Evaluating the Outcomes of a Service-Learning Based Course in an Engineering Education Program: Preliminary Results of the Assessment of the Engineering Projects in Community Service - EPICS.

Jason C. Immekus, Susan J. Maller, Sara Tracy, & William C. Oakes
Purdue University

Abstract

Design courses embedded in service-learning are rapidly emerging within the curricula of many engineering programs. The learning outcomes service-learning courses seek to promote are well aligned with the Accreditation Board for Engineering and Technology criteria 2000 (EC 2000)[1]. The Engineering Projects in Community Service (EPICS) program integrates engineering design with meeting the needs of the local community through a multi-disciplinary service-learning curricular structure. The EPICS courses can be counted for a wide range of courses in several disciplines, including capstone design in electrical and computer engineering and computer science. The approaches of EPICS to conceptualize and measure specific professional skills for program evaluation purposes are discussed. These include: social-responsibility, awareness of ethical issues, teamwork, and communication competence. Specifically, the theoretical framework used for scale construction, preliminary results, and evidence of the scales' psychometric properties are provided. The aim of this paper is to provide information regarding the use of self-report measures to assess program outcomes.

1. Introduction

Service-learning is the focus of considerable research and is a feature within many engineering programs. Within engineering education, design courses embedded in service-learning provide a way to promote students' development of technical and professional skills for solving applied problems. The ability to create learning environments for engineering students to apply mathematical and scientific principles when solving applied problems is critical for preparing students for careers in engineering[2]. The need for engineering programs to produce students proficient in these skills upon graduation is reflected in ABET EC 2000. Service-learning courses may provide engineering programs one way to promote program, institution, and accreditation outcomes.

Service-learning seeks to promote student learning in the form of experiential education. Jaccoby and Associates[3] define service-learning as, "a form of experiential education in which students engage in activities that address human and community needs together with structured opportunities intentionally designed to promote student learning and development…" (p. 5). Collectively, definitions of service-learning agree that it "joins two complex concepts: community action, the 'service,' and efforts to learn from that action and connect what is learned to existing knowledge, the 'learning'" (p. 2)[4]. Key factors of service-learning include reflection

"Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition

and reciprocity[3]. Reflection deals with augmenting students' understanding of the relationship between their learning and helping others; reciprocity addresses the interaction between students and the community members being served. Service-learning differs from volunteerism because it is explicitly centered on building students' educational experiences[5].

Service-learning is based on promoting the ideas and values of citizenship. Astin and Sax report that students who participate in service-learning during their undergraduate careers make notable social and academic gains[6]. Specifically, community involvement has been associated with enhancing students' appreciation of the people and organizations with whom they worked[7,8], feelings of social responsibility[9,10], and development of factual knowledge related to their field(s)[11]. Students who enroll in service-learning courses are characterized as being intrinsically motivated to assist others and having a desire to seek personal growth and self-actualization[12]. Collectively, empirical evidence suggests that service-learning promotes students' awareness of the broad issues that face society.

Service-learning is believed to promote academic and personal development because it combines students' academic learning with community involvement[3,12]. Within this framework, students are provided the opportunity to reflect on their service experiences to understand how their skills can benefit others[13]. Reflection can be structured into a service-learning program through journals, student meetings, and student meetings with community members[3,13], and can be leveled at specific domains (e.g., link between learning and service). Gathering information on the degree engineering students perceive the connection between their learning and community service as beneficial to their personal and academic development may shed light on how students, particularly females, approach science. For instance, Rosser[14] reports female scientists perceive a sense of connection to their research, value the opportunity to collaborate as opposed to compete, and enjoy working with both scientists and non-scientists in their research. Whether or not service experiences influence students' perceptions of engineering and promotes specific learning outcomes is an empirical question that warrants investigation.

Research is needed to further understand the relationship between service-learning and student outcomes[6,15]. For instance, Vogelgesan, Ikeda, Gilmartin, and Keup[13] indicate that research is needed to examine the effect service-learning may have on persistence. Limitations of previous research results include small sample sizes[7], emphasis on short-term effects[3], and research designs lacking a comparison group[15]. Longitudinal, multi-institutional research has been suggested as a way to investigate the lasting impact of service-learning on students' personal and academic outcomes[6,16]. Others have indicated the need for standardized instruments to evaluate service-learning programs[3]. Consideration of these issues may assist engineering educators to investigate how service-learning courses can be used to promote program, institutional, and accreditation outcomes.

Indeed, self-report instruments are one tool educators can use to measure program outcomes (e.g., teamwork, awareness of ethical issues). There are several features of these instruments make them attractive for program evaluation purposes. First, they can be designed to reflect the objectives of one's engineering program. Second, the process of administering and scoring the scales can be standardized. Third, their psychometric properties (e.g., reliability, validity) can be evaluated and used to guide subsequent scale revisions. Additionally, within single- and multi-institutional research, they provide a cost-effective way to collect student data in the presence of large sample sizes. Furthermore, theoretically grounded instruments with favorable psychometric properties yield scores that can be used to make meaningful inferences regarding the influences of service-learning programs on student outcomes.

Based on these considerations, this paper presents the ongoing research at Purdue University to develop self-report instruments as one component in the overall assessment strategy for the EPICS outcomes. First, the EPICS program and the learning outcomes it seeks to promote are described. Second, ABET EC 2000 Criterion 3 outcomes are mentioned to note their link with program outcomes. Third, the theoretical foundations of the instruments, as well as preliminary results based on our Fall 2004 data collection, are summarized. Additionally, evidence of the psychometric properties of the scales is presented. The paper concludes by discussing how engineering programs can use self-report instruments to evaluate program outcomes.

1a. EPICS program

EPICS is an engineering-centered, service-learning program that integrates multi-disciplinary curricula with meeting the needs of the local community. Within EPICS, teams of eight to twenty undergraduate students are matched with community agencies requesting technical assistance[2, 17]. Under guidance from faculty and/or practicing engineers, students define, design, build, and deploy a usable product for their community partner (e.g., Habitat for Humanity)[2]. Student teams are multi-disciplinary, coming from 20 disciplines including Engineering, Science, and Liberal Arts. Students range from freshmen to senior status and can enroll in EPICS multiple semesters for one to two credit hours. EPICS was formally established in 1995[18] at Purdue University and is now a feature at fifteen undergraduate engineering programs (e.g., Pennsylvania State University, Puerto Rico-Mayaguez, University of Wisconsin, Madison) across the U.S. More information on EPICS can be found at the website *http://epicsnational.ecn.purdue.edu/*.

