DOCUMENT RESUME

ED 339 326 HE 025 111

AUTHOR Pike, Gary R.

TITLE Dimensions of Academic Growth and Development During College: Using Alumni Reports to Evaluate Education Programs. ASHE Annual Meeting Paper.

PUB DATE 31 Oct 91

NOTE 37p.; Paper presented at the Annual Meeting of the Association for the Study of Higher Education (Boston, MA, October 31-November 3, 1991).

PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Academic Achievement; Alumni; *College Outcomes Assessment; *Construct Validity; Data Analysis; *Educational Quality; Graduate Surveys; Higher Education; *Program Effectiveness; Research Design; *Research Methodology; *Student Development; Validity

IDENTIFIERS *ASHE Annual Meeting; University of Tennessee Knoxville

ABSTRACT

This study attempted to validate the use of academic growth and development items from Tennessee alumni surveys as measures of program quality and effectiveness at the University of Tennessee (UTK), Knoxville. The argument is made that it is essential that the instruments used to assess students educational outcomes be valid measures of the goals of the education program being evaluated and that the empirical structure of assessment data reflect the structure of the outcomes being measured as well as being sensitive to the educational experiences of students. The validation methodology focused on three aspects of construct validity: (1) construct representativeness; (2) structural fidelity; and (3) criterion relatedness. Survey analysis of two randomly-selected samples of 500 alumni from both 1988 and 1990 revealed that the academic growth and development items on the survey represent approximately 50 percent of the goals of the general education program at UTK, had a stable structure that is generally consistent with the structure of the UTK goals, and were significantly related to college experiences. Contains 40 references. (GLR)

DIMENSIONS OF ACADEMIC GROWTH AND DEVELOPMENT DURING COLLEGE: USING ALUMNI REPORTS TO EVALUATE EDUCATION PROGRAMS

Gary R. Pike
Associate Director
Center for Assessment Research and Development
University of Tennessee, Knoxville
1819 Andy Holt Avenue
Knoxville, TN 37996-4350
(615) 974-2350
FAX: (615) 974-2712
BITNET: PA108968@UTKVM1

Paper Presented at the
Annual Meeting of the Association for the Study of Higher Education,
Boston,
October 31, 1991

Texas A&M University
Department of Educational
Administration
College Station, TX 77843
(409) 845-0393

ASSOCIATION FOR THE STUDY OF HIGHER EDUCATION

This paper was presented at the annual meeting of the Association for the Study of Higher Education held at the Park Plaza Hotel & Towers in Boston, Massachusetts, October 31-November 3, 1991. This paper was reviewed by ASHE 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 ASHE conference papers.

ABSTRACT

Student academic development is widely recognized as a key outcome of a college education. This article reports the results of research designed to validate the use of alumni self-reports of academic development as measures of general education program quality at one institution. The validation criteria focused on three aspects of construct validity: content representiveness, structural fidelity, and criterion relatedness. An analysis of two randomly-selected samples of 500 alumni surveyed in 1988 and 1990 revealed that the academic development items covered approximately half of the target institution's general education goals, had a stable structure that was consistent with the structure of the institution's goals, and were significantly related to reported college experiences.

DIMENSIONS OF ACADEMIC GROWTH AND DEVELOPMENT DURING COLLEGE: USING ALUMNI REPORTS TO EVALUATE EDUCATION PROGRAMS

Student academic growth and development is widely recognized as a key outcome of a college education, particularly those aspects of the college experience referred to as "general education" (Astin 1987, 1991; Miller 1988). Researchers have found that students' perceptions of their academic growth and development during college are related to a variety of positive educational experiences and outcomes, including involvement, quality of effort, persistence, and satisfaction (Astin 1985, 1987; Pace 1988, 1990; Pascarella and Terenzini 1991). Understandably, almost two-thirds of the colleges and universities involved in evaluating and improving their education programs rely on measures of student academic growth and development in their assessment efforts (South Carolina Higher Education Assessment Network 1990). In Tennessee, questions about academic growth and development during college are an integral part of a statewide alumni survey that is used to evaluate the quality and effectiveness of education programs and award millions of dollars in public funds for higher education (Banta 1988).

Despite the importance of the academic growth and development items included in the Tennessee alumni survey, almost no research has been conducted to assess the validity of using these questions to evaluate and improve education programs. In the only study to date, Pike (1990) examined a subset of these questions using data from five institutions and found that the items represented four dimensions: (1) development of verbal skills; (2) development of mathematics skills; (3) growth in knowledge about the humanities; and (4) personal development. He also found that the dimensions were differentially related to academic, social, and job satisfaction. Although suggestive, Pike's research did not directly address questions about the validity of using the academic growth and development items to evaluate education programs.

This paper reports the results of research designed to validate the use of the academic growth and development items from the Tennessee alumni survey as measures of program quality and effectiveness at one institution, the University of Tennessee, Knoxville (UTK). In addition to providing data about the use of a specific set of items at a single institution, this study describes a validation methodology that can be employed to evaluate the use of other measures at other institutions.

VALIDATION METHODOLOGY

The validation methodology used in this study is based on the work of Samuel Messick (1989) which focuses on the accuracy and appropriateness of score interpretation and use, and has been employed in evaluating several paper-and-pencil achievement tests at UTK (Banta and Pike 1989; Pike 1989a, in press; Pike and Banta 1989). Because it focuses on the validity of score interpretation and use, Messick's approach is context specific. What is a valid interpretation or use of one set of items at one institution may not be valid for a different use or for a different institution (Millman 1988).

Validation Context

At UTK, the context for validating the interpretation and use of the academic growth and development questions from the Tennessee alumni survey is defined by the State's performance funding guidelines and by the University's internal use of alumni survey data for strategic planning and program review. The former is an externally-imposed, accountability-based mandate, while the latter is internally driven and improvement oriented.

As early as 1975, the Tennessee Higher Education Commission (THEC) began discussing the possibility of basing a part of state funding for higher education on performance criteria, and in 1983 the Commission established a set of performance funding guidelines which currently provide a financial supplement of slightly more than 5 percent of an institution's budget for instruction based on the results of a series of evaluation activities (Banta 1988; Levy 1986; Pike and Banta 1989). In 1987, the THEC voted to continue the program and established new criteria for evaluating institutional performance in the areas of program accreditation, major field assessment, undergraduate general education outcomes, satisfaction surveys, actions for improvement, and developing or pilot testing new assessment instruments (Banta 1988).

The standard governing satisfaction surveys determines 15 percent of the total possible performance funding award an institution may receive (Banta 1988). Thus far, only the Tennessee alumni survey has been used to award funds via this standard. Questions regarding perceptions of academic growth and development during college comprise more than one-third of the survey items used in the award procedure. The amount of an award is determined by calculating institutional and statewide means for each question on the survey. (Separate statewide means are calculated for two- and four-year institutions.) If an institution's mean is above the statewide mean for a given question, that institution is considered to be successful on that question. The total number of institutional successes is calculated and funds are awarded based on a formula developed by the THEC. (Regarding this process, it is worth noting that the award criteria assumes that academic growth and development items on the alumni survey represent a unidimensional construct and that all items carry equal weight.) At UTK, the total performance funding award based on alumni survey results approaches $1 million.

In addition to gathering data for performance funding, staff at UTK collect and disseminate outcomes information for use in strategic planning and periodic academic program reviews (Banta and Fisher 1989). Data concerning alumni perceptions of academic growth and development during college play an important role in planning and program reviews, particularly as they relate to the University's general education program. At UTK, general education outcomes are assumed to be the product of three factors: (1) the general education curriculum; (2) the curriculum in the major field; and (3) involvement in extracurricular activities (Coordinating Committee on General Education 1981).

