DOCUMENT RESUME

ED 424 802		HE 031 597

AUTHOR	Pike, Gary R.
TITLE	The Constant Error of the Halo in Educational Outcomes Research. AIR 1998 Annual Forum Paper.
PUB DATE	1998-05-00
NOTE	50p.; Paper presented at the Annual Forum of the Association for Institutional Research (38th, Minneapolis, MN, May 17-20, 1998).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	College Freshmen; College Seniors; *College Students; Higher Education; *Learning Experience; *Learning Theories; Models; *Outcomes of Education; *Self Evaluation (Individuals); Student Attitudes; Tables (Data); Universities
IDENTIFIERS	*AIR Forum; College Student Experiences Questionnaire; *Halo Effect; University of Missouri Columbia

ABSTRACT

This study examined whether halo error--the masking of college gains by general gains in intellectual development--influenced students' ratings of their learning and development during college. A total of 1,084 first-time freshmen at the University of Missouri-Columbia (MU) completed the College Student Experiences Questionnaire (CSEQ) during the winter 1996 semester, and 1,604 seniors completed the MU Senior Survey in the winter 1996 or winter 1997 semesters. It was found that, for freshmen, halo error generally accounted for more than one-half of the explained variance in students' ratings; for seniors, halo error accounted for one-quarter to one-half of the explained variance in ratings. The results indicated that halo error in students' ratings of their learning and development during college affected observed relationships between reports of college experience and gains. For freshmen, a comparison of traditional and halo models revealed that halo error tended to mask differential effects of college experience. Although the results were less pronounced for seniors, the results still indicated a lack of differentiation in traditional models of college effects. Six data tables and two figures are appended. (Contains 48 references.) (MDM)

ED 424 802

THE CONSTANT ERROR OF THE HALO IN EDUCATIONAL OUTCOMES RESEARCH

Gary R. Pike
Director, Student Life Studies
1 Parker Hall
University of Missouri-Columbia
Columbia, MO 65211
Telephone: (573) 882-2887
Fax: (573) 884-4936
Email: slspike@showme.missouri.edu

HE 031 597

Paper presented at the annual meeting of the Association for Institutional Research, Minneapolis, MN, May 1998.

for Management Research, Policy Analysis, and Planning

This paper was presented at the Thirty-Eighth Annual Forum of the Association for Institutional Research held in Minneapolis, Minnesota, May 17-20, 1998.
This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of AIR Forum Papers.

Dolores Vura
Editor
AIR Forum Publications

Abstract

Students' reports of their learning and development play an important role in research and assessment in higher education. Assessment research frequently asks students questions about gains made during college to identify dimensions of gains and then examines relationships between college experiences and gains. A growing body of research suggests that correlations between ratings of gains and college experiences may be an artifact of a constant error of the halo. The present research examines whether halo error underlies Students' self reports of gains, the significance of the halo error, and the effect of halo error on relationships between college experiences and educational outcomes. Results confirm that halo error may be an important component in students' ratings of their learning and development. Moreover, halo error may obscure relationships between college experiences and educational outcomes.

THE CONSTANT ERROR OF THE HALO IN EDUCATIONAL OUTCOMES RESEARCH

Faced with a highly competitive and sometimes hostile environment, colleges and universities have increasingly relied on the assessment of students' educational outcomes to provide information for internal improvement and to provide data on institutional effectiveness for external audiences (Ewell, 1991; Sanders & Chan, 1996). Surveys of enrolled students play a prominent role in assessment research because they provide information that satisfies the twin needs of quality improvement and accountability (Banta, 1988; Sanders & Chan, 1996; Williford & Moden, 1993). In fact, at least two national reports have recommended that surveys be an integral part of a national assessment of college student learning (Ewell, Lovell, Dressler, & Jones, 1994; Halpern, 1994).

Research on the appropriateness of using surveys as part of a national assessment of college student learning has found that surveys can serve as proxies for achievement test scores and provide important information about educational processes that are related to educational outcomes (Pike, 1995, 1996). This same research notes that surveys are not without their limitations. Pike found that students' responses to survey questions about gains made during college were influenced by a strong, general gains factor. Furthermore, the findings suggested that the general factor might mask important differential relationships between students' reported college experiences and their perceptions of gains during college. In an earlier study, Pike (1993) found evidence of the same general gains factor in surveys of enrolled students and alumni. He suggested that this general factor might represent Thorndike's (1920) constant error of the halo (also termed halo error). The present research examined whether halo error influenced students' ratings of their learning and development during college and whether halo error masked important relationships between students' reported college experiences and their perceptions of gains during college.

Background

The Use of Surveys in Assessment Research

Surveys of enrolled students and alumni play an important role in many campus assessment programs. The ACE/Winthrop study (1990), for example, found that almost two-thirds of the colleges and universities involved in outcomes assessment made extensive use of survey research. Similar results have been reported by the Clearinghouse for Higher Education Assessment Instruments (Bradley, Draper, & Pike, 1993). Survey research has also been a critical component in the National Study of Student Learning (NSSL). The NSSL has used surveys to document current levels of student attainment and to identify those aspects of the college experience that are related to student learning and development (Pascarella, Whitt, Nora, Edison, Hagedorn, & Terenzini, 1996).

Many of the surveys currently used in assessment and educational research present students with a list of educational outcomes (e.g., the ability to write effectively or the ability to use mathematics in everyday life), then ask the students to evaluate the extent to which they have made gains in the various outcome domains during college. In addition to questions about gains, many surveys ask students to report on their levels of involvement in various campus activities and use of university programs and services (e.g., involvement in clubs/organizations and use of the library).

The growing popularity of survey research is directly related to the utility of surveys in providing summative data about student learning that can be communicated to external constituents and in providing formative data about the quality of students' educational experiences (Ewell, Lovell, Dressler, & Jones, 1994). Perhaps most important, these surveys provide information about the relationships between students' college experiences and their learning and development (Williford & Moden, 1993). This relational data is particularly relevant for quality improvement initiatives because it can be used to identify how changes in students' college experiences can enhance their learning and development (Pike, 1995).

The ability to use surveys to collect summative data for accountability and formative data for improvement has prompted at least two national reports to recommend that survey research play an important role in a national assessment of college student learning (Ewell, Lovell, Dressler, & Jones, 1994; Halpern, 1994). In their review, Ewell and his colleagues noted that surveys of currently enrolled students provide a significant source of information for research on college impact, and that the results of these surveys have been used to inform national policies. They concluded that students' reports of their gains during college could serve as proxies for a direct assessment of college student learning and also provide important information about levels of student involvement, frequency of faculty-student and peer interaction, and quality of instruction.

As part of a federal study design workshop, Halpern (1994) also recommended the use of survey research both to provide baseline data about college outcomes and to guide the development of a direct assessment of student learning. She noted that the results of a skills survey would provide critical information about the relationships between process variables (e.g., emphasis on teaching and class size) and the development of students' critical thinking skills.