EPICS involves a long-term commitment on the part of faculty, students, and community agencies. Based on the needs of community agencies, student projects can last less than or longer than the 15-week academic semester. In some instances, complex projects can extend over several years. The time involved to design, build, and deliver a product requires students to consider the implications of their work throughout the design process. This includes considering how the product will be used by the community agency. Students rely on current team members' expertise and progress made by preceding team members to deliver a tangible product to their community partner. The EPICS framework is aligned with Dewey's[19] four criteria for tasks to be "truly educative": (1) stimulate interest, (2) be intrinsically worthwhile, (3) present problems that stimulate curiosity and seeking of new information, and (4) extend over time to promote personal development.

The EPICS program is based on the belief that students learn the steps involved in the design process by working on multidisciplinary teams to solve applied problems. Outcomes associated with participation in EPICS include various lessons in engineering, including the role of the partner, or "customer," in defining an engineering project; the necessity of teamwork; the difficulty of managing and leading large projects; the need for skills and knowledge from many different disciplines; an awareness of ethical issues; and the art of solving technical problems. They also learn many valuable lessons in citizenship, including the role of community service in our society, the significant impact their engineering skills can have on the community, and the connection between assisting others and their own substantial growth as individuals, engineers, and citizens. In a variety of ways, EPICS provides students a learning experience that is closely aligned to the environment they will enter following graduation.

EPICS seeks to promote students' ability to apply their skills to enable community agencies to better serve society. Through their experiences in EPICS, students are required to use their technical and professional skills to work through the engineering design process. Collaboration among students, community agencies, and faculty are critical for teams to successfully provide deliverable products. Therefore, professional skills associated with working on multi-disciplinary teams, communication, and awareness of societal needs, among many, are areas in which students are postulated to benefit. The learning objectives of EPICS are fundamentally connected to the objectives of the engineering program at Purdue University and ABET EC 2000.

1b. Defining and Assessing Criterion 3 Outcomes

Although the outcomes that accredited engineering programs must promote in their graduates are explicitly stated in Criterion 3, they are "intentionally undefined"[20]. Ultimately, how these outcomes are conceptualized, measured, and assessed depend on one's engineering program. As such, outcomes may be defined differently across programs. However, as engineering programs across institutions or academic areas are aware of others' similar efforts, that knowledge can be used to augment their own work, in the form of shared resources and coordinated efforts.

The professional skills of understanding global and societal contexts (3.h), knowledge of contemporary issues (3.j.), understanding professional and ethical responsibility (3.f), teamwork (3.d), and communication competence (3.g) are several learning outcomes that EPICS and Criterion 3 seek to encourage. These professional skills are used to illustrate how EPICS is defining and constructing operational measures of these outcomes.

Social Responsibility. An understanding of the impact of engineering solutions in a global and societal context (3.h.) and a knowledge of contemporary issues (3.j.) are professional outcomes that can, in part, be represented by social responsibility. Within our research, social responsibility is conceptualized as the degree an individual feels connected to the community and seeks to use his/her skills to help others [21,22]. Those with a strong belief in social responsibility value community service and view it as an important component of their lives. It has been conceptualized as a continuous construct and represented in terms of the following five dimensions: exploration, clarification, realization, activation, and internalization[21]. At the earlier stages, students seek introductory and often superficial involvement in community service. At the higher levels, including realization and internalization, individuals are characterized as feeling deep involvement for particular social causes and will modify their career choices to reflect their level of commitment[21]. Based on the nature of EPICS and empirical evidence suggesting that students who self-select into service-learning courses may have previous service experience[23], our scale focuses on the dimensions of realization and internalization.

Ethics. Practicing engineers often encounter ethical issues, yet ABET and engineering corporations consider ethics to be inadequately reflected in engineering education[24]. Ethical awareness refers to the level of attention students give to the social implications of their products[25], both as practicing engineers and throughout the design process. For professional engineers, ethical issues need to be considered at all phases of product design and implementation[25,26]. However, the engineering design process presents ethical issues distinct from those routinely encountered by engineering professionals[26,27]. For example, engineers experience ethical dilemmas related to public health, safety, and welfare, as well as globalization and technology, within many aspects of their jobs[24]. Ethical issues encountered within the design process include making tradeoffs between safety and economic considerations and understanding that ethical concerns arise at multiple points during the design process[27]. Ethical awareness is particularly relevant to the EPICS program, as student teams generate authentic products to be used by community agencies. Reflection, via self-reporting, is one method through which engineering educators can provide students with the tools necessary to be ethically and socially responsible[25].

Teamwork. For the purposes of EPICS, a team is a group of individuals who see themselves and are seen by others as a social entity, which is interdependent because of the tasks performed as members of a group[28]. Accordingly, this interdependence allows the team to experience increased productivity. This definition of teamwork reflects the way students work together in EPICS. Specifically, student teams must be able to manage rotating team members, maintain a budget, adhere to assigned roles, and establish and attain short- and long-terms goals. Aspects of teamwork considered critical within EPICS include: composition, interdependency, norms. These dimensions were selected based on the design of EPICS, theory, and empirical evidence suggesting that these sub-domains are related to team effectiveness[20,29,31]. For instance, composition addresses the productivity of teams with mixed traits, such as background diversity and membership stability; interdependency is leveled at students' ability to coordinate and collaborate in their efforts; and roles assesses the processes and roles (e.g., community liaison, team leader) agreed upon by the group.

Communication Competence. For our purposes, communication competence is defined as, "the ability to choose among available communicative behaviors in order that he may successfully accomplish his own interpersonal goals during an encounter while maintaining the face and line of his (or her) fellow interactants within the constraints of the situation"[32]. Thus, communication competence includes any or all of the mechanisms (i.e., written, oral) students use to transmit their oral and written messages to technical and non-technical audiences. Our interpersonal communication scale is comprised of the following subscales: Interaction Management, Environmental Control, Conflict Resolution[33]. The Interaction Management and Environmental Control subscales assess students' abilities to negotiate topics for discussion and achieve predetermined conversational goals; and Conflict Resolution deals with students' abilities to overcome interpersonal conflicts and maintain the team's effectiveness. Another subset of items measure oral and written competencies.