When the focus of either internal or external evaluation efforts is on general education at UTK, the goals for that program provide the standard against which outcomes data are compared. In 1979, UTK's Chancellor established a task force to examine the general

education curriculum and identify its goals. The Coordinating Committee on General Education (1981) identified three interrelated sets of general education goals: (1) basic skills; (2) knowledge; and (3) judgments and attitudes. The basic skills domain included goals related to verbal communication, computation, foreign language, computer skills, and problem solving. The knowledge domain included goals covering the topics of aesthetics, the scientific basis for life, technology, history, foreign culture, economics, and the social sciences. The judgments and attitudes domain included goals dealing with values, political and social dynamics, personal wholeness, life-long learning, and experience in learning. These goals were adopted by the University in 1982 and were subsequently used to guide the redesign of the general education curriculum when UTK changed from a quarter to a semester calendar in 1986 (Banta and Pike 1989).

Validation Criteria

Although the results of a validation study are context specific, the theory and methods underlying Messick's approach transcend the context within which it is used. Regarding his approach, Messick has argued that validity is a unitary concept with construct validity at its core. He has not discounted the importance of other types of judgments (e.g., content and criterion-related validity), but has indicated that they must be bolstered by evidence of construct validity. For example, judgments about content validity do not provide evidence about whether an instrument actually measures the domains its content seems to represent. Likewise, evidence of criterion-related validity begs the question of whether the criterion measure is itself a valid indicator of its construct. In both cases, evidence of construct validity is required.

The first step in validating the interpretation or use of an instrument is the identification of the construct the instrument is intended to measure. Once the construct has been identified, specific tests of construct validity can be employed. Loevinger (1957) has identified three components of construct validity, *substantive, structural,* and *external,* that can be used to guide validation research. The *substantive* component of construct validity focuses on instrument content, comparing it to the content of the construct the instrument is assumed to represent. Messick (1989) has termed this facet of validity *content representativeness.*

The *structural* component of construct validity focuses on the empirical structure of a measure and is subdivided by Loevinger into questions of *structural fidelity* and *inter-item structure.* Questions regarding *structural fidelity* focus on the dimensionality of a measure in order to determine if empirical structure is consistent with the assumed structure of the construct. Questions related to *inter-item structure* deal with the internal consistency of the measure. Not mentioned by either Loevinger or Messick, but nevertheless important, is the stability of structure over time.

The *external* component of construct validity corresponds to the traditional concept of criterion-related validity (Loevinger 1957). Messick (1989) has argued that the theory of a construct carries with it implicit and explicit assumptions about the relationship of a construct to other constructs. Evaluating the *external* component of construct validity involves

comparing empirical relationships among measures to the theory of the construct. Thus, the external component of construct validity provides an indication of *criterion relatedness*. Pike (1989b) has argued that when a multidimensional instrument is being used to suggest program improvements, it is important that the dimensions be differentially related to external variables. Otherwise, the dimensions will have little practical utility.

Taken as a whole, the preceding discussion suggests that the following research questions should be used to guide the evaluation of the academic growth and development items from the Tennessee alumni survey:

1. To what extent does the content of academic growth and development items from the Tennessee alumni survey correspond to the content of the general education goals at UTK?

2. To what extent does the structure of the academic growth and development items correspond to the structure of UTK's general education goals, and is the structure of the academic growth and development items both reliable and stable?

3. To what extent are the dimensions underlying the academic growth and development questions on the alumni survey related to external measures of college experiences, such as coursework and involvement?

RESEARCH METHODS

Subjects

The subjects for this research consisted of two randomly-selected samples of 500 alumni each. The first sample was drawn from 1451 UTK alumni who graduated in 1986 and completed the Tennessee alumni survey in 1988. The second sample was drawn from 1501 UTK alumni who graduated in 1988 and completed the alumni survey in 1990. The response rates for the two surveys were 51 and 52 percent respectively.

Approximately 52 percent of the alumni in the 1988 sample were male and 97 percent were white. These alumni were, on average, 26.2 years old when they were surveyed, and 94 percent reported attending UTK full time. Of the 1990 sample, 51 percent were male and 95 percent were white. These alumni averaged 26.9 years of age at the time of the survey, and 92 percent reported that they had attended UTK full time.

Instrument

Academic growth and development items from the alumni survey consisted of twenty-one items, many of which were drawn from an alumni survey developed by the American College Testing Program (1989). These items are listed in Table 1. In 1988 and 1990, UTK alumni were asked to indicate the degree to which their education added to their

skills in each area. Three response options were provided: "very little," "somewhat," and "very much."

In addition to the academic growth and development items, responses to fourteen questions about college experiences at UTK were included in the data analysis. These questions are also listed in Table 1. Six of these questions dealt with library usage ("used the library as a quiet place to read or study," "developed a bibliography or references for a term paper," and "used library reference materials") and involvement in art, music, or theatre ("went to an art gallery or exhibit on campus," "attended a concert or other musical performance on campus," and "saw a play on campus"). Each question had four response options: "never," "seldom," "occasionally," and "often." The remaining eight college experience questions dealt with student-faculty contact ("availability of your advisor," "willingness of your advisor to help," "availability of faculty to help students outside of class," and "availability of faculty to talk informally") and quality of coursework ("quality of courses for providing a good general education," "quality of courses for preparing for employment," "quality of instruction in the major," and "quality of courses in the major in preparing for graduate or professional school"). Again, four response options were provided: "poor," "fair," "good," and "excellent."

Data Analysis

In order to evaluate the content representativeness of the twenty-one academic growth and development items, the author classified each question using the UTK general education goals. Despite the fact that these goals were assumed to be interrelated, each item was initially classified as representing one and only one goal.

Confirmatory factor analysis was used to evaluate the structural fidelity of the academic growth and development items from the alumni survey (Jöreskog and Sörbom 1989). Because it assesses the congruence between a hypothetical model and empirical data, confirmatory factor analysis was particularly appropriate for this phase of the data analysis (James, Mulaik, and Brett 1982; Long 1983; Messick 1989).

In this study, several models were specified and tested for goodness-of-fit to the observed data. Initially, four *a priori* models were evaluated. The first model contained twenty-one uncorrelated measured variables corresponding to the academic growth and development questions on the alumni survey. This model represented a null model against which all other models could be compared (Marsh, Balla, and McDonald 1988). The second model contained a single factor, and all of the academic growth and development items were assumed to have significant loadings on this single factor. This model corresponded to the scoring scheme utilized by the Tennessee Higher Education Commission in making performance funding awards (Pike 1990). The third model posited a three-factor structure representing the three sets of general education goals at UTK. The researcher's classification of academic growth and development items in the content analysis served to define the pattern of factor loadings for this model. The final model contained four factors. The first two factors were derived by dividing the basic skills domain into verbal and quantitative skills,

while the remaining factors were identical to those in the third model. This model was included because research has shown that verbal and quantitative skills are qualitatively different (Lohman, 1989).

Based on results from tests of the *a priori* models, specification search procedures were used to make minor modifications to improve goodness-of-fit (Sörbom 1989). The use of specification searches in confirmatory factor analysis has been criticized by MacCallum (1986) because they may undermine a parsimonious explanation of the data and distort estimates of the factor loadings. However, MacCallum's own research indicates that specification searches can be useful if they are restricted using prior knowledge and limited to four or five modifications (MacCallum 1986; Sylvia and MacCallum 1988). In this study, specification searches were based on an examination of modification indices, component chi-squared values, and significance tests of the factor loadings to insure that only significant changes would be made (Pike 1989b).