In two studies, Pike (1995, 1996) examined the appropriateness of using students' self reports of college experiences and gains during college in a national assessment of student learning. In the first study, Pike (1995) examined the relationships between students' scores on the College Basic Academic Subjects Examination (College BASE) and their self-reports of gains on a locally developed survey modeled after Pace's (1990) College Student Experiences Questionnaire (CSEQ). Multitrait-multimethod analyses revealed that both College BASE scores and students' reports of their gains in learning and development measured the same constructs. However, Pike also found evidence of a general, method-specific, measurement factor underlying survey responses. In the second study, Pike (1996) examined the relationship between students' College BASE scores and self reports of learning that were developed using the test specifications for College BASE. Results provided even stronger support for the claim that both self reports and test scores measured the same outcomes. Again, however, Pike found evidence of a general factor underlying students' ratings of their learning and development. Based on the results of both studies, Pike (1996) concluded that the presence of a general factor underlying self reports could inflate relationships among educational outcome domains. He also noted that the presence of a general gains factor might mask important relationships between educational outcomes and college experiences. This latter effect could be particularly damaging to quality improvement efforts because they depend on correctly identifying relationships between students' college experiences and their educational outcomes.

Evidence of generalized ratings of learning and development is not limited to research by Pike. Numerous studies have found evidence of positive intercorrelations among outcome domains that is indicative of a higher-order general factor. For example, Terenzini, Pascarella, and Lorang (1982) developed a self-report measure of student learning and development that represented four domains: personal growth, academic process, academic content, and future preparation. The researchers reported that the correlations among factor scores ranged from .37 to .58. Likewise, research using the College Student Experiences Questionnaire has consistently found that five domains underlie students' reports of their gains during college

(Kuh, Vesper, Connolly, & Pace, 1997; Pace, 1987). This research found moderate positive correlations across all domains, with the development of intellectual skills being most highly correlated with other learning and development dimensions. Similar results have been reported using data from the ACT Alumni Survey (Graham & Cockriel, 1989; Valiga, 1982) and the alumni survey used by public colleges and universities in Tennessee (Pike, 1990, 1992).

Evidence that generalized perceptions of gains may mask relationships between college experiences and gains can be found in research showing that students' perceptions of gains tend to be positively correlated with each other and with reports of their college experiences. In a study of almost 2500 former students who responded to Tennessee's state alumni survey, Pike (1990), reported that gain dimensions were positively correlated and that positive relationships between self reports of college experiences and perceived gains dominated the findings. Likewise, Davis and Murrell (1993) examined the CSEQ scores of students at 11 institutions and found a general pattern of positive relationships between quality of effort and reported gains in learning and development. Indirect evidence that generalized perceptions of gains may mask important relationships between college experiences and gains comes from the work of Pascarella and Terenzini (1991). Based on their review of 20 years of research, much of it survey research, related to 10 outcome domains, Pascarella and Terenzini concluded that the effects of college were cumulative and generalized.

Although previous research has not focused specifically on the general factor underlying students' self reports of gains in learning and development, a previous study by Pike (1993) does provide important information about the nature of this general factor. Using data from individuals who completed surveys as seniors and again two years after graduation, Pike examined the extent to which reports of learning and development during college were related to satisfaction with college. Rather than indicating a dominant relationship between satisfaction and reported gains, the research suggested that students' generalized perceptions of their learning and development during college tended to influence their ratings of gains in specific outcome domains (e.g., verbal, quantitative, and personal outcomes). Pike noted that this tendency toward a general rating was similar to the halo error in ratings described by Thorndike (1920). He suggested that if trained raters have a tendency toward generalized ratings of others, then college students and alumni might also have a tendency toward generalized ratings of their own gains during college.

Although Pike's (1993) study suggested that halo error suffuses students' ratings of learning and development during college, educational outcomes research has not attempted to assess the nature and the consequences of halo error in self reports of learning and development. While an exhaustive review of research on halo error is beyond the scope of this study, an overview of the literature concerning halo error is helpful in understanding consistencies in students' ratings of their learning and development.

Halo Error

The tendency toward consistency in raters' evaluations of others was documented nearly a century ago by Wells (1907) who examined ratings of authors using multiple literary criteria and an overall merit criterion. Wells found that there was a tendency for raters to allow their overall ratings of merit to color their specific ratings of literary qualities. Thorndike (1920) examined data on employee performance appraisals and ratings of Army officers. He found that evaluations of individuals across a variety of performance dimensions tended to be highly and uniformly correlated. He concluded that high correlations among rating scales were the product of a "constant error toward suffusing ratings of special features with a halo belonging to the individual as a whole" (Thorndike, 1920, p. 25). In other words, raters tended to rely on general perceptions, even when they were asked to evaluate specific characteristics of individuals. Based on his review of research, Cooper (1981) used the term "ubiquitous halo" to reflect the fact that research findings have indicated that halo error suffuses all types of ratings across a variety of contexts.

In his 1920 study, Thorndike noted that the magnitude of halo error seemed to be surprisingly large. Consistent with Thorndike's results, Symonds (1925) found evidence of a large halo error in teachers' ratings of pupils. He noted that halo error increased the magnitude of positive correlations among rating scales from 0.17 to 0.25. Symonds also noted that the effects of halo error were most pronounced when the traits being evaluated were abstract or difficult to measure.

Cooper (1981) noted that theory and research since Thorndike has tended to view halo effects as sources of error to be minimized. Significantly, Cooper found little evidence to suggest that halo error has produced inaccurate ratings. In fact, four studies he reviewed found that halo error was positively related to the accuracy of ratings. Cooper (1981) referred to this as a paradox in the rating process.

At least two reasons have been advanced to explain how halo error in ratings can improve accuracy. The first line of reasoning suggests that there are at least two types of halo—true and illusory. True halo, or true consistency in ratings, occurs because the behaviors of the ratees are, in fact, related. For example, performance in math truly may be related to performance in science. Illusory halo, on the other hand, inheres in raters and reflects raters' inability to differentiate among specific characteristics of the individuals being evaluated. Consistency in ratings can improve the accuracy of ratings if the halo is true. Only if the halo is illusory will consistency produce inaccurate evaluations (Murphy, Jako, &

Anhalt, 1993). A second reason that halo can improve the accuracy of ratings is statistical. If the focus of the evaluation is on ranking or rating individuals, halo error enhances reliable variance among individuals, which will improve the accuracy of ratings from a purely statistical point of view (Murphy, Jako, & Anhalt, 1993).

Murphy and Balzer (1986) and Murphy, Jako and Anhalt (1993) noted that halo error can create problems when the focus of the research is on the rating criteria and the relationship of specific ratings to external variables. Significantly, the various dimensions of student outcomes are the focus of most assessment and quality improvement research. Thus, halo error may not be a problem in individual performance evaluations, but it can be a serious problem in educational assessment research focusing on the differential relationships between students' college experiences and their learning and development.

Two general approaches have be used to identify and control for the effects of halo error. The first approach uses the training of raters to identify and control for halo error. Research indicates that greater familiarity with rating scales, an understanding of the conceptual underpinnings of those scales, and increased time for observation aids raters in overcoming rating errors such as halo (Bernardin & Pence, 1980). While evidence of the usefulness of rater training is widely available, Landy and Farr (1980) noted that efforts to control for halo error in ratings have not been entirely successful.

The second approach to controlling for halo error is statistical and many of the techniques that have been employed make use of correlational procedures (Fisicaro & Lance, 1990; Lance & Woehr, 1986; Landy, Vance, Barnes-Farrell, & Steele, 1980; Mossholder & Giles, 1983; Myers, 1965). While specific statistical techniques differ from one author to another, most correlational approaches begin by identifying the dominant general factor representing halo error. The effects of the general factor on specific factors is then removed statistically to provide halo-free evaluations.