Oral presentations conducted at the end of each semester provide student teams the opportunity to describe the technical assistance they provided to their community partners. Specifically, students are expected to discuss how they approached and navigated the design process to produce a deliverable product. Audience members, including faculty, teaching assistants, and community members attending students' presentations complete a rating scale to evaluate the team's ability to communicate. The presentation performance rating scale is based on Dannels'[34] ethnographic study of a mechanical engineering senior design course. Specifically, Dannels determined that several competencies are valued by engineering professionals in oral communication for both technical and non-technical audiences[34]. The subscales of the EPICS presentation rating scale consist of Visual Sophistication, Results-Oriented Approach, Translation of Ideas, Content Knowledge, and Delivery. For example, the Results-Oriented Approach subscale focuses on whether students made the presentation's thesis and implications clear to the audience; Translation of Ideas addresses whether presenters provided explanations of design decisions, background information, and technical jargon that were appropriate for the audience.

Self-report measures and rating scales completed by faculty, teaching assistants, and students are being used to operationalize these outcomes. Scales to assess these outcomes have been developed, implemented, and refined according to theory, feedback from experts, and pilot administrations of the instruments. Next, we summarize preliminary data and evidence of the scales' psychometric properties based on our Fall 2004 data collection.

2. Methods

2a. Participants

Participants were drawn from the 265 students registered for EPICS for the Fall 2004 semester, at Purdue University. Data collected on 178 students (67.2%) at the onset of the semester indicated that they had completed 3 years in college (senior status) (range=1 to 5 years) and had participated in EPICS for 2 semesters (range=1 to 5 semesters). Student majors were as follows: 74% Engineering, 11% Computer Science, 7% Liberal Arts, 2% Technology, and 6% other. Of the 265 registered EPICS students, 203 participated in the December 2004 cohort, a response rate of 76.6%.

2b. Instrumentation

The development of self-report instruments for program evaluation purposes is designed to meet our long-term goal of longitudinal, multi-institutional research. A common set of self-report instruments would permit within- and between-student comparisons based on data collected from the same instruments. It is recognized that ABET EC 2000 states that self-report instruments cannot be used as sole indicators of students' attainment of Criterion 3 outcomes. However, self-report instruments that possess favorable psychometric qualities allow for research results based on representative test scores. For each scale, students recorded their responses on a 5-point Likert scale (e.g., 1=strongly disagree to 5=strongly agree).

Social Responsibility. The social responsibility scale consists of 13 items and includes two subscales: Realization and Internalization. The Realization subscale contains six items; Internalization consists of seven items. The total scale score represents a student's level of social responsibility, or the degree they feel community service is a central aspect of their life.

Ethics. The ethical awareness scale consists of 18 items and includes two subscales: Professional Impact and Design Process. The Professional Impact subscale includes seven items and the Design Process subscale contains six items. The total scale score represents a student's level of awareness of ethical issues in engineering.

Teamwork. The teamwork scale is a 15-item measure comprised of the following three subscales: Composition, Interdependency, and Roles. The number of items in each subscale ranges from four (Composition) to six (Interdependency) items.

Communication Competence. The communication competence scale is a 26-item measure and includes the following subscales: Interaction Management, Environmental Control, Conflict Resolution, Oral, and Written. The subscales consist of four (Written) to six (Interaction Management) items. Three subscales (Interaction Management, Environmental Control, and Conflict Resolution) correspond to interpersonal communication. The Oral and Written subscales correspond to expressive communication. Scores across the three interpersonal subscales will be used to represent a student's level of interpersonal communication, whereas scores across the two expressive subscales will indicate one's expressive communication.

The presentation performance rating scale is a 19-item scale rated by audience members (e.g., faculty, students, community partners) attending team presentations. The rating scale consists of the following subscales: Visual Sophistication, Results-Oriented Approach, Translation of Ideas, Content Knowledge, and Delivery. The scale was piloted during the December 2004 end-of-semester presentations. Results of the pilot administration of the presentation performance rating scale will not be addressed in this paper.

2c. Scale Descriptives and Psychometric Properties

The *Standards for Educational and Psychological Measurement*[35] state that evidence of an instrument's psychometric properties (e.g., reliability, validity) must be provided to substantiate the use of obtained scores. This section presents preliminary evidence of reliability and validity of the scales based on our Fall 2004 data collection.

Table 1 shows the internal reliability estimates (Cronbach's coefficient alpha) and descriptive statistics of the scales. Scores are reported as sum scores. For instance, the highest score a student could report on the overall Social Responsibility scale is 65. The means and medians for each scale were roughly equal, indicating student scores were approximately symmetrical.

Item analysis is being used to identify items that contribute to the functioning of each scale. Specifically, items to retain in the scales are based on descriptive statistics, item discriminations, and internal reliability estimates. For instance, within the Team scale, the Composition subscale originally consisted of five items. However, one item within this measure (e.g., "I have had to manage changing team members to avoid disruption of the team's progress") was deleted due to a low correlation (.24) with the total scale score. Deletion of this item resulted in an increased internal consistency reliability estimate. As such, our research is focused on identifying items that are contributing to the overall functioning of each scale.

Reliability. Reliability refers to the degree an instrument produces consistent scores[36,22]. Specifically, reliability indicates the degree test scores are not affected by random measurement error[37]. Internal consistency reliability indicates the degree an instrument's items function together to yield a test score[22]. High internal consistency reliability estimates (e.g., above .90) are desired[38]. As shown in Table 1, overall scale reliabilities are acceptable. With the exception of the Composition subscale and Written Expression, subscale reliabilities were acceptable.

Table 1. Number of Items, Reliabilities, and Descriptive Statistics of EPICS Scales.