Because the stability of the structure of academic growth and development items was an important issue in this study, separate analyses were conducted for the 1988 and 1990 samples. The results were then cross-validated using procedures suggested by Jöreskog (1971) and Alwin and Jackson (1981). Goodness-of-fit tests provided an indication of whether parameters were stable (invariant) across the two samples. Three models of invariance were evaluated: (1) invariance of the factor loadings; (2) invariance of the factor loadings and uniquenesses; and (3) invariance of the factor loadings, uniquenesses, and inter-factor correlations. At a minimum, stability of the factor structure requires that factor loadings be invariant across groups (Marsh and Grayson 1990; McGaw and Jöreskog 1971). Other authors have argued that invariance of uniquenesses, as well as factor loadings, should be required to demonstrate the stability of measurement structure (Byrne, Shavelson, and Muthén 1989). Adding the requirement that inter-factor correlations also be invariant represents the most restrictive model of factor stability.

Using item responses, rather than scale scores, in confirmatory factor analysis can create serious problems when responses are not normally distributed (Bernstein and Teng 1989; Jöreskog and Sörbom 1989). In this study, a preliminary analysis of the growth and development questions revealed that response patterns were not normally distributed and were all negatively skewed. To avoid problems created by departures from normality, matrices of polychoric correlations were calculated for both the 1988 and 1990 samples, and these matrices were analyzed using weighted least squares techniques (Jöreskog and Sörbom 1989).

The relatively large samples sizes in the analyses made it difficult to identify an acceptable model based on traditional chi-squared goodness-of-fit tests (Bentler and Bonett 1980). To overcome this problem, a Tucker-Lewis Index was calculated for each model (Tucker and Lewis 1973). The Tucker-Lewis Index is relatively immune to the effects of sample size and offers the additional advantage of providing a measure of the reliability of the factor solution (Marsh, Balla, and McDonald 1988).

Despite its advantages, the Tucker-Lewis Index tends to encourage researchers to accept overly complex models. That is, the greater the number of free parameters in the model, generally the better the index of model fit (McDonald and Marsh 1990). Because the parsimony of a model was also an important concern, the Parsimonious Goodness-of-Fit Index was used to guide model selection. This index provides a penalty function for freeing parameters and provides a measure of the extent to which a model is able to account for all of the information in an observed variance-covariance matrix (Mulaik et al. 1989).

Evaluating the external component of construct validity involved testing the relationships between alumni reports of college experiences and perceptions of academic growth and development during college. Initially, confirmatory factor analysis was used to identify the structure underlying the college experience questions and to test the invariance of that structure across the 1988 and 1990 samples. For this preliminary analysis, matrices of polychoric correlations were calculated and analyzed using weighted least squares techniques.

Based on the results of the confirmatory factor analysis, a structural equation model was specified and tested. In the structural equation model, the factors underlying the college experience questions were assumed to be exogenous variables, while the factors underlying the academic growth and development questions were assumed to be endogenous variables. Because the stability of the relationships between college experience and academic growth and development variables across the two samples was of interest, a multigroup structural equation model was employed. Here again, matrices of polychoric correlations were calculated and analyzed using weighted least squared techniques.

RESULTS

Content Representativeness

Content analysis revealed that all three sets of UTK's general education goals were represented by the academic growth and development questions from the Tennessee alumni survey. However, coverage of the goals within each domain was uneven. For example, goals related to verbal skills, mathematics skills, science for life, knowledge of foreign cultures, personal wholeness, and life-long learning were represented by two or more questions, while goals concerning foreign language skills, computer skills, technology, history, economics, social sciences, values, political and social dynamics, and experience in learning were not represented by any questions. Overall, eight of seventeen goals (47 percent) were covered by questions from the alumni survey. The number of academic growth and development items measuring each goal is presented in Table 2.

Three of five basic skills goals (60 percent) were represented by academic growth and development items. Three items, "speaking effectively," "writing effectively," and "understanding written information," were classified as measuring verbal skills, while computation skills were measured by "understanding graphic information," "ability to understand mathematical concepts," and "ability to use mathematics in everyday life." The

question "defining and solving problems" represented problem solving and critical thinking skills.

Three of seven knowledge goals (43 percent) were measured by academic growth and development questions. Aesthetics was represented by the question "understanding and appreciating the arts," while the questions "understanding the interaction between people and the environment" and "understanding and applying scientific principles and methods" were classified as measuring the science for life goal. "Getting along with people of different races and ethnic groups," "appreciation of different cultures," and "understanding different philosophies and cultures" were classified as measuring knowledge about foreign cultures.

Only 40 percent of the goals concerning judgments and attitudes were covered by academic growth and development questions from the alumni survey. The items "ability to grow and learn as a person" and "learning on you own" were classified as examples of life-long learning. Six items, "practical skills necessary to obtain employment in your field," "ability to lead or guide others," "ability to adjust to new job demands," "self-confidence in expressing your ideas," "planning and carrying out projects," and "working cooperatively in a group," were classified as examples of personal wholeness. The question concerning practical skills for employment was particularly difficult to classify because none of the general education goals at UTK were employment related. The similarity of this item to the question about adjusting to new job demands was the basis for its classification in the personal wholeness category.

Structural Fidelity

As previously noted, separate confirmatory factor analyses were performed for the 1988 and 1990 samples in order to determine whether the structure of the alumni academic growth and development questions corresponded to the structure of the UTK general education goals. In the final set of analyses, the stability of the factor structure across the two samples was evaluated. Goodness-of-fit results for all three sets of analyses are presented in Table 3.

An examination of the data on model fit for the 1988 sample revealed that moving from the null model to the one-, three-, or four-factor models significantly improved goodness-of-fit, as indicated by the decrease in residual chi-squared values and the improvement in the Goodness-of-Fit Indices. The component chi-squared statistics ($\Delta\chi^2$) also revealed that moving to more complex models resulted in statistically significant improvements at each step in the analyses. Likewise, both the Tucker-Lewis Index and the Parsimonious Goodness-of-Fit Index revealed that a four-factor model provided a better representation of the data than any of the other *a priori* models. However, the Tucker-Lewis Index value for the four-factor model was well below the .90 cutoff suggested by Bentler and Bonett (1980) as an indication of a satisfactory model. Consequently, a specification search was conducted to determine how the pattern of factor loadings should be altered.

An examination of the significance tests (t-values) for the parameters in the four-factor model revealed that all of the factor loadings were statistically significant ($p<.001$) and should be retained. Examination of the modification indices indicated that goodness-of-fit could be improved if question 19, "understanding and applying scientific principles and methods," was free to load on the factor representing quantitative skills, as well as the knowledge factor. Adding this parameter to the model produced a chi-squared improvement of 389.55 ($\Delta df=1$; $p<.001$). The Tucker-Lewis Index for this model was .851, indicating a substantial improvement in model fit, but also indicating a need to continue the specification search.

An examination of the t-values at this point in the specification search revealed that the loading for question 19 on the knowledge factor was no longer statistically significant ($p>.05$). Eliminating this parameter decreased the chi-squared goodness-of-fit value by only 0.20 ($\Delta df=1$; $p>.05$). Both the Tucker-Lewis Index and the Parsimonious Goodness-of-Fit Index increased slightly for this model.

