rate (Kirlinger, 1973).

Question four examined if combining the properties of a qualitative-quantitative format; that is when converting an ABC letter grade to its numerical equivalent (i.e., A = 4 points, B = 3 points... etc.) can produce meaningful data. Findings indicate the efficacy of this process and offers a chart detailing an expected ABC effect size with relative interpretation. Table two, highlights the findings when examining 26 SA areas in relationship to expected effect size. Effect size describes the average deviation between the ABC numerical means over a two year interval.

Table 2.

ABC Response-Format: Effect Size and Relative Interpretation

EFFECT SIZE	CLASSIFICATION	NORMAL	CURVE POSITION
.10	NO SIGNIFICANT IMPACT		— MEAN SCORE —
.15	LOW IMPACT	.5	STANDARD DEVIATION
.20	MEDIUM IMPACT	1.0	STANDARD DEVIATION
.25	HIGH IMPACT	1.5	STANDARD DEVIATION
.30	SIGNIFICANT IMPACT	2.0	STANDARD DEVIATION

Note. Effect size can be a positive or negative number. The classification: SIGNIFICANT IMPACT represents an established theoretical point for most research designs (i.e., a score two standard deviations from an expected mean is significant). A modification of the actual mean and standard deviation ensued for application ease (i.e., real mean was .11 with a standard deviation of .10).

Discussion

This study has several implications for SA professionals engaged in assessment. Foremost, the descriptive findings support any SA decision to employ a grading system approach in developing a response format. The findings associated with the ABC method strongly suggest the assessment viability. Two feasible reasons for the apparent ABC success in evaluation are familiarity and catharsis. The first concept is obvious; for, no exhaustive narrative is necessary to explain the rating system. As for catharsis, student participation probably acts as an emotional release; that is, assigning a grade when contrasted against receiving one.

Though a literature review found no relevant studies addressing turn-around time; it is a pragmatic function. No assumption is made concerning validity issues and length of study; but rather, this study documents the apparent ease in participant completion. The decreased amount of time a researcher spends collecting data increases allotted time for analysis. So, a reduction in the interval period between administration, scoring and interpretation is not viewed as a disadvantage. Likewise, the increased percentage in return rate dramatically influences the ability to draw valid conclusions concerning population characteristics. In a similar fashion, increasing the data pool size for random assignment can only enhance statistical validity and reliability measures. If the ABC response-format decreases response-sets as findings indicate, the probability of an accurate assessment increases. As a caution, findings strongly suggest a correction process for biased response-sets.

When assigning meaning to a grade received in SA, Table II offered some working guidelines in assessment. For example, if a SA area received a composite GPA of 2.80 in the first year of assessment and a 3.20 the following year; an effect size would be .4. Hence, findings would suggest that area in SA made a significant impact in programming or service. A second option derived from the effect size table, is in employing your actual composite university GPA as a theoretical baseline and predetermine a testable significance level. For instance, if your university composite academic GPA is 2.95, your specified SA areas could establish their individual significance level at a .25 effect size. Then receiving only a GPA of 3.15 or greater would indicate successful programming or service. It becomes apparent that there are many options available when employing the ABC response format in terms of analysis. Furthermore, since the ABC response-format is initially a qualitative assign-

ment (i.e., A, B, C, D, or F) it still possesses quantitative properties (i.e., A = 4 points, B = 3 points... etc.); consequently, traditional analysis is possible.

Although this current study provided some important information in terms of understanding the nature of response-formats in SA; two limitations need reviewing. First, only one university participated in this study, thus the ability to generalize results are suspect. Second, this participant university has an active SA research program. On that account, effects demonstrated may reflect program growth and development in assessment beyond the actual impact of a selected response-format. For those reasons, future study should examine the nature and impact of SA response-formats in other universities. Likewise, examining the nature of response-sets in SA assessment would add to the body of professional literature addressing validity measures.

References

- Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, 18(3) 205-215.
- Cornesky, R., McCool, S., Byrnes, L., and Weber, R. (1991). *Implementing total quality management in higher education*. Newbury Park, Calif.: Sage.
- Cress, C. (1996). *Assessment and testing: Measuring up to expectations*. ERIC Clearinghouse for Community Colleges, #391, 559.
- Edwards, A.L. (1957). *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts.
- Garg, R.K. (1996). The influence of positive and negative wording and issue involvement on responses to Likert scales in marketing research. *Journal of the Market Research Society*, 38(3) 235-246.

- Kerlinger, F.N. (1973). *Foundations of behavioral research* (2nd ed.).
New York: Holt, Reinehart, and Winston.
- Krejcie, R. and Morgan, D. (1970). Small sample techniques. *Education and Psychological Measurement*, 30, 607-610.
- Lee, R.M., and Fielding, N.G. (1995). Computing for qualitative research: Options, problems, and potential. In N.G. Fielding and R.M. Lee (eds.), *Using computers in qualitative research*. London: Sage.
- Likert, R.A. (1932). Technique for measurement of attitudes. *Archives Psychological*, 40.
- McNeil, K.A., Kelly, F.J., and McNeil, J.T. (1975). *Testing research hypotheses using multiple linear regression*. Southern Illinois University Press.
- Newman, I. (1976). *Basic procedures in conducting survey research*. The University of Akron, Akron, Ohio.
- Oppenheim, A.N. (1966). *Questionnaire design and attitude measurements*. New York: Basic Books.
- Patton, M.Q. *Qualitative evaluation and research methods*. (2nd ed.). Newbury Park, Calif.: Sage.
- Thurstone and Chane (1929). *The measurement of attitude*. Chicago: University of Chicago Press.
- Upcraft, M.L., and Schuh, J.H. (1996). *Assessment in student affairs: A guide for practitioners*. San Francisco: Jossey-Bass Inc.