Drawing on the work of Schmid and Lehman (1957) and Gold and Muthén (1991), Pike proposed comparing a hierarchical factor model to a traditional model of educational outcomes in order to identify halo error. He further suggested that the hierarchical model should be used to statistically control for the effects of halo. Figure 1 presents a traditional factor model of educational outcomes. In this model relationships among students' ratings of learning and development ($V_1 \ldots V_9$) are the product of a series of factors or outcome domains ($F_1 \ldots F_3$). In this model, the various outcome domains are assumed to be intercorrelated.

Insert Figure 1 about here

The hierarchical factor model depicted in Figure 2 assumes that relationships among specific ratings ($V_1 \ldots V_9$) are the product of a general halo factor (H) and specific performance factors ($S_1 \ldots S_3$). Due to constraints that must be imposed on the model in order for it to be identified, the loadings of observed ratings on the halo factor are fixed to unity, while the variances of the specific performance factors are also fixed at unity. In addition, educational outcome dimensions are assumed to be uncorrelated. One desirable feature of this model is the fact that the effects of halo error is uniform across ratings. Another desirable feature of the model is that the variance in the specific performance factors is uncontaminated by halo error. This model was used in the present research both to identify halo error in students' ratings of their learning and development and to evaluate the effects of halo error on the relationships between students' reported college experiences and their perceptions of their learning and development during college.

Insert Figure 2 about here

Criteria for Evaluating the Effects of Halo Error

The three questions that guided Pike's (1993) research provide a set of criteria for evaluating the nature and consequences of halo error in students' ratings of their learning and development during college. The first and most basic criterion focused on whether Pike's (1993) model of halo error provides an acceptable representation of the relationships between observed ratings. That is, can the covariances among students' ratings of their learning and development be satisfactorily explained by a general halo factor and specific performance factors (i.e., the hierarchical factor model)? With regard to this criterion, it is important to understand that statistical analyses seldom provide a definitive answer to the question of which model, traditional or halo, is correct (Mulaik & Quartetti, 1997). Instead, statistical tests provide an indication of the reasonableness of the halo model, recognizing that the traditional model may also provide a reasonable representation of the data.

The second criterion for evaluating halo error focused on the magnitude of the halo error. In other words, does halo error account for a meaningful proportion of the variance in students' ratings of their learning and development during college? Although meaningful is a relative term, Symonds noted that halo error increased correlations among specific rating scales by as much as 0.25. In Symonds research the increase in the magnitude of the correlation was nearly as great as the original correlation. This means that halo error explained nearly as much of the covariance as did the true relationship. While requiring that halo error account for as much covariance as true relationships among gain factors is a clear indication of a substantial effect, such a criterion may be unrealistic. In the present research, halo error was deemed to have a substantial impact on relationships among self reported gain scores if the ratio of variance explained by the halo factor compared to the variance explained by the specific (i.e., content) factors was 0.33. The 0.33 criterion is equivalent to saying that the halo factor is one-third the magnitude of the specific factors and that it accounts for one-quarter of the explained variance in students' self reports of gains in learning and development during college.

The third criterion focused on the consequences of halo error in educational assessment research. Specifically, does halo error affect relationships between students' reports of their educational experiences and their ratings of their learning and development during college? Results indicating that halo error obscures relationships between specific college experiences and specific educational outcomes would provide evidence that halo error can be harmful in assessment research. In addition to Pike's three criteria, a fourth question remains to be answered. To what extent is halo error ubiquitous? That is, do the nature and consequences of halo error transcend student characteristics and specific assessment instruments? Answering this question requires a cross validation of results from one study to another.

Research Methods

Participants

The setting for the present research was the University of Missouri–Columbia (MU). MU is a Carnegie Research I institution enrolling approximately 17,000 undergraduate and 5,500 graduate and professional students. In order to examine the generalizability of halo error across different rating scales and types of students, data from two separate studies were evaluated. In the first study, 3,000 first-time freshmen were administered the College Student Experiences Questionnaire (CSEQ) (Pace, 1990) during the Winter 1996 semester. After an initial mailing, a reminder postcard, and a follow-up mailing, 1084 students had returned surveys—a response rate of slightly more than 36%. Complete data were available for 1000 students. An analysis of background characteristics revealed that respondents were more likely to be female (67.9%) than were students who did not respond (52.0%). The mean entering ACT composite score for respondents (25.9) was significantly higher than the mean ACT score for nonrespondents (25.0). Likewise, the cumulative grade point average of respondents (3.03) was significantly higher than the grade point average of nonrespondents (2.71). While statistically significant, these differences accounted for less than 5% of the variance in background measures. No significant differences were found for ethnicity or major.

In the second study, the MU Senior Survey was mailed to approximately 3,000 enrolled seniors in both Winter 1996 and Winter 1997. Samples were drawn so that the same students did not respond in both years. After an initial mailing, a postcard reminder, and a subsequent mailing, slightly more than 900 students returned the survey each year—approximately a 30% response rate. Combining responses from both administrations of the survey, complete data were available for 1,604 seniors. Again, an analysis of participants' background characteristics revealed that respondents were more likely to be female than were nonrespondents (57.9% and 47.1%, respectively). Respondents had a slightly higher mean ACT composite score (25.6) than students who did not respond (25.2). The mean grade point average of respondents (3.10) was also significantly higher than the mean grade point average of nonrespondents (2.93). Again, differences between respondents and nonrespondents accounted for less than 5% of the variance in background measures, and no differences were found for ethnicity or academic major.

Instruments

The CSEQ, which was administered to freshmen in the first study, is based on Pace's (1987) view that students learn what they do. The survey focuses on the quality of student effort and gains made during college (Kuh, Vesper, Connolly, & Pace, 1997). As Ewell and his colleagues noted, one of the strengths of the CSEQ is its utility in identifying relationships between student effort and gains in learning and development (Ewell, Dressler, Lovell, & Jones, 1994).

In the present research, three questions each were used to represent four dimensions of gains. Gains in personal development were represented by questions about understanding yourself, understanding other people, and being able to function as a team member. Gains in science and technology were represented by questions about understanding the nature of science, understanding new scientific and technical developments, and becoming aware of the consequences of new development in science and technology. Gains in intellectual skills were represented by questions about thinking analytically and logically, putting ideas together, and learning independently. The fourth dimension, gains in general education, was represented by questions about seeing the importance of history, broadening appreciation for literature, and becoming aware of different philosophies and cultures. For each gains item on the CSEQ students are asked to rate the extent to which they have gained or made progress. Four response options are provided: very little, some, quite a bit, and very much.

Six college experience constructs were drawn from quality of effort questions in the CSEQ. These constructs focused on use of the library; course effort; involvement in art, music, and theatre; experience in writing; effort in science; and the intellectual content of conversations. Each of the college experience constructs was represented by three questions. The questions used in this research were initially selected based on factor analyses of the individual quality of effort scales (Pace & Swayze, 1992). Factor analyses of the data in this study were used to screen and select items for inclusion in the analyses.

In the second study of MU seniors, three dimensions of gains were examined. Gains in appreciation of diversity were represented by questions about getting along with different people, appreciating different cultures, and understanding different philosophies and cultures. Gains in communication skills were represented by questions about expressing ideas confidently, speaking clearly and effectively, and writing clearly and effectively. Gains in science were represented by questions about understanding and applying scientific principles, understanding new scientific and technical developments, and becoming aware of new applications in science and technology. For each question, students were asked how large a contribution their college experiences had made to their learning and development. Response options were very great, great, moderate, little, and none.