A further examination of the modification indices revealed that the "defining and solving problems" item (question 14) should be free to load on the judgments and attitudes factor, as well as the mathematics factor. This new model produced an improvement in chi-squared goodness-of-fit of 243.50 ($\Delta df=1$; $p<.001$), a Tucker-Lewis Index of .888 and a Parsimonious Goodness-of-Fit Index of .732.

Both the modification indices and t-values indicated that freeing question 12, "understanding graphic information," to load on both the verbal and the quantitative skills factors would produce the most significant improvement in goodness of model fit. The component chi-squared value for this modification was 196.07 ($\Delta df=1$; $p<.001$). The Tucker-Lewis Index and Parsimonious Goodness-of-Fit Index for this model were .917 and .740 respectively. Further model modifications, such as freeing question 17 to load on the judgments and attitudes factor, produced decreases in the Parsimonious Goodness-of-Fit Index.

The second subtable in Table 3 presents the results of the goodness-of-fit tests for the 1990 data. Consistent with the results for the 1988 alumni, analyses of the *a priori* models revealed that moving from the null model to the one-, three-, and four-factor models significantly improved the fit of the model to the observed data. Once again, however, the Tucker-Lewis Index for the four-factor model was below the recommended .90 cutoff, indicating the need for a specification search.

Examination of the modification indices revealed that the greatest improvement in model fit could be achieved by freeing question 19, "understanding and applying scientific principles and methods," to load on the quantitative skills factor. This modification produced a chi-squared improvement of 309.29 ($\Delta df=1$; $p<.001$), a Tucker-Lewis Index of .898, and a Parsimonious Goodness-of-Fit Index of .739.

Following the pattern established in the analysis of the 1988 alumni sample, the t-value for the loading of question 19 on the knowledge factor indicated that eliminating this parameter would do little to reduce overall goodness-of-fit. Although the 6.99 increase in chi-squared

was statistically significant ($\Delta df=1$; $p<.01$), the change in the Tucker-Lewis Index as minimal and the Parsimonious Goodness-of-Fit Index actually increase slightly to .742.

Consistent with the results for the 1988 sample, allowing question 14 to be free to load on the judgments and attitudes factor produced a significant increase in chi-squared goodness-of-fit ($\Delta\chi^2=118.54$; $\Delta df=1$; $p<.001$). Improvements were also noted in the Tucker-Lewis Index (.919) and the Parsimonious Goodness-of-Fit Index (.746). Although this model exceeded the recommended .90 cutoff for the Tucker-Lewis Index, modification indices suggested that freeing question 12 to load on the verbal skills factor would improve goodness-of-fit. The component chi-squared value for this modification was 97.36 ($\Delta df=1$; $p<.001$), and both the Tucker-Lewis Index and the Parsimonious Goodness-of-Fit Index showed measurable improvements. Although freeing question 17 to load on the fourth factor also produced improvements in chi-squared and the Tucker-Lewis Index, the Parsimonious Goodness-of-Fit Index actually declined slightly for this modification.

In summary, analyses for both the 1988 and 1990 samples indicated that model [8] provided the best explanation of the observed relationship among the academic growth and development items. Model [8] was a four-factor model in which "understanding and applying scientific principles and methods" was found to represent quantitative skills rather than knowledge, "defining and solving problems" was found to represent both quantitative skills and judgments and attitudes, and both verbal and quantitative skills were represented by "understanding graphic information."

The third subtable in Table 3 presents the results of the tests of measurement invariance across the 1988 and 1990 samples. As indicated in the subtable, a null model, representing uncorrelated measured variables, provided a very poor explanation of the observed data ($\chi^2=14{,}071.61$; $df=441$; $p<.001$). In contrast, a model in which the pattern of factor loadings (consistent with model [8]) was invariant across the 1988 and 1990 alumni samples provided a much better representation of the data ($\chi^2=1235.58$; $df=362$; $p<.001$). The model representing invariance of the factor pattern produced a Tucker-Lewis Index of .922 and Parsimonious Goodness-of-Fit Indices of .775 and .785 for the 1988 and 1990 samples.

Requiring that factor loadings be invariant across the two samples produced a statistically significant change in chi-squared goodness-of-fit ($\Delta\chi^2=102.33$; $\Delta df=23$; $p<.001$). However, the change in the Tucker-Lewis Index was quite small. Most important, the Parsimonious Goodness-of-Fit Indices for the 1988 and 1990 alumni samples improved substantially. The third model, in which both factor loadings and uniquenesses were invariant, did not produce a significant decrease in chi-squared goodness-of-fit ($\Delta\chi^2=0.07$; $\Delta df=21$; $p>.05$). The Tucker-Lewis Index and Parsimonious Goodness-of-Fit Indices increased substantially. Requiring that inter-factor correlations also be invariant produced a significant change in chi-squared of 23.44 ($\Delta df=6$; $p<.001$), but it did not alter the Tucker-Lewis Index. Parsimonious Goodness-of-Fit Indices for the 1988 and 1990 samples increased measurably.

Table 4 presents the weighted least squares estimates for the factor loadings, uniquenesses, and inter-factor correlations for the model in which all of these parameters were invariant across the 1988 and 1990 samples. Weighted least squares estimates for the factor loadings and inter-factor correlations were all statistically significant ($p < .001$). Also included in Table 4 are the squared multiple correlations (R^2) for the growth and development questions.

An examination of the factor loadings for the four questions comprising the verbal skills domain indicated that verbal skills was most strongly identified with the question concerning "understanding written information" (.922). "Speaking effectively" (.730) and "writing effectively" (.715) had the next highest factor loadings, while the factor loading for "understanding graphic information" was much lower (.428).

The quantitative skills dimension was defined primarily by two questions, "ability to use mathematics in everyday life" (.931) and "ability to understand mathematical concepts" (.874). "Understanding and applying scientific principles and methods" also contributed significantly to this dimension (.687). The question concerning "understanding graphic information had approximately the same loading on the quantitative dimension (.418) as it had on the verbal dimension (.428), while the factor loading for "defining and solving problems" was much lower for the quantitative factor (.295) than for the judgments and attitudes factor (.612).

All five of the questions comprising the knowledge domain had relatively high factor loadings, ranging from .792 to .585. The highest factor loading was for "appreciation of different cultures," and the lowest factor loading was for "understanding and appreciating the a The fact that "understanding the interaction between people and the environment" had the second highest factor loading on the knowledge factor suggested that this was not simply a knowledge of foreign cultures dimension.

With one exception, all of the questions included within the judgments and attitudes domain had factor loadings in excess of .600. The one exception was the question related to "practical skills necessary to obtain employment in your field" (.488). It would seem that the difficulties encountered in trying to classify the question about practical skills for employment during the content analysis carried over into the investigations of structural fidelity.

Criterion Relatedness

The first step in evaluating the criterion relatedness of the growth and development dimensions involved identification of the criterion variables, specifically dimensions of college experiences. Initially, confirmatory factor analysis was used to determine if four dimensions of college experience (library usage, involvement in art, music, and theatre, faculty interaction, and coursework) could account for the relationships among these questions. Separate analyses were conducted for the 1988 and 1990 samples, and then a two-group

confirmatory factor analysis was used to test the invariance of the measurement model across samples. The results of these analyses are presented in the three subtables of Table 5.

The results presented in the first two subtables in Table 5 strongly supported the assumption that relationships among the college experience questions could be explained by four underlying dimensions. For the 1988 sample, the four-factor model produced a residual chi-squared goodness-of-fit statistic of 408.21 ($df=71$; $p<.001$). Although this results was statistically significant, both the Tucker-Lewis Index and the Parsimonious Goodness-of-Fit Index indicated that the four-factor model provided a satisfactory representation of the observed data. Likewise, analysis of a four-factor model for the 1990 sample produced a statistically significant chi-squared value ($\chi^2=399.86$; $df=71$; $p<.001$), but an acceptable Tucker-Lewis Index (.948). The Parsimonious Goodness-of-Fit Index (.754) also indicated that a four-factor model provided a satisfactory explanation of the 1990 data.