Scale	Subscale	No. Items	Alpha	*M* (*SD*)	*Mdn*	Range
Social Responsibility	Realization	6	.88	21.73 (4.48)	23	7-30
	Internalization	7	.88	25.37 (5.33)	26	9-35
	Overall	13	.92	47.09 (8.96)	48	16-65
Ethical Awareness	Professional Awareness	7	.89	23.20 (4.63)	24	6-30
	Design Process	6	.92	27.21 (5.67)	28	7-35
	Overall	13	.95	47.09 (8.96)	52	16-65
Teamwork	Composition	4	.75	14.81 (3.03)	15	4-20
	Interdependency	6	.91	22.66 (4.44)	24	6-30
	Roles	5	.91	22.98 (4.81)	24	7-30
	Overall	15	.94	60.34 (11.40)	63	21-80
Interpersonal Communication	Interaction Management	6	.92	18.54 (3.78)	19	6-25
	Environmental Control	5	.84	17.69 (3.59)	18	7-25
	Conflict Resolution	6	.90	22.63 (4.55)	24	7-30
Expressive	Oral	5	.88	18.49 (4.16)	20	6-25
	Written	4	.74	13.46 (3.14)	14	4-20
	Overall	26	.95	95.00 (17.46)	98	36-130

Note. *M*=Mean, *SD*=Standard Deviation, *Mdn*=Median.

Validity. Validity is the degree an instrument measures what it was designed to measure[39]. Test scores with evidence of validity provide engineering educators with meaningful information for evaluating program outcomes. Although there are distinct types of validity (e.g., content, criterion, construct), they all ultimately serve as construct validity evidence[40].

Content validity. Content validity refers to the correspondence between test content and its associated domain. Typically, content validity is ensured by (a) thoroughly reviewing the content domain literature, and (b) review of items by content experts. For the EPICS scales, a team of researchers (graduate students and faculty) thoroughly reviewed the relevant literature. Further, a panel of students and faculty familiar with the measured outcomes assisted in item and test development. Faculty within Purdue University's Communication Department served as consultants on the communication competence measures; engineering faculty from multiple institutions provided feedback on the scales at the 2004 EPICS National Conference.

Construct validity. Construct validity addresses whether a scale measures what it was designed to measure. Factor analytic procedures are the most commonly used methods to assess an instrument's factor structure. Factor analysis procedures fall under two categories: exploratory (EFA) and confirmatory (CFA) factor analysis. EFA is data driven; CFA is theory driven[38]. EFA is useful when there is no *a priori* information available to suggest the relationship between observed variables (e.g., items) and latent traits (e.g., communication competence). Because our scales are based on theoretical models, CFA is being used to document evidence of construct validity. Within CFA, various "fit statistics" are used to evaluate the degree to which a specified measurement model represents the actual data.

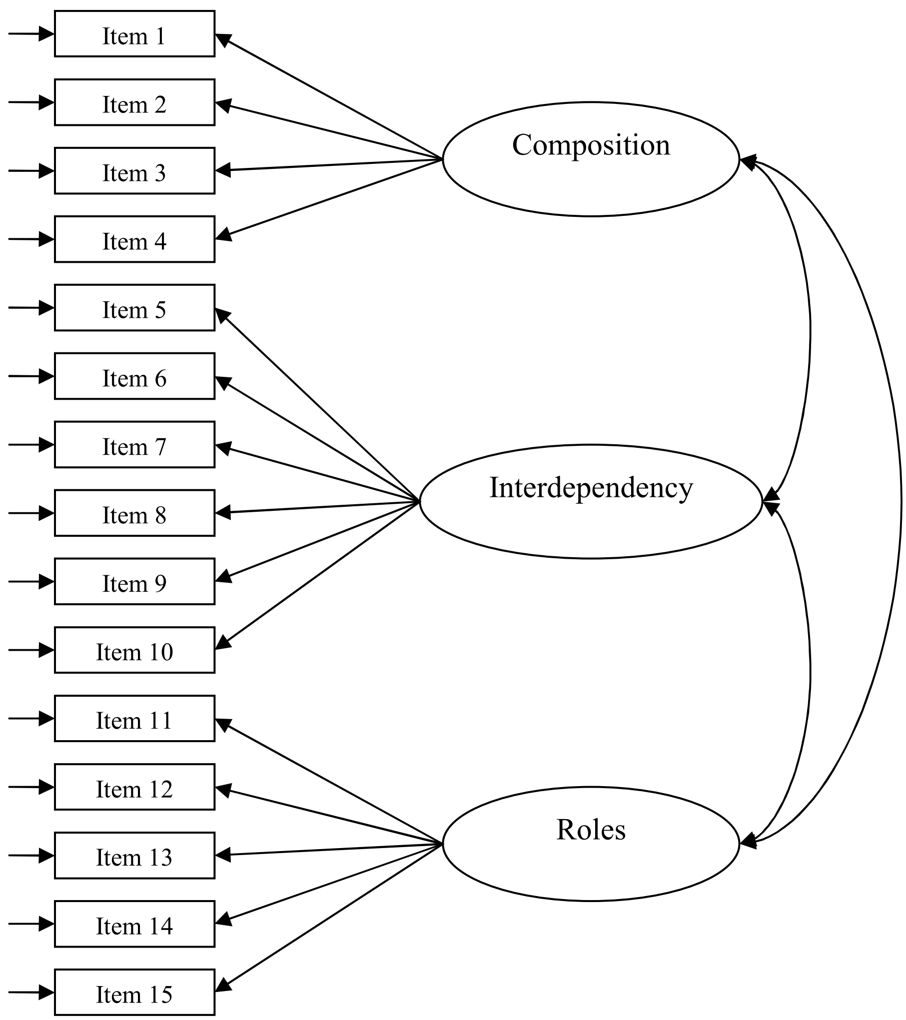
Current evaluation and refinement of our scales are in the preliminary stages. For didactic purposes, two examples are provided to demonstrate how CFA is being used to investigate each scale's factor structure. Specifically, competing theoretical CFA models of the teamwork and communication scales are presented. First, first-order CFA models of the scales were examined. Within these models, items are explained by first-order factors, shown in Figure 1. Next, second-order CFA models were investigated (see Figures 2 & 3). That is, the relationships among the items are posited to be explained by first-order factors, which are accounted for by higher-order factors. Second-order models are preferred over first-order models because they can be used to explain first-order factors in terms of second-order factors[41,42]. As a result, they represent a more parsimonious description of the data[43].