Seven college experience constructs were also included in this study. The constructs were selected based on previous research at MU using similar items (Eimers & Pike, 1997; Pike, Schroeder, & Berry, 1997). The present research included seven college-experience constructs: academic integration, social integration, institutional commitment, external encouragement, affinity of values, peer influence, and faculty influence. Preliminary factor analyses were again utilized to screen the items used to represent college experience constructs.

Data Analyses

Identical sets of data analyses were performed in both studies and paralleled the three-step procedure employed by Pike (1993). Each step in the procedure provided an answer to one of the three criteria for assessing halo error. The final question concerning the generalizability of results was addressed by comparing the results of the two studies in the research.

The most fundamental criterion which halo error must satisfy is that it exists. In order to infer the existence of halo error in students' reports of their learning and development, two models were specified and tested using confirmatory factor analysis procedures in the LISREL 8 computer program (Jöreskog & Sörbom, 1993). The preferred method of dealing with responses to individual questions in a confirmatory factor analysis is to first calculate the matrix of polychoric correlations among the items and then analyze the matrix using weighted least squares procedures (Jöreskog & Sörbom, 1993). Because the magnitude of polychoric correlations is uniformly greater than product moment correlations or covariances, it was believed that use of polychoric correlations might artificially contribute to the existence of halo error in the analyses. Consequently, least squares covariances were calculated and analyzed using weighted least squares estimation procedures.

The first model that was specified included questions about students' perceptions of their gains during college. In the model covariances among responses to the gains questions were assumed to be the product of a series of correlated outcomes factors. For the first study of freshmen, four outcomes factors were specified, and for the second study, three outcomes factors were specified. The second model included the same specific factors as in the traditional model, but it also included a halo factor that was assumed to uniformly influence covariances among the gains questions. The extent to which these models accurately represented the data was assessed using the traditional chi-square goodness of fit statistic, the root mean square error of approximation, and the expected cross-validation index (CVI) (Jöreskog & Sörbom, 1993). More familiar incremental fit indices were not used in this research because the data were estimated using weighted least squares techniques and because the models being evaluated were not hierarchically nested (Sugawara & MacCallum, 1993).

Because the chi-square goodness of fit statistic is strongly influenced by sample size, it was expected that the large number of students in both studies, would tend to produce significant chi-square results when non-significant results were desired. In contrast, both the root mean square error of approximation and Browne and Cudeck's (1989) (CVI) are less subject to the effects of sample size (Browne & Cudeck, 1993). According to Browne and Cudeck (1993), the root mean square error of the approximation should generally be less than 0.05 and should not be greater than 0.08. The CVI proposed by Browne and Cudeck (1989) ranges from 0 to infinity with smaller values representing a better fitting model. Two important advantages of the CVI are that it allows comparisons to be made between models that are not hierarchically nested and that it rewards more parsimonious models (Sugawara & MacCallum, 1993; Williams & Holahan, 1994). In this phase of the analysis, both the root mean square error of approximation and the Cross Validation Index (CVI) were the primary measures used to assess the relative fit of the traditional and halo models.

In addition to assessing and comparing the fit of the two models, parameter estimates were used to further substantiate the existence of halo error. In particular, the magnitude of the correlations between outcome domains in the traditional model was of interest because a constant error of the halo could not exist without a general pattern of positive inter-factor correlations. The standardized factor loadings of gains questions on the halo factor also provided an indication of the existence of halo error. As with interpretation of traditional factor analysis results, loadings of .40 or greater should be considered substantively important.

Presuming that the existence of halo error could be inferred from the first step in the data analysis, the second step involved assessing the magnitude of the halo error of student ratings. Gold and Muthén (1991) suggested that index of dimensionality (i.e., the ratio of the variance explained by the general factor to the average variance explained by the specific factor) is a useful method of assessing the substantive importance of the general factor. A value of 0.50 for the index of dimensionality would indicate that the general halo factor accounts for one-half as much variance as does the specific factor, while a value of 1.00 would indicate that both the halo and specific gain factors account for equal amounts of variance. Index of dimensionality values greater than 1.00 would indicate that halo error accounts for more of the variance than does the specific factor. For the purposes of this research, values of 0.33 or greater were taken as evidence of substantial halo error in students' ratings of their learning and development during college.

The third step in the data analysis involved evaluating the effect of halo error on relationships between college experience measures and ratings of gains made during college. For this step, both college experience and gains data were included in the analysis. Consistent with procedures employed in the first step, two models were specified. The first (i.e., traditional) model for freshmen included four gain factors and six quality of effort factors. The traditional model for seniors included three gain factors and seven college experience factors. For both studies, all factors in the traditional model were correlated.

The second model evaluated in this phase of the analysis included a halo factor, specific gain factors, and college experience factors. Like the halo model evaluated in Step 1, the halo factor was assumed to have a uniform effect on gains items, and the halo and specific-gain factors were assumed to be uncorrelated. In the halo model, college experience factors were correlated with each other and with the halo and specific-gain factors.

Consistent with procedures employed in Step 1 of both studies, the chi-square goodness of fit test, root mean square error of approximation, and CVI were used to evaluate model fit. Again, because of the large number of participants in both studies, it was expected that chi-square statistics would be less useful in assessing model fit than would the root mean square error of approximation and the CVI. In addition to assessing goodness of model fit, correlations between college experience and gain factors were examined and comparisons made between the two models. These comparisons were used to assess whether the presence of a halo factor substantively altered relationships between experience and gain factors. As a final step in the data analysis, results were compared across the two studies in order to assess the generalizability of findings.

Results

Study I: MU Freshmen

Confirmatory factor analysis of the 12 CSEQ gains questions using the traditional, four-factor, model produced a chi-square goodness of fit statistic of 221.40 with 48 degrees of freedom. Although this chi-square value was statistically significant ($p < 0.001$), both the root mean square error of approximation (0.060) and the cross-validation index (0.28) indicated that the model provided an acceptable representation of the observed data. Further substantiating this conclusion, all factor loadings were statistically significant ($p < 0.001$). Squared multiple correlations for the measured variables ranged from 0.28 (broadening appreciation for literature) to 0.81 (understanding new scientific and technical developments), with virtually all of the squared multiple correlations being greater than 0.50.

Table 1 presents the correlations among the four gain factors in the traditional model. All correlations were found to be highly significant

($p < 0.001$). Consistent with previous findings, all correlations were modest and positive. The lowest correlation was found between gains in personal development and gains in science and technology (0.41). In fact, gains in science and technology had the lowest correlations with all other gain factors. Correlations among the personal development, intellectual skills, and general education factors were all greater than 0.65.

Insert Table 1 about here

Analysis of the halo model for the 12 CSEQ gains questions also produced a highly significant chi-square statistic ($\chi^2 = 269.22$; $df = 53$; $p < 0.001$). However, both the root mean square error of approximation (0.064) and the CVI (0.32) indicated that this model also provided an acceptable representation of the observed data. All factor loadings were statistically significant in the halo model ($p < 0.001$). Taken as a whole, these results suggested that a general factor underlies MU freshmen responses to the 12 CSEQ gains questions used in this research, and that the more restrictive halo model provided nearly as accurate a representation of the observed data as did the less restrictive traditional model of the relationships among gains.