The final subtable in Table 5 presents the results of the tests of invariance of the measurement model across the 1988 and 1990 samples. The null model, specifying no significant relationships among the observed variables provided a very poor representation of the observed data ($\chi^2=16,246.66$; $df=196$; $p<.001$). A measurement model in which the pattern of factor loadings (consistent with model [2]) was invariant produced a substantially better fitting model ($\chi^2=808.07$; $df=142$; $p<.001$). Requiring that the factor loadings themselves be invariant across the 1988 and 1990 samples produced a significant increase in chi-squared ($\Delta\chi^2=48.40$; $df=14$; $p<.001$); however, the Tucker-Lewis Index (.945) and the Parsimonious Goodness-of-Fit Indices for the 1988 and 1990 samples (.766 and .767 respectively) improved slightly for this model. Results for a model in which both factor loadings and uniquenesses were invariant across samples did not alter chi-squared goodness-of-fit, and both the Tucker-Lewis Index (.951) and the Parsimonious Goodness-of-Fit Indices for each sample increased substantially (.834 and .836 for the 1988 and 1990 samples respectively). Requiring that inter-factor correlations also be invariant resulted in a chi-squared increase of 71.90 ($df=6$; $p<.001$). Although the Tucker-Lewis Index declined slightly (.948), the Parsimonious Goodness-of-Fit Indices for the two samples increased (.861 and .863 respectively).

Overall, these results indicated that a model describing four dimensions of college experiences provided an acceptable representation of the relationships among the fourteen college experience questions. Furthermore, both the factor loadings and uniquenesses for the four-factor model were invariant across the 1988 and 1990 samples. Less certain is whether the inter-factor correlations should also be considered to be invariant over time.

Table 6 presents the weighted least squares estimates for the factor loadings of the college experience items. Also included in the table are weighted least squares estimates of the uniquenesses and estimates of explained variance for each item. An examination of the factor loadings for the factor representing use of the library revealed that "used library reference materials," was most strongly identified with this dimension (.979), followed by "developed a bibliography or set of references for a term paper" (.756) and "used the library as a quiet place

to read or study" (.558). The three items representing involvement in art, music, and theatre all had remarkably similar factor loadings: .673 for "went to an art gallery or exhibit on campus," .772 for "attended a concert or other musical performance on campus," and .644 for "saw a play on campus."

Four questions defined the dimension representing faculty-student interaction. Most strongly associated with this dimension was "willingness of your advisor to help" (.929), followed closely by "availability of your advisor" (.912). Also strongly related to this dimension were "availability of faculty to help students outside of class" (.883) and "availability of faculty to talk informally" (.856). The four questions representing the coursework dimension also had remarkably similar factor loadings: .783 for "quality of instruction in the major," .781 for "quality of courses in the major in preparing for graduate or professional school," .777 for "quality of courses for providing a good general education," and .740 for "quality of courses for preparing for employment."

Based on the results of the analyses of academic growth and development and college experiences items, a structural equation model was specified and tested. In the model, the four dimensions of college experiences were assumed to influence the four academic growth and development dimensions. The three models were specified and tested. The first model was a null model in which the thirty-five measured variables (fourteen college experience variables and twenty-one academic growth and development variables) were assumed to be unrelated. This model produced an extremely poor representation of the observed data (χ^2=29,637.17; df=1225; p<.001).

The second model used in this phase of the analysis included the measurement structures for the college experience and academic growth and development dimensions. In this model the measurement structure (factor loadings and uniquenesses) was assumed to be invariant across the 1988 and 1990 samples, but the effects of college experiences on academic growth and development were not the same across samples. This model produced a chi-squared goodness-of-fit statistic of 3488.65 (df=1134; p<.001), a Tucker-Lewis index of .910, and Parsimonious Goodness-of-Fit Indices of .854 and .861 for the 1988 and 1990 alumni samples. While requiring that the effects of college experiences on academic growth and development be invariant significantly increased the chi-squared value ($\Delta\chi^2$=66.22; df=16; p<.001), the Tucker-Lewis Index did not change (.910), and the Parsimonious Goodness-of-Fit Indices for the 1988 and 1990 samples increased slightly (.865 and .871 respectively). These results strongly suggested that both the measurement model and the structural equation model could be assumed to be invariant across groups.

Table 7 presents the weighted least squares coefficients representing the effects of college experiences on academic growth and development. Also included are the squared multiple correlations (estimates of explained variance) for the four academic growth and development dimensions. An examination of the coefficients in Table 7 revealed that the development of verbal skills was influenced by three different college experiences: the quality of coursework (.280), library usage (.253), and faculty-student interaction (.077).

involvement in art, music, and theatre was not significantly related to the development of verbal skills (.046).

Three college experience variables also influenced the development of quantitative skills. Again, the strongest influence was found to be quality of coursework (.417). Unlike growth of verbal skills, the remaining three educational experience variables has a negative effect on the development of mathematics skills. Both library usage and faculty-student interaction had significant negative effects on the development of quantitative skills (-.192 and -.065 respectively). Involvement in art, music, and theatre was also negatively related to quantitative skills (-.018), but this effect was not statistically significant.

Growth in knowledge, primarily about he humanities, was influenced most strongly by involvement in art, music, and theatre (.322), followed closely by perceived quality of coursework (.210). Library usage was also significantly related to growth in the knowledge domain (.068). Faculty-student interaction was not significantly related to growth in knowledge (.023).

Alumni reports of their development in the judgments and attitudes domain was most strongly related to the perceived quality of coursework at UTK (.394). Both library usage and involvement in art, music, and theatre were also related to growth in judgements and attitudes, although the size of the effects were much smaller (.051 and .055 respectively). Faculty-student involvement did not make a unique contribution to growth in judgments and attitudes (.006).

Examination of the squared multiple correlations (estimates of explained variance) revealed that a substantial amount of the variability in the four academic growth and development scales remained to be explained. The four college experiences dimensions, at best, explained slightly more than one-third of the variance in the judgments and attitudes domain and, at worst, explained only about 20 percent of the variance in the quantitative skills domain.

DISCUSSION

The results of this research can be summarized as follows:

1. Content analysis of the academic growth and development items from the Tennessee alumni survey indicated that approximately half (47%) of the UTK general education goals were covered by these items. The academic growth and development items provided the best coverage of the UTK goals related to basic skills (60%), while approximately 40 percent of the goals related to knowledge outcomes and judgments and attitudes were covered by questions on the alumni survey.

2. Confirmatory factor analysis revealed that a modified four-factor model of perceived academic growth and development provided the best explanation

of the observed data. This model differed from the UTK general education goals and the results of the content analysis in that the basic skills domain was split into dimensions of verbal skills and quantitative skills. In addition, two of the questions on the alumni survey were found to represent more than one type of outcome. Also, one of the questions ("understanding and applying scientific principles and methods") was found to represent quantitative skills rather than knowledge outcomes. The confirmatory factor analysis also revealed that the measurement structure for the four-factor model was stable across the 1988 and 1990 alumni samples.