For each of the analyses presented, maximum likelihood based on a covariance matrix was used for parameter estimation. Various fit statistics were used to examine the fit of the models. These included: Satorra-Bentler's (*S-B*) chi-square statistic (χ^2)[44], ratio of chi-square to degrees of freedom (χ^2/df), Root Mean Error of Approximations (*RMSEA*), Standardized Root Mean Square Residual (*SRMR*), and Comparative Fit Index (*CFI*). The chi-square to degrees of freedom ratio suggests how much larger chi-square is than would be expected, with values less than 3.00 indicating good fit[45]. The *RMSEA* provides a measure of the discrepancy between the actual and estimated variance-covariance matrix per degree of freedom[46,47], with values equal to or less than .05 indicating good model fit and values less then .08 indicating reasonable fit[46]. The SRMR indicates the average deviation between the actual and predicted correlation, with values below .09 indicating acceptable fit[47]. The *CFI* provides a measure of the discrepancy between a restricted and null model in relation to the fit of the null model[48.49]. Its values range between 0-1, with values above .95 suggesting adequate fit[47]. LISREL 8.53[50] was used to examine the CFA models.

Teamwork Scale

The first analysis investigated a three factor model of the teamwork scale. Specifically, it was of interest to test whether the items comprising the scale could be described by three first-order factors (e.g., Composition, Interdependency, & Roles). Figure 1 shows the CFA model investigated. Within the figure, the observed variables (items) are represented as squares; the unobserved, latent traits are shown as circles. The arrows connecting the latent traits and items are called factor loadings, or pattern coefficients, which indicate the relationship between the observed and latent variables. Factor loadings are central to investigating construct validity[51]. The arrows corresponding to the items represent error variances, or "score unreliability" [41]. The arrows between the latent traits indicate that the constructs are correlated. It is desirable to have high factor loading parameter values and low error variance values. Within Figure 1, for instance, the latent construct Composition is hypothesized to explain the relationships among the first four items.

Figure 1. Hypothesized First-Order CFA Model of Teamwork Scale

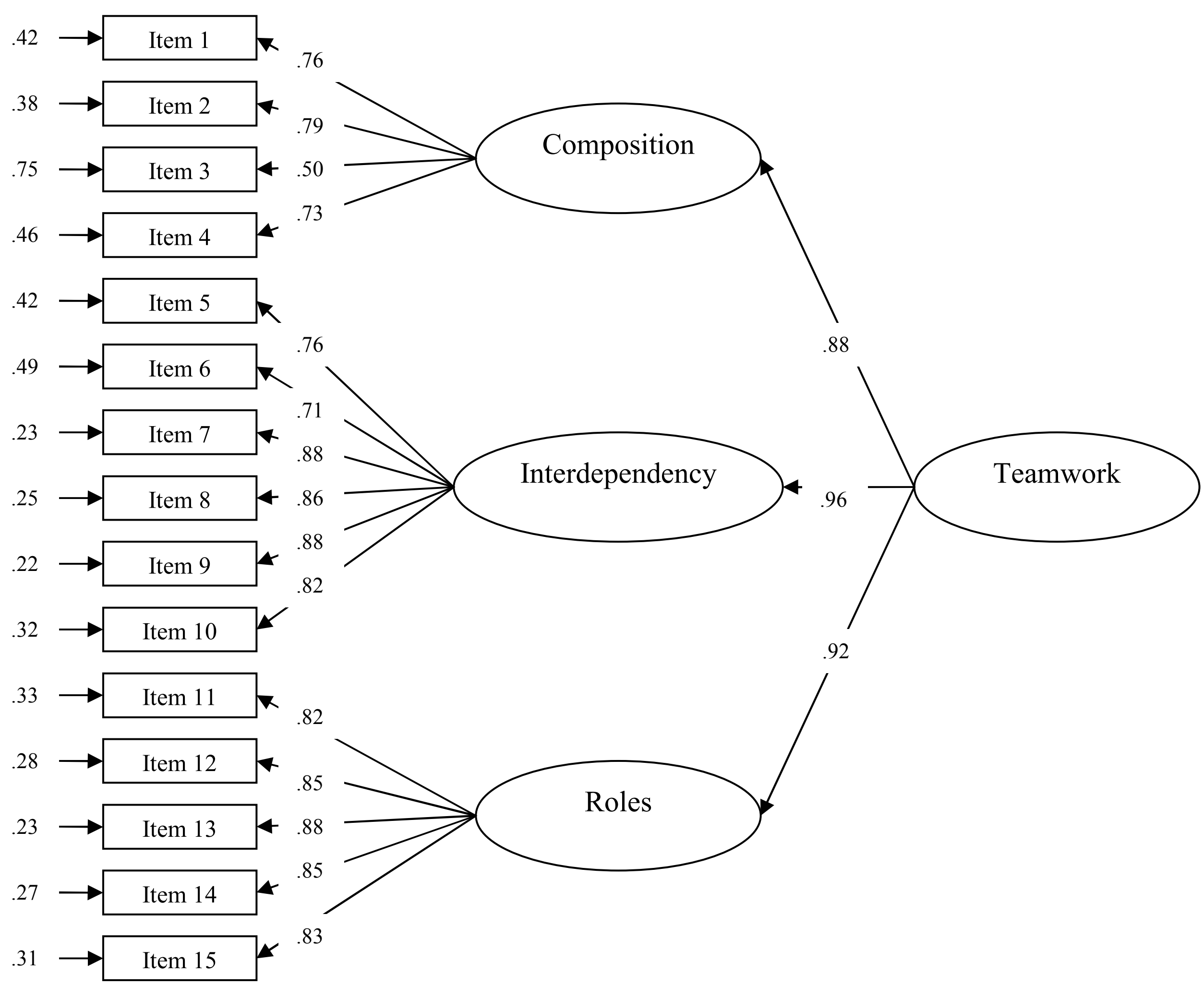


In addition, a second-order CFA model of the teamwork scale was investigated (final model shown in Figure 2). To analyze the fit of this model, a two-step procedure was used. First, an independence model (constrained model) was fit to the data. This is a baseline model that specifies a second-order factor with the paths connecting the first-order factors (e.g., Composition, Interdependency, and Roles) and the second-order factor (Teamwork) fixed to zero. Next, a second model (free model) was estimated with the paths between the first- and second-order factors freely estimated. A chi-square difference test[52,53] can be used to evaluate whether the free model provides a better fit to the data than the independence model.

The hypothesized three-factor model fit the data (*S-B* $\chi^2[87]=80.32$, $p=.68$, $\chi^2/df=.92$, *RMSEA*=.00, *CFI*=.98, *SRMR*=.08). Next, it was of interest to examine the plausibility of a higher-order CFA model.