The results of the second step of the analysis provided further support for the viability of the halo model. Table 2 presents the indices of dimensionality for the halo factor relative to the four specific gain factors in the model. All of the indices of dimensionality exceed the 0.33 criterion established for this research. To facilitate interpretation of the indices, estimates of the proportion of explained variance accounted for by the halo factor are also included. Relative to the gains in personal development factor, the halo factor accounted for substantially more than half (60%) of the explained variance in the variables representing personal development, while the halo factor accounted for nearly half (47%) of the explained variance in the variables representing gains in science and technology. The halo factor accounted for approximately 70% of the explained variance in intellectual skills measures, and slightly more than 75% of the explained variance in general education measures. Overall, the effects of the halo factor were substantial, with the halo factor having the least effect on what was the most concrete outcomes dimension—gains in science and technology.

Insert Table 2 about here

In the third phase of the analysis traditional and halo models which included both college experience and gain factors were specified and tested. The chi-square goodness of fit statistic for the traditional model was statistically significant ($\chi^2 = 1180.25$; $df = 360$; $p < 0.001$), but both the root mean square of approximation (0.048) and the cross-validation index (1.39) indicated that the model provided an acceptable representation of the data. All factor loadings were also statistically significant ($p < 0.001$). Fit indices for the halo model were virtually identical to those for the traditional model. Again, the chi-square coefficient was statistically significant ($\chi^2 = 1188.98$; $df = 359$; $p < 0.001$), but the root mean square of approximation (0.048) and the cross-validation index (1.40) suggested that the model was acceptable. All factor loadings in the halo model were statistically significant ($p < 0.001$).

Table 3 presents the correlations between the college experience and gain factors for both the traditional model and the halo model. An examination of the correlations between college experiences and traditional gain factors revealed a consistent pattern of positive relationships. For example, correlations between gains in personal development and quality of effort variables ranged from 0.20 (quality of effort in topics of conversations) to 0.37 (quality of effort in writing experiences). Likewise, correlations between gains in intellectual skill development ranged from 0.26 (quality of effort in art, music, and theatre) to 0.47 (quality of effort in writing experiences). Correlations between gains in general education and college experience measures ranged from 0.33 (quality of effort in science and technology) to 0.52 (quality of effort in course learning). Only for gains in science and technology was there evidence of a clear differential effect for college experiences. Reported gains in science and technology was strongly correlated with quality of effort in science (0.65), while correlations between gains in science and the other college experience measures ranged from 0.13 (quality of effort in art, music, and theatre) to 0.29 (quality of effort in library experiences).

Insert Table 3 about here

Greater differentiation was found in the correlations between college experience measures and the gain factors that included a measure of halo error. First, all of the college experience measures except quality of effort in art, music, and theatre had significant positive correlations with the halo factor. In the case of personal development, both quality of effort in science and technology and quality of effort in topics of conversation were unrelated to gains. The gains in science and technology factor was positively correlated with all of the college experience measures, but only quality of effort in science and technology had a substantial effect on gains. Interestingly, correlations between quality of effort measures and gains in intellectual skills and gains in general education tended to be larger in the halo error model than in the traditional model. Quality of effort in science and technology was negatively related to gains in intellectual skills and gains in general education, although these correlations were not statistically significant.

Study II: MU Seniors

Confirmatory factor analysis of the nine gains questions from the MU Senior Survey using the traditional model of educational outcomes produced a chi-square goodness of fit statistic of 126.75 (df = 24; $p < 0.001$). While this value was statistically significant, both the root mean square error of approximation (0.052) and the cross-validation index (0.11) suggested that the traditional model provided an acceptable representation of the observed data. Also, all factor loadings in the model were statistically significant. Squared multiple correlations for the measured variables ranged from 0.35 for getting along with people from different ethnic and cultural groups to 0.82 for understanding new scientific and technical developments.

The results of the second study also provided support for the appropriateness of the halo error model. This model produced a statistically significant chi-square value of 133.66 (df = 26; $p < 0.001$), but both the root mean square error of approximation (0.051) and the cross-validation index (0.11) suggested that the model adequately represented observed covariances among gains questions. Again, all factor loadings were statistically significant ($p < 0.001$), and squared multiple correlations for the gains questions ranged from 0.40 for getting along with people from different ethnic and cultural backgrounds to 0.84 for understanding new developments in science and technology.

Table 4 presents the correlations among the three gain factors from the senior survey. Consistent with results from the first study, all correlations among gain factors were positive and significant, ranging from 0.23 to 0.38. It is important to note, that the magnitude of the correlations among seniors' responses to gains questions was substantially lower than the magnitude of the correlations for freshmen. Indeed, the largest correlation among gain factors for seniors (0.38) was less than the smallest correlation among gain factors for freshmen (0.41).

Insert Table 4 about here

Indices of dimensionality suggested that halo error represented a substantial, but less important component in seniors' estimates of gains than the estimates of gains made by freshmen. These results are summarized in Table 5. While two of the indices exceeded the 0.33 criterion, effects were less pronounced than for freshmen. While indices of dimensionality for freshmen were generally in excess of 1.00, none of the indices for seniors were greater than 1.00. For seniors, halo error accounted for between 22% and 43% of the explained variance in gains.

Insert Table 5 about here

An examination of the relationships between seniors' college experiences and their reported gains also provided limited support for the halo error model. Correlations between college experiences and gain factors for both the traditional and halo error models are presented in Table 6. Consistent with the results for freshmen, college experience factors were positively correlated with all traditional gain factors. Gains in diversity had the lowest correlation with academic integration (0.09) and the highest correlation with peer influence (0.25). Gains in communication skills had the lowest correlation with institutional commitment (0.14) and the highest correlation with peer influence (0.51), while gains in science and technology had the lowest correlation with external encouragement (0.11) and the highest correlation with social integration (0.31).

Insert Table 6 about here

Correlations between college experience measures and gain factors in the halo model showed some differentiation, but not as much as for freshmen. For example, institutional commitment was not related to halo error, and there was little evidence of differential relationships between college experiences and gains when halo error was removed. Correlations between college experiences and gains in diversity ranged from 0.08 for academic integration to 0.22 for institutional commitment. Correlations between college experiences and gains in communication ranged from -0.17 for external encouragement to 0.48 for peer influence. Correlations between college experience measures and gains in science and technology ranged from -0.06 for external encouragement to 0.31 for peer influence.

Discussion

The results of the two studies in this research can be summarized as follows:

- Confirmatory factor analysis of both freshman and senior responses provided support for the existence of a constant error of the halo underlying students' ratings of gains in their learning and development during college. For both freshmen and seniors, the more restrictive halo error model provided an acceptable representation of observed covariances. The fact that correlations among gain factors were consistently positive in the traditional model, coupled with significant factor loadings for the halo model, also supported the appropriateness of the halo error model.

- Analyses also indicated that halo error accounted for a substantial proportion of the variance in students' ratings of gains in their learning and development. For freshmen, halo error generally accounted for more than one-half of the explained variance in students' ratings, while halo error accounted for one-quarter to one-half of the explained variance in seniors' ratings.
- Results of both studies indicated that the presence of halo error in students' ratings of their learning and development during college affects observed relationships between reports of college experiences and gains. For freshmen, a comparison of traditional and halo models revealed that halo error tended to mask differential effects of college experiences. Although results were less pronounced for seniors, results still indicated a lack of differentiation in traditional models of college effects.
- Overall, the results of both studies provided limited support for the generalizability of halo effects. For both freshmen and seniors, the halo error model provided nearly as good a representation of the observed covariances as did the traditional model of gains. However, the contribution of halo error to regularities in students' ratings was much less pronounced for seniors. Similarly, comparisons of correlations between college experiences and gains revealed that halo error was less of a factor for seniors.