3. Analysis of the structural relationships between college experience dimensions and the dimensions of academic growth and development revealed that college experiences were differentially related to perceptions of academic growth and development. The development of verbal skills was positively influenced by quality of coursework, library usage, and faculty-student interaction, while the development of quantitative skills was positively influenced by the quality of coursework and negatively influenced by library usage and faculty-student interaction. Growth in the knowledge domain was positively influenced by involvement in art, music, and theatre, as well as by quality of coursework and library usage. Likewise, development of attitudes and judgment was positively related to coursework, library usage, and involvement in art, music, and theatre.

Before examining the implications of the research findings, it is important to consider the limitations of this study. Most importantly, this research provides evidence about the validity of the academic growth and development items on the Tennessee alumni survey for only one institution, the University of Tennessee, Knoxville. Results should not be used to make judgments about the validity of these items at other institutions, nor should they be used to make judgments about the validity of other measures of academic growth and development.

The conclusions from this research are also limited in terms of the evidence they provide about the validity of the academic growth and development items as used at UTK. Because of the need to examine the invariance of academic growth and development scales over time, coupled with the very large computer memory requirements for calculating polychoric correlations and weighted least squares estimates, only subsets of the UTK alumni responding to the surveys in 1988 and 1990 were used. While these samples were representative of all respondents in 1988 and 1990 in terms of the variables considered in this study, it may be that the samples were atypical in other important respects. Future research, with all alumni respondents, is needed in order to determine if current uses of the academic growth and development items from the Tennessee alumni survey are valid.

Although the relationships between reported college experiences and the academic growth and development scales were statistically significant, only one-fifth to one-third of the variance in these scales was explained by the four college experiences variables included in this study. Based on this finding, it is clear that these experiences are not the only, and perhaps

not the most important, factors influencing perceptions of academic growth and development. Again, future research should be conducted to identify other factors associated with alumni perceptions of academic growth and development, such as coursework, demographic characteristics, and levels of entering ability.

Irrespective of the limitations just discussed, the present research does demonstrate that the validation methodology described in this paper is workable and can yield useful information about whether an assessment instrument accurately represents the outcomes considered important at an institution. Furthermore, the results of this research, coupled with the results of efforts to validate paper-and-pencil tests of achievement in general education, amply demonstrate that the standards of construct validity described in this paper can be used with a variety of different types of assessment measures.

In addition to serving as a criterion for selecting assessment instruments, the validation methodology presented in this paper can be used to improve the design and functioning of an assessment program. For example, evaluating the substantive component of construct validity by examining the content representativeness of an outcomes measure requires that students, faculty, and administrators familiarize themselves with the goals of an assessment program. This reexamination of educational goals can be at least as valuable as evaluating students to determine what they have learned. Likewise, evaluating the structural component of construct validity can also improve the practice of assessment by providing insight into score meaning and by giving assessment practitioners an idea of the confidence that can be place in scores. Finally, studies designed to evaluate the external component of construct validity can help guide improvement efforts and serve as a model for future research on how college affects students. By designing an on-going program to examine the relationships between students' educational experiences and outcomes measures, assessment practitioners can monitor the effects of program changes, while continually evaluating the validity of their assessment instruments.

Despite its limitations, the present research does provide qualified support for the use of the academic growth and development questions from the Tennessee alumni survey to evaluate the quality and effectiveness of the general education program at UTK. These academic growth and development questions do cover the three domains of general education outcomes at UTK, have an underlying structure that is generally consistent with the three outcomes domains, and are differentially related to reports of college experiences.

The most important qualification in using these items to evaluate the UTK general education program is that the academic growth and development questions only cover about half of the outcomes deemed important at UTK. Content areas not covered by the academic growth and development questions include foreign language and computer skills, knowledge in the areas of technology, history, economics, and the social sciences, and judgments and attitudes related to values and political and social dynamics. Moreover, the content areas, of problem solving and aesthetics are covered by only a single item. Given the limitations in content coverage, it seems prudent to limit judgments about attainment of general education goals at UTK to the three general domains of basic skills, knowledge, and judgments and

attitudes. Evaluations of specific outcomes should be made only when the results of the alumni survey can be bolstered by additional evidence from objective test scores, performance appraisals, and surveys of client groups.

Based on the fact that only one-fifth to one-third of the variance in academic growth and development scales can be explained by college experiences, it seems prudent to be very cautious in suggesting program changes based on the results of the alumni survey. It is possible that the observed relationships between college experiences and academic growth and development scales are an artifact of students' background characteristics or other aspects of the general education program. Defining the specific relationships between background characteristics, curriculum, extracurricular activities, and academic growth and development is a subject for future research, and it is a topic that must be addressed before results from the academic growth and development scales are used to suggest extensive revisions in the general education program.

One clear implication of the present research is that the award formula used in performance funding decisions is not appropriate for UTK. Results of the confirmatory factor analysis revealed that alumni perceptions of academic growth and development are not unidimensional and that all items eliciting those perceptions do not carry equal weight. A more accurate representation of general education program effectiveness at UTK can be achieved only if the academic growth and development items on the Tennessee alumni survey are weighted and used to calculate summed scores representing verbal skills, quantitative skills, knowledge outcomes, and judgments and attitudes.

CONCLUSION

This paper makes the very basic argument that it is essential that the instruments used to assess students educational outcomes be valid measures of the constructs they are designed to represent. Validity in this case requires that the content of the assessment measures accurately reflect the goals of an education program, that the empirical structure of assessment data reflect the structure of the outcomes being measured, and that assessment measures be sensitive to the educational experiences of students. This research also indicates that the continuing validation of assessment instruments can be an expensive, time-consuming, and technically-demanding process. However, these are costs that must be borne if assessment is to realize its potential and serve as a catalyst for improving the quality of American higher education. In a nutshell, the quality and effectiveness of assessment and program improvement are a direct result of the quality and effectiveness of the information upon which they are based.

REFERENCES

Alwin, Duane F., and David J. Jackson. "Applications of Simultaneous Factor Analysis to Issues of Factorial Invariance." In *Factor Analysis and Measurement in Sociological Research*, edited by David J. Jackson and Edgar F. Borgatta, 249-279. Beverly Hills: Sage, 1981.

American College Testing Program. *The ACT Evaluation Survey Service: User's Guide*. 4th ed. Iowa City: American College Testing Program, 1989.

Astin, Alexander W. "Involvement: The Cornerstone of Excellence." *Change* 17 (July/August 1985): 35-39.

Astin, Alexander W. *Achieving Educational Excellence: A Critical Assessment of Priorities and Practices in Higher Education*. San Francisco: Jossey-Bass, 1987.

Astin, Alexander W. *Assessment for Excellence: The Philosophy and Practice of Assessment and Evaluation in Higher Education*. New York: Macmillan, 1991.

Banta, Trudy W. "Assessment as an Instrument of State Funding Policy." In *Implementing Outcomes Assessment: Promise and Perils*, edited by Trudy W. Banta, 81-94. New Directions in Institutional Research, no. 59. San Francisco: Jossey-Bass, 1988.

Banta, Trudy W., and Homer S. Fisher. "Utilizing Research on Postsecondary Student Outcomes: Evidence of Effectiveness." Paper presented at the annual meeting of the American Educational Research Association, San Francisco: March 1989. Photocopy of the typescript in my possession.

Banta, Trudy W., and Gary R. Pike. "Methods for Comparing Outcomes Assessment Instruments." *Research in Higher Education* 30 (October 1989): 455-469.

Bentler, Peter M., and Douglas G. Bonett. "Significance Tests and Goodness of Fit in the Analysis of Covariance Structures." *Psychological Bulletin* 88 (November 1980): 588-606.