As expected, the independence model did not fit the data (*S-B* $\chi^2[90]=229.45$, $p<.001$, $\chi^2/df=2.55$, *RMSEA*=.08, *CFI*=.92, *SRMR*=.43). For the purposes of this study, it was designed to serve as a disconfirmable, baseline model[41]. The second-order model with the paths between the first- and second-order factors freely estimated fit the data (*S-B* χ^2[87]=80.32, *p*=.68, χ^2/df =.92, *RMSEA*=.00, *CFI*=.98, *SRMR*=.08). As expected, the fit of this model was the same as the three factor model. The chi-square of the independence model was statistically significant coupled with poor fit statistics (e.g., *SRMR*), indicating that the model was misspecified. Consequently, a chi-square difference test[52,53] between the independence model and free model was not conducted. As a result, the free model was accepted as the final model. Figure 2 shows the final model with parameter estimates.

Figure 2. Second-Order CFA Model of Teamwork Scale



Communication Scale

The second CFA investigated the factor structure of the communication scale. The first model tested a five factor CFA model, with the following latent traits accounting for the relationships among the 26-items: Interaction Management, Environmental Control, Conflict Resolution, Oral and Written. Subsequently, a second-order model was tested to determine whether higher-order Interpersonal and Expressive factors accounted for the relationships among the five first-order factors. Within this model, the three interpersonal factors of Interaction Management, Environmental Control, and Conflict Resolution were hypothesized to represent the higher-order factor Interpersonal; the first-order factors of Oral and Written were postulated to represent the second-order factor Expressive.

The hypothesized five-factor model fit the data (*S-B* χ^2[289]=367.59, *p*=.001, χ^2/df =.1.27, *RMSEA*=.04, *CFI*=.94, *SRMR*=.06). Next, it was of interest to examine whether the inclusion of two second-order factors, Interpersonal and Expressive, fit the data. The independence model specified the presence of two second-order factors with paths between the first- and second-order factors fixed to zero. This baseline model was used to examine whether establishing the relationships between the first- and second-order factors results in improved model fit.

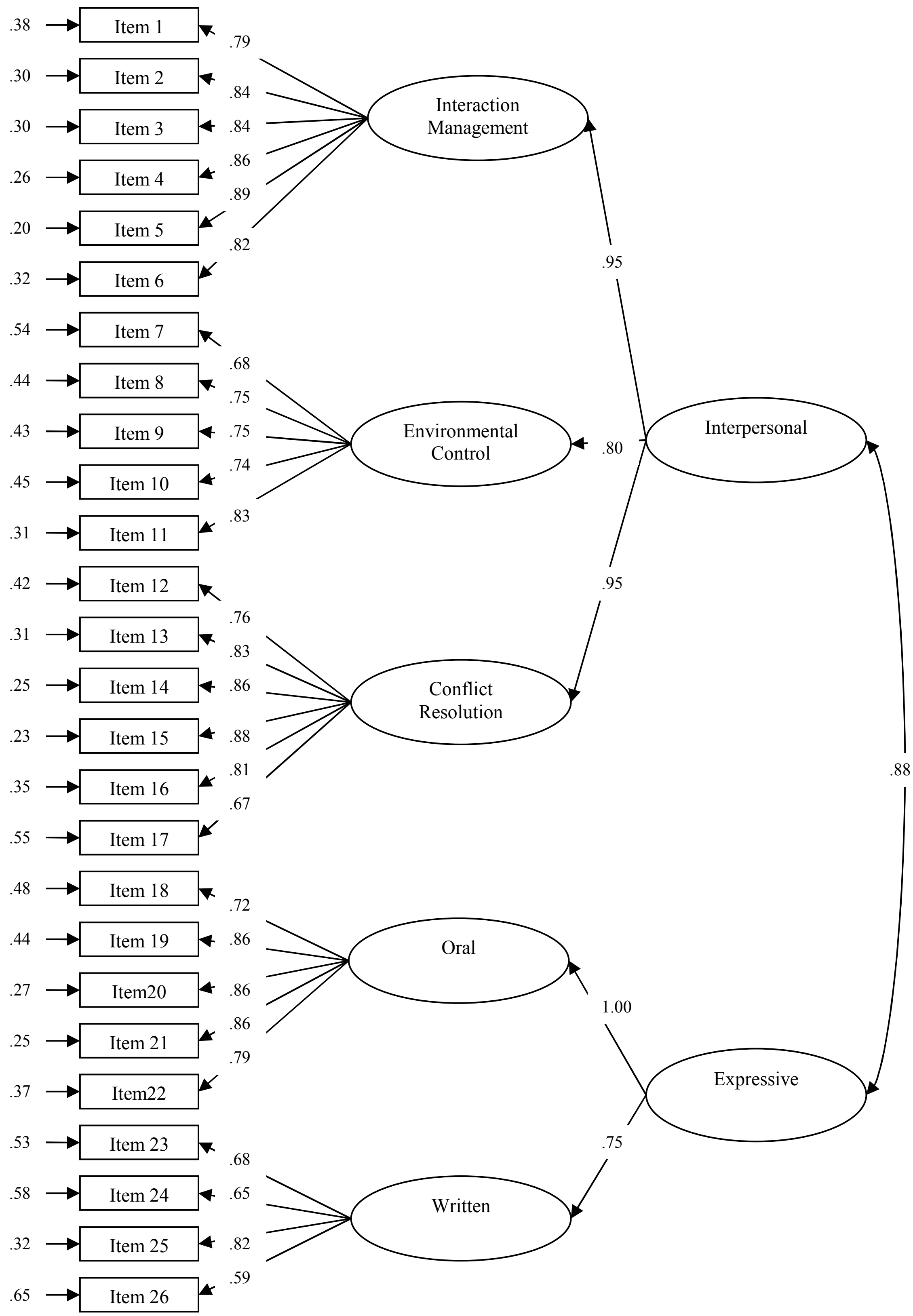
As expected, the independence model did not fit the data (*S-B* χ^2[300]=933.99, *p*<.001, χ^2/df =3.11, *RMSEA*=.10, *CFI*=.89, *SRMR*=.43). The second-order model with the paths between the first- and second-order factors freely estimated fit the data (*S-B* χ^2[294]=391.71, *p*<.001, χ^2/df =1.32, *RMSEA*=.04, *CFI*=.94, *SRMR*=.07). The chi-square of the independence model was statistically significant and other fit statistics indicated poor model fit (e.g., *SRMR*), suggesting that the model was misspecified. Consequently, a chi-square difference test[52,53] between the independence model and free model was not conducted. As a result, the free model was accepted as the final model. As such, Figure 3 depicts the final second-order CFA model of the Communication scale with parameter estimates.