While the results of this research may have important implications for assessment and educational research, care should be taken not to over generalize the results of the two studies. First and foremost, this research was limited to students at a single university. Had the research been conducted at another institution, it is possible that the results would have been different. In all fairness, however, the purpose of this research was not to make generalizations about student performance either at MU or at other institutions. Instead, this research focused on relationships among and between student ratings, and this and other research has found consistent positive relationships among students ratings of gains in learning and development across a variety of institutions. Additional research at a different types of institutions is needed to assess the stability of halo effects across contexts, but the fact that students' ratings of their learning and development show consistent positive relationships is well established.

A second limitation of the present research involved the questions selected for inclusion in both studies. While the items selected were intended to be representative of the types of questions included on most surveys of college outcomes, it is still possible that different types of questions would have produced different results. In addition, the use of different items for freshmen and seniors limited generalizations about changes in halo effects over time. In this research, it is simply not possible to say with certainty whether weaker effects for seniors represent differences in items or developmental differences in students. Clearly additional research is needed to evaluate the stability of halo error across outcome domains and over time.

Perhaps the most important limitation of the present research is that it was not possible to ascertain whether the observed regularities in students' ratings of gains in learning and development were the product of true or illusory halo. That is, it was not possible to determine whether consistency in ratings was due to the fact that outcome dimensions were related or whether regularities were due to the inability of respondents to clearly differentiate among the outcome domains. Therein lies the rub. Whether consistency in students' ratings of gains is due to relationships between the outcome domains being rated or whether it is due to the inability of raters to adequately distinguish between outcome domains is fundamental to the validity of the claim that the effect of college on students is general and cumulative. If halo is "true," then such claims are justified. If halo is "illusory," then evidence that college outcomes are general and cumulative may be the result of limitations in the ability to measure those outcomes.

While research to distinguish between true and illusory halo is critical for assessment research, it will not be easy. In psychology, research on halo error has attempted to manipulate the "true" nature of the individual being rated. Consistencies in raters' evaluations are true halo when the evaluations converge with the actual characteristics of the individual being rated. Halo is illusory when consistencies do not converge with the characteristics of the individual. It is difficult to see how this experimental approach can be applied to the assessment of college outcomes. Certainly random assignment to groups is not possible. Using test scores as an indicator of true gains would bring halo research full circle to Pike's (1995, 1996) studies showing that a general factor underlies self reports and that this general factor is unrelated to test scores. However, Pike reported that test scores were also suffused with a general factor and that this factor was unrelated to the general factor underlying self-reported gains. Which is the "true" general effect?

Despite these limitations, the present research has important implications for assessing students' educational outcomes using self-reported gains in learning and development during college. First, these results suggest that researchers should exercise caution in interpreting students' reports of their learning and development. Given strong evidence that there are consistencies in students' evaluations of their learning and development, and given ambiguous evidence concerning the validity of these consistencies, the prudent course would seem to be to carefully examine relationships between college experiences and educational outcomes using both traditional and halo models. If halo effects are small, they may be discounted. If they are large, assessment practitioners will need to draw on information from other assessment methods to confirm or disconfirm survey results.

The results of this research also indicate that the hierarchical factor model can be useful in assessment research. One particular strength of this model is its ability to partition the variance in students' ratings of their learning and development into general and specific outcome domains. In the present research, the hierarchical model was able to differentiate among college effects in reasonable ways. For example, quality of effort in science and technology was more clearly related to gains in understanding science and technology in the hierarchical model than in the traditional outcomes model.

The hierarchical model also may be useful in examining data from objective measures of student performance (e.g., standardized test scores). As previously noted, research has shown that objective measures of college outcomes tend to be suffused by a general factor. Application of a hierarchical model to this data would allow researchers to partition variance in test scores into a general factor and specific outcomes paralleling the content domains of the test. Indeed, research by Pike (1992) found that the general factor identified by a hierarchical factor analysis of test scores was strongly related to the entering ability of the students, but unrelated to patterns of course taking. In contrast, the specific factors were related to patterns of coursework and unrelated to entering ability. Pike's findings suggest that the hierarchical model may provide researchers with an opportunity to assess the value added by college without resorting to longitudinal research designs.

The findings of the present research may also provide support for the claim that improving students' self assessments is an important outcome of college. In the present research, halo error was more pronounced for freshmen than it was for seniors. In addition, halo error was more likely to mask differential relationships between college experiences and gains for freshmen. Research on the training of raters has shown that familiarity with rating scales and increased opportunities for observation can at least partly offset errors such as halo. College experiences provide students' with an opportunity to be trained in self rating as they are evaluated by others and encouraged to reevaluate their own work in light of others' evaluations. Differences in results for freshmen and seniors in the present research seem to be consistent with a training perspective. Presumably seniors have had greater opportunities to be evaluated and to evaluate their own work than freshmen. Hence they are better trained and less subject to halo error than freshmen. From a training perspective, a key element in efforts to minimize illusory halo is the quality of feedback provided to students and the opportunities students have to reflect on their performance. Clearly, assessment to serve the learner (e.g., the Alverno model of assessment and self assessment) can have positive effects on student learning and on the accuracy with which student learning is measured.

Conclusion

For a variety of reasons, colleges and universities make extensive use of survey research about gains in learning and development during college to assess and improve the quality and effectiveness of their education programs. The present research suggests that assessment professionals should exercise caution when using ratings of gains to differentiate among outcomes. The difficulty is that drawing distinctions among outcomes frequently is essential for improving educational quality. Limitations in the ability to measure outcomes precisely are not a sufficient reason to abandon assessment. As Curry and Hager (1987, p. 57) observed: "To assess outcomes, we must overcome enormous problems of procedure and analysis, but we cannot refuse to look at what the instruments enable us to see."

References

ACE/Winthrop national assessment results release (1990, Winter). The Exchange, 1(5): 1.

Banta, T. W. (1988). Assessment as an instrument of state funding policy. In T. W. Banta (Ed.), Implementing outcomes assessment: Promise and perils (New Directions for Institutional Research Series, No. 59, pp. 81-94). San Francisco: Jossey-Bass

Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: New response sets and decreasing accuracy. Journal of Applied Psychology, 65, 60-66.

Bradley, J. L., Draper, G. W, & Pike, G. R. (1993). Assessment measures: Survey of assessment practice by the Clearinghouse for Higher Education Assessment Instruments. Assessment Update: Progress, Trends, and Practices in Higher Education, 5(5): 14-15.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. Multivariate Behavioral Research, 24, 445-455.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bolen & J. S. Long (Eds.), Testing structural equation models (pp. 136-162). Newbury Park: Sage.

Cooper, W. H. (1981). Ubiquitous halo. Psychological Bulletin, 90, 218-244.

Curry W., & Hager, E. (1987)Assessing general education: Trenton State College. In D. Halpern (Ed.), Student outcomes assessment: What institutions stand to gain (New Directions for Higher Education Series, No. 59, pp. 57-65), San Francisco: Jossey-Bass.

Davis, T. M., and Murrell, P. H. (1993). A structural model of perceived academic personal and vocational gains related to college student responsibility. Research in Higher Education, 34, 267-290.

Eimers, M. T., & Pike, G. R. (1997). Minority and nonminority adjustment to college: Differences or similarities? Research in Higher Education, 38, 77-97.

