Research Report

Development and Validity Evidence Supporting a Teamwork and Collaboration Assessment for High School Students

Xiaohua Zhuang

Carolyn MacCann

Lijuan Wang

Lydia Liu

Richard D. Roberts

October 2008

ETS RR-08-50

Development and Validity Evidence Supporting a Teamwork and Collaboration Assessment for High School Students

Xiaohua Zhuang
Rutgers University, New Brunswick, NJ

Carolyn MacCann
ETS, Princeton, NJ

Lijuan Wang
University of Notre Dame, IN

Lydia Liu and Richard D. Roberts
ETS, Princeton, NJ

October 2008

As part of its nonprofit mission, ETS conducts and disseminates the results of research to advance quality and equity in education and assessment for the benefit of ETS' constituents and the field.

ETS Research Reports provide preliminary and limited dissemination of ETS research prior to publication. To obtain a PDF or a print copy of a report, please visit:

http://www.ets.org/research/contact.html

Abstract

Various policy papers and research studies assert that teamwork is one of the most important skills for students to learn if they are to become meaningful contributors to the 21st century workforce. However, outside of organizational psychology and adult populations, few reliable assessments of this construct exist, with suitable validity evidence scant or nonexistent. To redress this imbalance, teamwork assessments for high school students were developed using multiple methods: self-report ratings, situational judgment testing, and teacher reports. Exploratory factor, confirmatory factor, and latent class analyses were used to determine the structure of the scales. Measures showed reasonable reliability and promising validity evidence, relating to each other and to academic achievement, while remaining relatively independent from personality. The advantages and disadvantages of each methodology and the potential applications for identification and intervention, selection, and evaluation of training programs are discussed. This report also serves as an archival document for the teamwork and collaboration assessments that have been developed at ETS for high school students.

Key words: Teamwork, noncognitive assessment, situational judgment tests (SJT), self- and other report

Acknowledgments

This research was supported, in part, by a Ford Partnership for Advanced Studies (PAS) research grant to ETS's Higher Education and School Assessments division and the final author. We would like to thank the following individuals who were involved in test construction, data collection, and analyses or who gave suggestions on an earlier draft of this manuscript: Anthony Betancourt, Amy Craig, Dan Eignor, Walter Emmerich, David Escoffrey, Nathan Kogan, Patrick Kyllonen, Mary Lucas, Gerard Mannarino, Jennifer Minsky, Frank Rijmen, Ralf Schulze, Elliott Smith, Jonathan Steinberg, Cindy Tocci, and Matthias von Davier. We also wish to thank the administrators, scientists, teachers, and students of the Ford PAS program, who gave willingly of their time. The views expressed here are those of the authors and do not reflect on ETS, the University of Notre Dame, Rutgers University, or Ford PAS.

Table of Contents

Appendixes

List of Tables

Page

List of Figures

Page

Prologue

The impetus for the development of a teamwork and collaboration assessment for high school students came from four independent sources (both internal and external to ETS, and both scientific and business facing) during early 2006:

1. The award of a small grant to ETS's Higher Education and School Assessments division by the Ford Partnership for Advanced Studies (PAS) to develop self- and teacher reports of this construct (see Appendix A).
2. Marketing research with community colleges' decision makers and stakeholders, which showed that teamwork and collaboration was an important variable that lacked reliable and valid assessments.
3. Ongoing research at ETS to develop noncognitive assessments using the situational judgment test (SJT) methodology.
4. Various scientific articles arguing that multiple assessment methods for the assessment of noncognitive qualities might address issues such as faking and response distortion (or limit the impact of these confounding variables, at the very least). These studies also showed that different methods provided incremental validity over and above an assessment used in isolation.

The literature review and study reported in the passages that follow aim to serve these multiple purposes without sacrificing scientific integrity. An ancillary aim of this report is to archive the teamwork and collaboration assessments. Please note that these instruments are subject to copyright; if you wish to use them please contact Richard D. Roberts, Center for New Constructs, R&D, MS 16-R, ETS, Rosedale Road, Princeton, NJ, 08541, USA. E-mail: RRoberts@ets.org.

Literature Review

> Teams are a fact of life. From medicine to aviation to the policeman on the beat, from management to modern warfare to the Superbowl clash, teams carry out much of the work in our world. (Brannick & Prince, 1997, p. 3)

Team-based settings have become a fact of life in many organizations as skill diversity and high levels of specific expertise increase among employees and the resulting knowledge

needs to be integrated (Kozlowski & Ilgen, 2006). Similarly, in the educational system, an emphasis on teamwork is increasing. Professors and teachers frequently assign projects that require student collaboration (Ahles & Bosworth, 2004; Cordes et al., 1995). Teamwork has been touted as one of the major skills comprising workforce readiness in the 21st century (Barton, 2007; Casner-Lotto & Barrington, 2006). Thus, to increase career opportunities, to prepare for the workforce, and to succeed in team environments, students must better understand and possibly improve their teamwork skills (and related competencies).

Although researchers and educators have placed increasing importance on teamwork, progress has been slight in assessing and training teamwork skills. Many studies have been conducted to measure the components of teamwork, but valid assessments still appear meager (Hoegl & Gemuenden, 2001; O'Neil, Wang, & Lee, 2003). In addition, most of the present assessments are targeted at business organizations or college students (Loughry, Ohland, & Moore, 2007; Morgeson, Reider, & Campion, 2005). Fewer studies assess teamwork components in the K–12 domain. The current study aims to develop a multimethod teamwork assessment system that (a) targets high school students, (b) consists of reliable and qualitatively distinct factors, (c) involves a variety of methods, and (d) has demonstrable validity evidence. These instruments could be used to identify high school students' teamwork skills, to design intervention programs around the assessment, and to provide career and personal guidance for students. Information on the different test development procedures that inform assessment efforts and closer examination of the teamwork construct in a high school population may also be useful for the future development of teamwork training curricula.