3. Conclusion

The conceptualization and measurement of student outcomes is critical for various program evaluation purposes. In light of Criterion 3 outcomes, engineering programs are faced with many challenges when defining and assessing program outcomes[20]. First, the outcomes must be defined in terms of one's engineering program. Second, assessment tools must be identified and developed in consideration of program outcomes. In the presence of large sample sizes in single- and multi-institutional research, self-report instruments provide engineering educators a flexible method to evaluate the relationships between program and student outcomes. As such, the intent of this paper was to describe the process in which the self-report instruments being designed to assess EPICS and Criterion 3 outcomes are being developed and evaluated at Purdue University.

Within this study, self-report instruments are being designed to evaluate whether an engineering design course based on service-learning is effective in promoting program and Criterion 3 outcomes. Self-report instruments go beyond simply asking students if they learned a topic or achieved an outcome and ask them to reflect on their knowledge and behaviors. They can be used to track students' academic and personal development over the course of their undergraduate careers. For instance, students' standing on specific skills prior to beginning the engineering program can be assessed through the administration of pre-tests. To examine whether growth has occurred, the scores of students who enrolled in one course (e.g., treatment) can be compared to those in another course (e.g., matched sample) to determine whether growth can be attributed to the dynamics of the engineering course of interest. As such, self-report instruments provide engineering programs a method to evaluate students' growth across a range of skills relevant to pursuing a career in engineering.

Figure 3. Second-Order CFA Model of Communication Competence Scale



There are several steps engineering educators should consider throughout scale construction. First, formal operational definitions of outcomes should be based on theory, empirical evidence, Criterion 3, and the goals of one's engineering program. Engineering educators have the option to select pre-existing instruments, adapt pre-existing instruments, or construct new instruments to meet their assessment needs[45]. In the event that a new scale will need to be developed, it should be designed in accordance with theoretical considerations, such as identifying the subscales that will be used to represent the learning outcome(s). Further, the instrument should be designed in consideration of the population it will be administered and how obtained scores will be used.

Within this paper, the steps being used to investigate the psychometric properties of our scales were presented. Item analysis is being used to delete, modify, and replace poor items (e.g., low discrimination). Internal reliability estimates (e.g., Cronbach's coefficient alpha) indicate that our scales are providing consistent scores. CFA is being used to investigate evidence of construct validity of the scales. As demonstrated in this paper, the teamwork and communication scales can be conceptualized as higher-order CFA models. Factor loadings indicated that the items were related to the constructs they were designed to measure. For the teamwork scale, error variances were relatively low, suggesting that the latent traits account for the majority of variance in the observed variables. Several of the error variances for the communication scale items were high, suggesting that the factors may not fully explain the variability in the items. This suggests that these items may need to be revised to more accurately measure the construct they were designed to measure. Nevertheless, results of the CFA analyses provide promising results regarding the functioning of the scales.

Bibliography

1. ABET (1999). *Criteria for Accrediting Engineering Programs*. The Engineering Accreditation Commission of the Accreditation Board for Engineering and Technology. http://www.abet.org/eac/eac.htm.
2. Oakes, William C., & Rud Jr., A. G. (2001). The EPICS model in engineering education: Perspective on problem-solving abilities needed for success beyond school. In H. Doerr & R. Lesh (eds.). *Beyond Constructivism: A models & modeling perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.
3. B. Jaccoby & Associates (1997). *Service-learning in higher education*. San Francisco, CA: Jossey-Boss.
4. Stanton, T. K., Giles, D. E., Jr., & Cruz, N. (1999). *Service-learning: A movement's pioneers reflect on its origins, practice, and future*. San Francisco: Jossey-Bass.
5. Nathan, J., & J. Kielsmeir (1991). The sleeping giant of school reform. *Phi Delta Kappan, 72*, 739-742.
6. Astin, A. W., & Sax, L. J. (1998). How undergraduates are affected by service participation. Journal of College *Student Development, 39*, 251-63.
7. Giles, D. E., Jr., & Eyler, J. (1994). The theoretical roots of service-learning in John Dewey: Toward a theory of service-learning. *Michigan Journal of Community Service Learning, 1*, 77-85.
8. Conrad, D., & Hedin, D. (1982). The impact of experiential education on adolescent development. *Child and Youth Services., 4*, 57-76.
9. Hamilton, S. F., & Fenzel, R. S. (1988). The impact of volunteer experience on adolescent social development: Evidence of program effects. *Journal of Adolescent Research, 3*, 65-80.
10. Newman, F. M., & Rutter, R. A. (1983). *The effects of high school community service programs on students' social development. Final report*. University of Wisconsin, Wisconsin Center for Educational Research: Madison.
11. Hamilton, S. F., & Zeldin, L. M. (1987). Learning civics in the community. *Curriculum Inquiry, 17*, 407-420.
12. Waterman, A. S. (1997). *Service-learning: Applications from the research*. Mahwah, NJ: Lawrence Erlbaum.