Bernstein, Ira H., and Gary Teng. "Factoring Items and Factoring Scales are Different: Spurious Evidence for Multidimensionality Due to Item Categorization." *Psychological Bulletin* 105 (May 1989): 467-477.

Byrne, Barbara M., Richard J. Shavelson, and Bengt O. Muthén. "Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance." *Psychological Bulletin* 105 (May 1989): 456-466.

Coordinating Committee on General Education. "Interim Report of the Coordinating Committee on General Education." University of Tennessee, Knoxville, 1981. Photocopy of the typescript in my possession.

James, Lawrence R., Stanley A Mulaik, and Jeanne M. Brett. *Causal Analysis: Assumptions, Models, and Data.* Beverly Hills: Sage, 1982.

Jöreskog, Karl G. "Simultaneous Factor Analysis in Several Populations." *Psychometrika* 36 (December 1971): 409-426.

Jöreskog, Karl G., and Dag Sörbom. *LISREL 7: A Guide to the Program and Applications.* Chicago: SPSS, 1989.

Levy, Robert A. "Development of Performance Funding Criteria by the Tennessee Higher Education Commission: A Chronology and Evaluation." In *Performance Funding in Higher Education: A Critical Analysis of Tennessee's Experience,* edited by Trudy W. Banta, 13-26. Boulder, Colo.: National Center for Higher Education Management Systems, 1986.

Loevinger, Jane. "Objective Tests as Instruments of Psychological Theory." *Psychological Reports* 3 (December 1957): 635-694.

Lohman, David F. "Human Intelligence: An Introduction to Advances in Theory and Research." *Review of Educational Research* 59 (Winter 1989): 333-373.

Long, J. Scott. *Covariance Structure Models: An Introduction to LISREL.* Sage University Paper Series on Quantitative Applications in the Social Sciences, no. 07-034. Beverly Hills, Sage, 1983.

MacCallum, Robert C. "Specification Searches in Covariance Structure Modeling." *Psychological Bulletin* 100 (July 1986): 107-120.

Marsh, Herbert W., John R. Balla, and Roderick P. McDonald. "Goodness-of-Fit Indices in Confirmatory Factor Analysis: The Effect of Sample Size." *Psychological Bulletin* 103 (May 1988): 391-410.

Marsh, Herbert W., and David Grayson. "Public/Catholic Differences in the High School and Beyond Data: A Multigroup Structural Equation Modeling Approach to Testing Mean Differences." *Journal of Educational Statistics* 15 (Fall 1990): 199-235.

McDonald, Roderick P., and Herbert W. marsh. "Choosing a Multivariate Model: Noncentrality and Goodness of Fit." *Psychological Bulletin* 107 (March 1990): 247-255.

McGaw, Barry, and Karl G. Jöreskog. "Factorial Invariance of Ability Measures in Groups Differing in Intelligence and Socioeconomic Status." *British Journal of Mathematical and Statistical Psychology* 24 (November 1971): 154-168.

Messick, Samuel. "Validity." In *Educational Measurement.* 3d ed., edited by Robert L. Linn, 13-103, New York: Macmillan, 1989.

Miller, Gary E. *The Meaning of General Education: The Emergence of a Curriculum Paradigm.* New York: Teachers College Press, 1988.

Millman, Jason. "Designing a College Assessment." In *Performance and Judgment: Essays on Principles and Practice in the Assessment of College Student Learning,* edited by Clifford Adelman, 9-38. Washington D.C.: Government Printing Office, 1988.

Mulaik, Stanley A., Lawrence R. James, Judith Van Alstine, Nathan Bennett, Sheri Lind, and C. Dean Stilwell. "Evaluation of Goodness-of-Fit Indices for Structural Equation Models." *Psychological Bulletin* 105 (May 1989): 430-445.

Pace, C. Robert. *Measuring the Quality of College Student Experiences.* Los Angeles: Center for the Study of Evaluation, 1988.

Pace, C. Robert. *The Undergraduates: A Report on Their Activities and Progress in College in the 1980s.* Los Angeles: Center for the Study of Evaluation, 1990.

Pascarella, Ernest T., and Patrick T. Terenzini. *How College Affects Students: Findings and Insights from Twenty Years of Research.* San Francisco: Jossey-Bass, 1991.

Pike, Gary R. "A Comparison of the College Outcome Measures Program (COMP) and the Collegiate Assessment of Academic Proficiency (CAAP) Exams." Center for Assessment Research and Development, University of Tennessee, Knoxville, 1989a. Photocopy of typescript in my possession.

Pike, Gary R. "Background, College Experiences, and the ACT-COMP Exam: Using Construct Validity to Evaluate Assessment Instruments." *Review of Higher Education* 13 (Fall 1989b): 91-117.

Pike, Gary R. "Dimensions of Alumni Perceptions of Cognitive and Affective Growth During College." Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1990. Photocopy of typescript in my possession.

Pike, Gary R. "The Components of Construct Validity: A Comparison of Two Measures of General Education Outcomes." *Journal of General Education* (in press).

Pike, Gary R., and Trudy W. Banta. "Using Construct Validity to Evaluate Assessment Instruments: A Comparison of the ACT-COMP Exam and the ETS Academic Profile." Paper presented at the annual meeting of the American Educational Research Association, San Francisco, March 1989. Photocopy of the typescript in my possession.

Silvia, E. Suyapa M., and Robert C. MacCallum. "Some Factors Affecting the Success of Specification Searches in Covariance Structure Modeling." *Multivariate Behavioral Research* 23 (July 1988): 297-326.

Sörbom, Dag. "Model Modification." *Psychometrika* 54 (September 1989): 371-384.

South Carolina Higher Education Assessment Network. "ACE/Winthrop National Assessment Survey Results Released." *The Exchange*, no. 5 (Winter 1990): 1, 13.

Tucker, Ledyard R., and Charles Lewis. "A Reliability Coefficient for Maximum Likelihood Factor Analysis." *Psychometrika* 38 (March 1973): 1-10.

Table 1

Questions from the Tennessee Alumni Survey

Academic Growth and Development Questions

[1] Practical skills necessary to obtain employment in your field.
[2] Getting along with people of different races and ethnic groups.
[3] Ability to grow and learn as a person.
[4] Ability to lead or guide others.
[5] Ability to adjust to new job demands.
[6] Self-confidence in expressing your ideas.
[7] Appreciation of different cultures.
[8] Planning and carrying out projects.
[9] Speaking effectively.
[10] Writing effectively
[11] Understanding written information
[12] Understanding graphic information
[13] Learning on your own
[14] Defining and solving problems
[15] Working cooperatively in a group
[16] Ability to understand mathematical concepts
[17] Understanding the interaction between people and the environment
[18] Understanding and appreciating the arts
[19] Understanding and applying scientific principles and methods
[20] Understanding different philosophies and cultures
[21] Ability to use mathematics in everyday life

Table 1 (Continued)

College Experience Questions

[1] Used the library as a quiet place to read or study.
[2] Developed a bibliography or set of references for a term paper.
[3] Used library reference materials.
[4] Went to an art gallery or exhibit on campus.
[5] Attended a concert or other musical performance on campus.
[6] Saw a play on campus.
[7] Availability of your advisor.
[8] Willingness of your advisor to help.
[9] Availability of faculty to help students outside of class.
[10] Availability of faculty to talk informally.
[11] Quality of courses for providing a good general education.
[12] Quality of courses for preparing for employment.
[13] Quality of instruction in the major.
[14] Quality of courses in the major in preparing for graduate or professional school.