Ewell, P. T. (1991). To capture the ineffable: New forms of assessment in higher education. In G. Grant (Ed.), Review of research in education (Vol. 17, pp. 75-125). Washington, D. C.: American Educational Research Association.

Ewell, P. T., Lovell, C. D., Dressler, P., and Jones, D. P. (1994). A preliminary study of the feasibility and utility for national policy of instructional "good practice" indicators in undergraduate education. Washington, D.C.: U.S. Government Printing Office.

Fisicaro, S. A., & Lance, C. E. (1990). Implications of three causal models for the measurement of halo error. Applied Psychological Measurement, 14, 419-429.

Gold, K. F. & Muthén, B. O. (1991, April). Extension of covariance structure analysis: Hierarchical modeling of multidimensional achievement data. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Graham, S. W., & Cockriel, I. (1989). College outcomes assessment factors: An empirical approach. College Student Journal, 23, 280-288.

Halpern, D. F. (1994). A national assessment of critical thinking skills in adults: Taking steps toward the goal. In A Greenwood (ed.), The national assessment of college student learning: Identification of the skills to be taught, learned, and assessed (pp. 24-64). Washington, D. C.: U. S. Government Printing Office.

Jöreskog K. G., and Sörbom, D. (1993). LISREL 8. Chicago: Scientific Software.

Kuh, G. D., Vesper, N. Connolly, M. R., & Pace, C. R. (1997). College Student Experiences Questionnaire: Revised norms for the third edition, 1997. Bloomington, IN: Center for Postsecondary Research and Planning, Indiana University Bloomington.

Lance, C. E., & Woehr, D. J. (1986). Statistical control of halo: Clarification from two cognitive models of the performance appraisal process. Journal of Applied Psychology, 71, 679-685.

Landy, F. J., & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87, 72-107.

Landy, F. J., Vance, R. J., Barnes-Farrell, J. L., & Steele, J. W. (1980). Statistical control of halo error in performance ratings. Journal of Applied Psychology, 65, 501-506.

Mossholder, K. W., & Giles, W. F. (1983). The use of partial correlation to control halo in performance ratings. Educational and Psychological Measurement, 43, 977-984.

Mulaik, S. A., & Quartetti, D. A. (1997). First order or higher order general factor? Structural Equation Modeling, 4, 193-211.

Murphy, K. R., & Balzer, W. K. (1986). Systematic distortions in memory-based behavior ratings and performance evaluation: Consequences for rating accuracy. Journal of Applied Psychology, 71, 39-44.

Murphy, K. R., Jako, R. A., & Anhalt, R. L. (1993). Nature and consequences of halo error: A critical analysis. Journal of Applied Psychology, 78, 218-225.

Myers, J. H. (1965). Removing halo from job evaluation factor structure. Journal of Applied Psychology, 49, 217-221.

Pace, C. R. (1987). CSEQ test manual and norms. Los Angeles: Center for the Study of Evaluation.

Pace, C. R. (1990). CSEQ: College Student Experiences Questionnaire (3rd Edition). Los Angeles: Center for the Study of Evaluation.

Pace, C. R. & Swayze, S. (1992). Psychometric supplement to the CSEQ Third Edition, 1990. Unpublished research report, University of California Los Angeles, Center for the Study of Evaluation.

Pascarella, E. T., and Terenzini, P. T. (1991). How college affects students: Findings and insights from twenty years of research. San Francisco: Jossey-Bass.

Pascarella, E. T., Whitt, E., Nora, A. Edison, M., Hagedorn, L., & Terenzini, P. T. (1996). What have we learned from the first year of the national study of student learning? Journal of College Student Development, 37, 182-192.

Pike, G. R. (1990, April). Dimensions of alumni perceptions of cognitive and affective growth during college. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Pike, G. R. (1991, October). Dimensions of academic growth and development during college: Using alumni reports to evaluate education programs. Paper presented at the annual meeting of the Association for the Study of Higher Education, Boston.

Pike, G. R. (1992, November). General and specific dimensions of educational outcomes: A preliminary investigation using hierarchical confirmatory factor analysis. Paper presented at the annual meeting of the Association for the Study of Higher Education, Minneapolis.

Pike, G. R. (1993). The relationship between perceived learning and satisfaction with college: An alternative view. Research in Higher Education, 34, 23-40.

Pike, G. R. (1995). The relationship between self reports of college experiences and achievement test scores. Research in Higher Education, 36, 1-22.

Pike, G. R. (1996). Limitations of using students' self-reports of academic development as proxies for traditional achievement measures. Research in Higher Education, 37, 89-114.

Pike, G. R., Schroeder, C. C., & Berry, T. R. (1997). Enhancing the educational impact of residence halls: The relationship between residential learning communities and first-year college experiences and persistence. Journal of College Student Development, 38, 609-621.

Sanders, L., & Chan, S. (1996, Spring). Student satisfaction surveys: Measurement and utilization issues. AIR Professional File, No. 59.

Schmid, J., & Lehman, J. M. (1957). The development of hierarchical factor solutions. Psychometrika, 22, 53-61.

Sugawara, N., and MacCallum, R. C. (1993). Effect of estimation method on incremental fit indexes for covariance structure models. Applied Psychological Measurement, 17, 365-378.

Symonds, P. M. (1925). Notes on rating. Journal of Applied Psychology, 9, 456-461.

Terenzini, P. T., Pascarella, E. T., and Lorang, W. (1982). An assessment of the academic and social influences on freshman year educational outcomes. Review of Higher Education, 5, 86-110.

Thorndike, E. L. (1920). A constant error in psychological ratings. Journal of Applied Psychology, 4, 25-29.

Valiga, M. J. (1982, May). Structuring the perceived outcomes of higher education. Paper presented at the annual meeting of the Association for Institutional Research, Minneapolis.

Wells, F. L. (1907) A statistical study of literary merit. Archives of Psychology, 1(7).

Williams, L. J., and Holahan, P. J. (1994). Parsimony-based fit indices for multiple-indicator models: Do they work? Structural Equation Modeling, 1, 161-189.

Williford, A. M., & Moden, G. O. (1993). Using assessment to enhance quality. In T. W. Banta (Ed.), Making a difference (pp. 40-53). San Francisco: Jossey-Bass.

Table 1:

Correlations Among CSEQ Gain Factors

	Personal Development	Science & Technology	Intellectual Development	General Education
Personal	1.00			
Science & Tech	0.41	1.00		
Intellectual Dev.	0.71	0.59	1.00	
General Educ.	0.67	0.50	0.77	1.00

Table 2:

Indices of Dimensionality and Proportions of Explained Variance for the CSEQ Halo Factor

	Index of Dimensionality	Proportion of Variance Explained
Personal Development	1.50	0.60
Science & Technology	0.89	0.47
Intellectual Development	2.36	0.70
General Education	3.03	0.75

Table 3:

Correlations Between CSEQ College Experience and Gain Factors for the Traditional and Halo Models