13. Vogelgesand, L. J., Ikeda, E. K., Gilmartin, S. K., & Keup, J. R. (2002). Service-learning and the first-year experience: Outcomes related to learning and persistence. In E. Zlotkowski (Ed.), *Service-learning and the first-year experience: Preparing students for personal success and civic responsibility* (Monograph No. 34) (pp. 15-26). Columbia, SC: University of South Carolina, National Resource Center for the First-Year Experience and Students in Transition.
14. Rosser, S. V. (1990). *Female-friendly science*. New York: Pergamon Press.
15. Blyth, D. A., Saito, R., & Berkas, T. (1997). A quantitative study of the impact of service-learning programs. In A. S. Waterman (Ed.), *Service-learning: Applications from the research*. Mahwah, NJ: Lawrence Erlbaum.
16. American Council on Education. (1993). *Outside the classroom: Students as employees, volunteers, and interns*. (Research Briefs, Issue 4, No. 1) Washington, DC: Author.
17. Oakes, W. C., Coyle, E. J., Fortek, R., Gray, J., Jamieson, L. H., Watia, J., & Wukasch, R. (2000). EPICS: Experiencing engineering design through community service projects. In *Proceeding of ASEE Annual Conference*. American Society for Eng. Education.
18. Oakes, W. C., Jamieson, L. H., & Coyle, E. J. (2001). EPICS: Meeting EC 2000 through service learning, In *Proceeding of ASEE Annual Conference*. American Society for Eng. Education.
19. Dewey, J. (1933). *How we think*. Boston: Heath.
20. Besterfeld-Sacre, M., Shuman, L. J., Wolfe, H., Atman, C. J., McGourty, J., Miller, R. L., Olds, B. M., & Rogers, G. M. (2000). Defining the outcomes: A framework for EC-2000. *IEEE Transactions on Education, 43*, 100-109.
21. Delve, C.L., Mintz, S. D., & Stewart, G. M. (1990). Promoting values development through community service: A design. *New Directions for Student Services, 4*, 7-29.
22. Olney, C., & Grande, S. (1995). Validation of a scale to measure development of social responsibility. *Michigan Journal of Community Service Learning*, 43-53.
23. Herkert, J. R. (2003). ABET's engineering criteria 2000 and engineering ethics: Where do we go from here? Retrieved March 2, 2004 from Case Western Reserve University, The Online Ethics Center for Engineering and Science Web site: http://onlineethics.org/essays/ education/herkert2.html
24. McLean, G. F. (1993). Integrating ethics and design. *IEEE Technology and Society Magazine, 12, 3,* 19-30.
25. Lynch, W. T. (1997). Teaching engineering ethics in the United States. *IEEE Technology and Society Magazine, 16, 4,* 27-36.
26. van Gorp, A., & van de Poel, I. (2001). Ethical considerations in engineering design process. *IEEE Technology and Society Magazine, 20, 3,* 15-22.
27. Guzzo, R. A. (1986). Group Decision Making and Group Effectiveness. In Goodman, P. S. (Ed.). Designing Effective Work Groups, 34-71. San Francisco, CA: Jossey-Bass.
28. Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology, 46*, 823-850.
29. Hyatt, D. E., & T. M. Ruddy (1997). An examination of the relationship between work group characteristics and performance: One more into the breech. *Personnel Psychology, 50*, 553-585.
30. Goodman, P. S., E. C. Ravlin, & L. Argote (1986). Current thinking about groups: Setting the stage for new ideas. In P. S. Goodman and Associates (Eds.), *Designing effective work groups*. San Francisco, CA: Jossey-Bass.
31. Wiemann, P. (1977). Explication and test of a model of communicative competence. *Human Communication Research, 3*, 195-213.
32. Rubin, R. B., & Martin, M. M. (1994). Development of a measure of interpersonal communication competence. *Communication Research Reports, 11,* 33-44.
33. Dannels, D. P. (2002). Communication across the curriculum and the disciplines: Speaking in engineering. *Communication Education, 51*(3), 254-268.
34. American Educational Researchers Association, American Psychological Association, and the National Council on the Measurement in Education (1999). *Standards for educational and psychological measurement*.
35. Norton, R. W. (1978). Foundation of a communicator style construct. *Human Communication Research, 4*, 99-112.
36. Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton.
37. Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement*. Monterey, CA: Brooks/Cole Publishing.

38. Crocker, L. & Algina, J. (1984). *Introduction to classical and modern test theory.* Fort Worth, TX: Harcourt Brace.
39. Thompson, B. (2003).Understanding reliability and coefficient alpha, really. In Bruce Thompson (Ed.), *Score Reliability: Contemporary thinking on reliability issues*. Thousand Oaks, CA: Sage.
40. Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*, 5-11.
41. Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington, DC: American Psychological Association.
42. Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and Simplis: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.
43. Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research, 23*, 51-67.
44. Satorra, A., & P. M. Bentler (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis* (pp. 399-419). Thousand Oaks, CA: Sage.
45. Bollen, Kenneth (1989). *Structural equations with latent variables*. New York, NY: Wiley.
46. Brown, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.): *Testing Structural Equation Models*. Sage Publications.
47. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.
48. Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling, 3*, 369-379.
49. Bentler, P. M. (1990). Comparative fix indexes in structural models. *Psychological Bulletin, 107*, 238-246.
50. Jöreskog, K., & Sörbom, D. (2002). *LISREL 8.53* [Computer Program]. Lincolnwood, IL: Scientific Software International.
51. Keith, T. Z. (1997). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 373-403). New York: Guilford.
52. Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In Heijmans, R.D.H., Pollock, D.S.G. & Satorra, A. (eds.), Innovations in multivariate statistical analysis. A Festschrift for Heinz Neudecker (pp.233-247). London: Kluwer Academic Publishers.
53. Satorra, Albert and Bentler, Peter M. (1999). "A Scaled Difference Chi-Square Test Statistic for Moment Structure Analysis." Universitat Pompeu Fabra, Department of Economics Working Paper No. 412. http://ssrn.com/abstract=199064.

Biographical Information

Jason C. Immekus, M.S., is an advanced doctoral student in the Educational Psychology program at Purdue University. His areas of specialization are applied measurement and research design. His research interests include test validity, structural equation modeling, and test development.

Susan J. Maller is an associate professor in the Educational Studies at Purdue University. Her areas of specialization include psychometrics and quantitative research methods with research interests in item response theory, structural equation modeling when used to examine test validity, and item and test bias. She is a Spencer Dissertation Fellow, Assoc. Editor of *Psychology in the Schools*, and on the Editorial boards of several national refereed journals.

Sara Tracy is a master's student in the Educational Psychology program at Purdue University. Her research interests are in parents' influence on young children's learning and development.

WILLIAM C. OAKES is an Associate Professor in the Department of Engineering Education at Purdue University and the Co-Director of the EPICS Program. He is a co-recipient of the 2005 NAE Bernard M. Gordon Prize for Innovation in Engineering and Technology Education and the 2004 NSPE Engineering Education Excellence Award. He is a former-chair of the ASEE IL/IN Section, and board member of FPD and ERM Divisions.