Table 2

Number of Academic Growth and Development Questions Representing Each General Education Goal

General Education Goal	Number of Questions
Basic Skills	*7*
Verbal communication	3
Computational skills	3
Foreign language skills	0
Computer skills	0
Problem solving	1
Knowledge	*6*
Aesthetics	1
Science for life	2
Technology	0
Western history	0
Foreign culture	3
Economics	0
Social Sciences	0
Judgments and Attitudes	*8*
Values	0
Political and social dynamics	0
Personal wholeness	6
Life-long learning	2
Experience in learning	0

Table 3

Goodness-of-Fit Results for the Growth and Development Questions

	1988 Sample Model	df	χ^2	GFI	Δdf	$\Delta\chi^2$	TLI	PGFI
[1]	Null model	210	7821.95 b	0.401				
[2]	1-Factor model	189	2376.05 b	0.818	21	5445.90 b	0.681	0.669
[3]	3-Factor model	186	1857.00 b	0.858	3	519.05 b	0.752	0.691
[4]	4-Factor model	183	1555.61 b	0.881	3	301.39 b	0.793	0.698
[5]	19,2 free	182	1166.06 b	0.911	1	389.55 b	0.851	0.718
[6]	19,3 fixed	183	1166.26 b	0.911	1	0.20	0.852	0.722
[7]	14,4 free	182	922.76 b	0.929	1	243.59 b	0.888	0.732
[8]	12,1 fFree	181	726.69 b	0.944	1	196.07 b	0.917	0.740

	1990 Sample Model	df	χ^2	GFI	Δdf	$\Delta\chi^2$	TLI	PGFI
[1]	Null model	210	6249.66 b	0.456				
[2]	1-Factor model	189	1548.25 b	0.865	21	4701.41 b	0.750	0.708
[3]	3-Factor model	186	1225.81 b	0.893	3	322.44 b	0.806	0.719
[4]	4-Factor model	183	1027.09 b	0.911	3	198.72 b	0.840	0.722
[5]	19,2 free	182	717.80 b	0.938	1	309.29 b	0.898	0.739
[6]	19,3 fixed	183	724.79 b	0.937	1	6.99 a	0.897	0.742
[7]	14,4 free	182	606.25 b	0.947	1	118.54 b	0.919	0.746
[8]	12,1 free	181	508.89 b	0.956	1	97.36 b	0.937	0.749

Table 3 (Continued)

	Invariance Model	df	χ^2	1988 GFI	1990 GFI	Δdf	$\Delta\chi^2$	TLI	1988 PGFI	1990 PGFI
[1]	Null model	441	14071.61 b	0.401	0.456					
[2]	Pattern invariant	362	1234.58 b	0.944	0.956	79	12836.03 b	0.922	0.775	0.785
[3]	Factor loadings invariant	385	1337.9[illegible] b	0.940	0.951	23	102.33 b	0.920	0.821	0.830
[4]	Uniquenesses invariant	406	1337.98 b	0.940	0.951	21	0.07	0.926	0.865	0.876
[5]	Factor correlations invariant	412	1361.42 b	0.940	0.950	6	23.44 b	0.926	0.878	0.888

[a]$p < .01$; [b]$p < .001$

Table 4

Weighted Least Squares Estimates for the Growth and Development Questions

	Question	Verbal	Quant.	Know.	Judg.	Unique.	R^2
[1]	Employment in field				0.488	0.762	0.238
[2]	Get along with people			0.724		0.476	0.524
[3]	Grow and learn				0.749	0.438	0.562
[4]	Lead or guide others				0.722	0.479	0.521
[5]	New job demands				0.654	0.572	0.428
[6]	Self-confidence				0.684	0.533	0.467
[7]	Appreciation cultures			0.792		0.372	0.628
[8]	Planning projects				0.738	0.456	0.544
[9]	Speak effectively	0.730				0.467	0.533
[10]	Write effectively	0.715				0.489	0.511
[11]	Understand written	0.922				0.150	0.850
[12]	Understand graphic	0.428	0.418			0.540	0.460
[13]	Learn on own				0.715	0.489	0.511
[14]	Solving problems		0.295		0.612	0.367	0.633
[15]	Work cooperatively				0.651	0.576	0.424
[16]	Math concepts		0.874			0.237	0.763
[17]	People & environment			0.732		0.464	0.536
[18]	Understand arts			0.585		0.658	0.342
[19]	Understand science		0.687			0.527	0.473
[20]	Understand cultures			0.697		0.515	0.485
[21]	Use math everyday		0.931			0.133	0.867

All factor loadings are significant at the $p < .001$ level

Table 5

Goodness-of-Fit Results for the College Experiences Questions

	1988 Sample Model	df	χ^2	GFI	Δdf	$\Delta\chi^2$	TLI	PGFI
[1]	Null model	91	8031.98 b	0.303				
[2]	4-Factor model	71	408.21 b	0.965	20	7623.77 b	0.946	0.753
	1990 Sample							
	Model	df	χ^2	GFI	Δdf	$\Delta\chi^2$	TLI	PGFI
[1]	Null model	91	8214.68 b	0.298				
[2]	4-Factor model	71	399.86 b	0.966	20	7814.82 b	0.948	0.754

	Invariance Model	df	χ^2	1988 GFI	1990 GFI	Δdf	$\Delta\chi^2$	TLI	1988 PGFI	1990 PGFI
[1]	Null model	196	16246.66 b	0.303	0.298					
[2]	Pattern Invariant	142	808.07 b	0.965	0.966	54	15438.59 b	0.943	0.699	0.700
[3]	Factor loadings invariant	156	856.47 b	0.962	0.964	14	48.40 b	0.945	0.766	0.767
[4]	Uniquenesses invariant	170	856.47 b	0.962	0.964	14	0.00	0.951	0.834	0.836
[5]	Facotr correl. invariant	176	928.37 b	0.959	0.961	6	71.90 b	0.948	0.861	0.863

[a]$p < .01$; [b]$p < .001$

Table 6

Weighted Least Squares Estimates for the College Experience Questions

	Question	Library	AMT	Faculty	Courses	Unique.	R^2
[1]	Used library to study	0.558				0.689	0.311
[2]	Developed bibliography	0.756				0.428	0.572
[3]	Used refer. materials	0.979				0.042	0.958
[4]	Went to art gallery		0.673			0.547	0.453
[5]	Attended concert		0.772			0.404	0.596
[6]	Saw a play		0.644			0.586	0.414
[7]	Availability advisor			0.912		0.169	0.831
[8]	Advisor help			0.929		0.136	0.864
[9]	Faculty help			0.883		0.221	0.779
[10]	Faculty talk informally			0.856		0.226	0.734
[11]	Courses gen. educ.				0.777	0.396	0.604
[12]	Courses employment				0.740	0.453	0.547
[13]	Quality instruction				0.783	0.387	0.613
[14]	Courses grad. school				0.781	0.390	0.610

All factor loadings are significant at the $p < .001$ level

Table 7

Weighted Least Squares Estimates of the Effects of College Experiences on Growth and Development

College Experiences	Verbal	Quantitative	Knowledge	Judgments
Library Usage	0.253[b]	-.192[b]	0.068[a]	0.051[a]
Art, Music, Theatre	0.046	-.018	0.322[b]	0.055[a]
Faculty Interaction	0.077[a]	-.065[a]	0.023	0.006
Quality of Courses	0.280[b]	0.417[b]	0.210[b]	0.394[b]
Squared Multiple Correl.	0.254	0.212	0.309	0.343

[a]$p < .01$; [b]$p < .001$