	Traditional				Halo				
	Person.	Science	Intell.	Gen.Ed.	Person.	Science	Intell.	Gen.Ed.	Halo
Library Experiences	0.31^c	0.29^c	0.43^c	0.43^c	0.26^c	0.19^b	0.59^c	0.61^c	0.19^b
Course Learning	0.31^c	0.27^c	0.46^c	0.52^c	0.26^c	0.17^b	0.54^c	0.90^c	0.20^b
Art, Music, Theatre	0.26^c	0.13^c	0.26^c	0.39^c	0.38^c	0.12^b	0.38^c	0.85^c	0.09
Experience Writing	0.37^c	0.24^c	0.47^c	0.46^c	0.24^c	0.12^b	0.43^c	0.55^c	0.27^c
Science Experiences	0.25^c	0.65^c	0.30^c	0.33^c	-0.03	0.48^c	-0.14	-0.12	0.48^c
Topics Conversation	0.20^c	0.23^c	0.36^c	0.49^c	0.02	0.05	0.22^b	0.60	0.29^c

$^a p < 0.05$; $^b p < 0.01$; $^c p < 0.001$

BEST COPY AVAILABLE

Table 4:

Correlations Among MU Senior Survey Gain Scales

	Appreciation of Diversity	Communication Skills	Science & Technology
Appreciation Diver.	1.00		
Communication Skills	0.23	1.00	
Understanding Science	0.28	0.38	1.00

Table 5:

Indices of Dimensionality and Proportions of Explained Variance for the MU Senior Survey Halo Factor

	Index of Dimensionality	Proportion of Variance Explained
Appreciation Diversity	0.36	0.27
Communication Skills	0.77	0.43
Science & Technology	0.29	0.22

Table 6:

Correlations Between MU Senior Survey College Experience and Gain Factors for the Traditional and Halo Models

	Traditional			Halo			
	Diversity	Communic.	Science	Diversity	Communic.	Science	Halo
Academic Integration	0.09^{c}	0.38^{c}	0.20^{c}	0.08	0.39^{c}	0.13^{a}	0.12
Social Integration	0.21^{c}	0.38^{c}	0.31^{c}	0.14^{b}	0.25^{c}	0.18^{b}	0.29^{c}
Institutional Commitment	0.19^{c}	0.14^{c}	0.15^{c}	0.22^{c}	0.19^{b}	0.15^{b}	-0.01
External Encouragement	0.24^{c}	0.15^{c}	0.11^{c}	0.10	-0.17	-0.06	0.43^{c}
Affinity of Values	0.24^{c}	0.16^{c}	0.22^{c}	0.19^{b}	0.03	0.16^{b}	0.16^{b}
Peer Influence	0.25^{c}	0.51^{c}	0.23^{c}	0.17^{b}	0.48^{c}	0.09	0.31^{c}
Faculty Influence	0.21^{c}	0.17^{c}	0.14^{c}	0.16^{b}	0.10	0.05	0.17^{b}

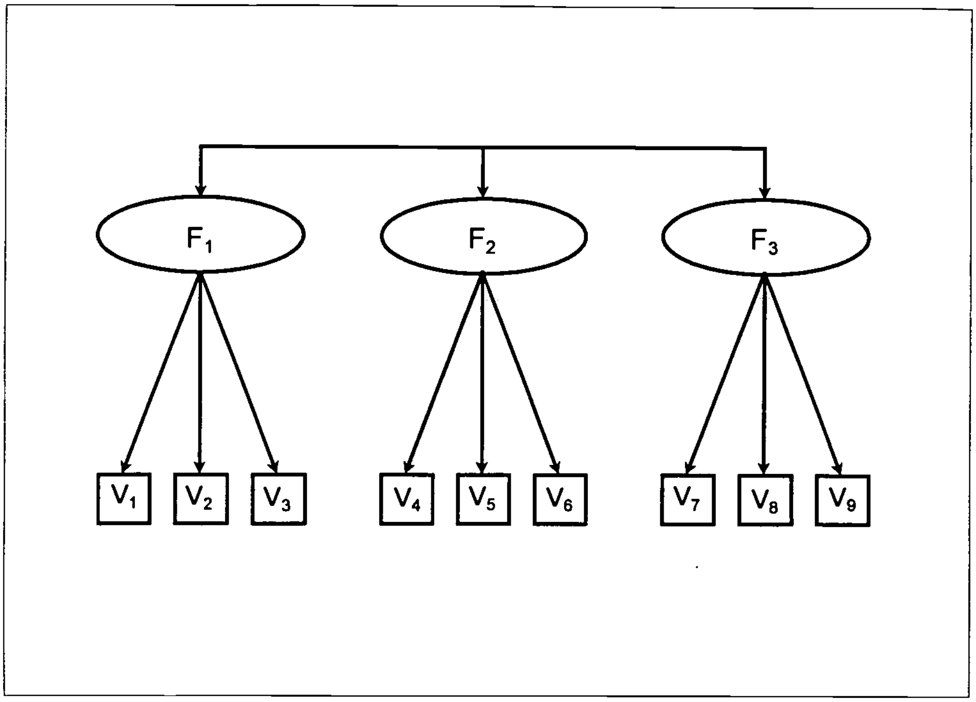
$^{a}p < 0.05$; $^{b}p < 0.01$; $^{c}p < 0.001$

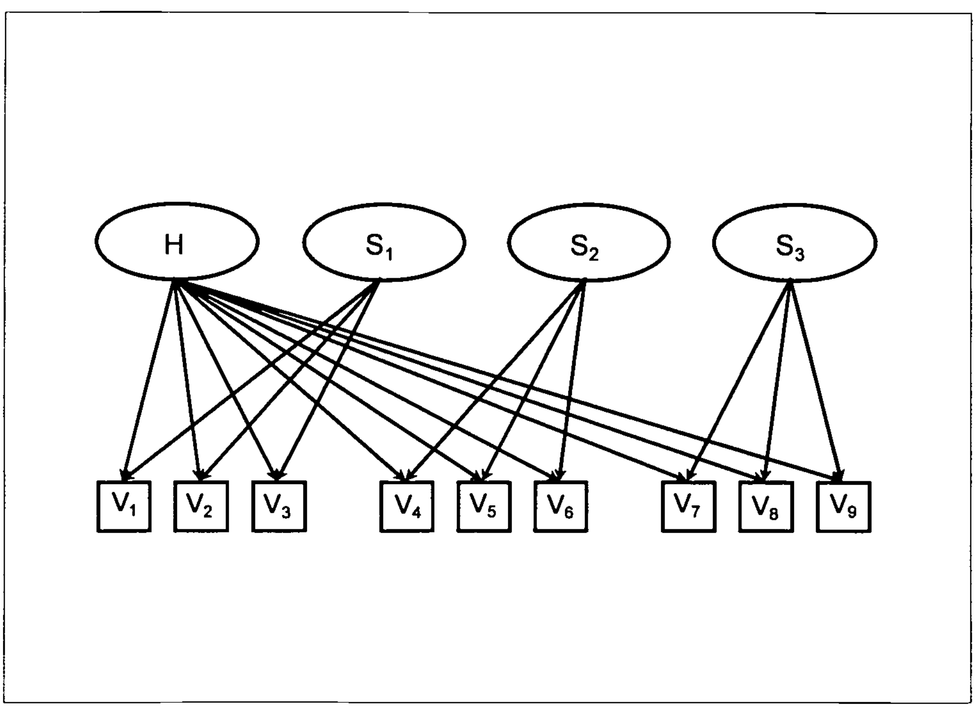
Figure 1:

Traditional Model of Educational Gains

Figure 2:

Hierarchical (Halo) Model of Educational Gains


F1
F2
F3
V1
V2
V3
V4
V5
V6
V7
V8
V9


H
S1
S2
S3
V1
V2
V3
V4
V5
V6
V7
V8
V9

NOTICE

REPRODUCTION BASIS

☒ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").