Conceptualizing Teamwork

Although a number of different definitions of a *team* were found in the literature, three common elements defined a team (e.g., Guzzo & Dickson, 1996; Hackman 1987; Kozlowski & Ilgen, 2006): (a) two or more individuals working together toward a common goal, (b) such individuals occupying different roles within the team, and (c) such individuals thus being interdependent.

Teamwork has been conceptualized at both the team and the individual level, with a team's performance thus conceptualized as both (a) the psychological variables that relate to the performance of the *team* (e.g., team diversity, leadership climate; Carpenter, 2002; Chen, Kirkman, & Kanfer, 2007) and (b) the characteristics of the *individual* that make him or her a

good team player (e.g., personality, value, ability, or conflict-solving skills; Driskell, Goodwin, Salas, & O'Shea, 2006; Kichuk & Wiesner, 1998; Marks, Mathieu, & Zaccaro, 2001; Stevens & Campion, 1999). Because the goal of the study was to assess individuals' teamwork-related characteristics, we focused on the characteristics of individuals.

A number of different models accounting for individual differences in teamwork characteristics have been put forward in the literature (see, e.g., Campion, Medsker, & Higgs, 1993; Janz, Colquitt, & Noe, 1997; Marks et al., 2001; O'Neil, Chung, & Brown, 1997; Stevens & Campion, 1994). For illustrative purposes, we elaborated on two of these models, providing a description of the proposed dimensions of teamwork.

Stevens and Campion (1994) suggested the following five-component model of teamwork involving knowledge, ability, and skills:

1. Conflict solving—the ability to recognize and encourage useful conflicts and to employ appropriate conflict resolution strategies when conflicts are not useful
2. Collaborative problem solving—the ability to identify situations requiring group problem solving and decision making
3. Communication—listening skills and a willingness and ability to develop open and supportive communication
4. Goal setting and performance management—setting acceptable and appropriate goals and providing feedback
5. Planning and task coordination—the ability to coordinate activities with other team members

O'Neil et al. (1997) presented an alternative six-factor model of teamwork:

1. Adaptability—the identification of problems in team activities and the provision of appropriate feedback
2. Coordination—the organization of the team so as to best accomplish a task
3. Decision making—the integration of information to seek and select the best solution
4. Interpersonal characteristics—cooperation and the promotion of a smooth working relationship with team members
5. Leadership—the direction of others, as well as the planning and assignment of tasks

6. Communication—the efficient exchange of information with team members

As is obvious from these two models, conceptualizations of teamwork may differ in the details but have a number of correspondences. Based on several different conceptualizations, including these two models, teamwork might be classified into five content areas (Campion et al., 1993; Janz et al., 1997; Marks et al., 2001; O'Neil et al., 1997; Stevens & Campion, 1994):

1. Task-related process skills (collaborative problem solving, decision making, planning and task coordination, strategy formulation, coordination, goal setting, performance management)
2. Cooperation with other team members (adaptability, interpersonal skills)
3. Influencing team members through support and encouragement (confidence building, social support)
4. Resolution of conflicts or disagreements among team members via negotiation strategies (conflict solving, communication)
5. Guidance and mentorship of other team members (leadership, helping others)

Process skills appear more cognitively loaded than the other four components of teamwork and might perhaps be more appropriately measured with maximum performance tasks or aligned with cognitive ability assessment. In this study, we concentrated only on the more behavioral aspects of teamwork that were appropriate to assess with typical performance. For this reason, we limited our content coverage of individual differences in teamwork to the latter four content areas: (a) cooperation with others, (b) advocating and influencing others, (c) resolving conflict/negotiating, and (d) guiding others.

Assessing Teamwork

Several different types of teamwork assessments have been developed based on the different models outlined in the preceding account. For example, Stevens and Campion (1999) developed a five-dimension SJT. By contrast, Loughry et al. (2007) and O'Neil et al. (2003) developed multidimensional self-report assessments of teamwork skills and competencies. However, these measures were developed for adult populations and had some psychometric limitations (e.g., very high correlations between teamwork dimensions, making it difficult to distinguish subdomains). Thus, an attempt to develop a psychometrically sound measure of

teamwork for high school students represented an expansion of the field. In addition, previous studies have not considered relationships between different methods of measurement (i.e., SJTs, self-reports, and other reports). Given that the practical concerns with each type measured in isolation (i.e., how to score SJTs and how to control for response distortion in self-reports), an examination of the relationship between different methods seemed warranted. For this reason, teamwork measures for high school students were developed in self-report, SJT, and teacher-report formats.

Self-report ratings. The most convenient and cost-efficient method, self-reports are traditionally used in assessing teamwork (O'Neil et al., 2003). Of course, such a method depends on students' capabilities for self-knowledge: Students must have the necessary psychological-mindedness to accurately gauge their own levels of teamwork. However, self-reports have an advantage over teacher reports in that students may consider their behaviors across all the different settings in which they are involved, whereas teachers only observe students in one setting: the classroom.

Situational judgment tests (SJTs). This method is often used as a simulation for job performance or for performance on particular job-related tasks in order to select applicants in certain fields (e.g., Kyllonen & Lee, 2005). These tests may show less adverse impact on ethnic minorities (e.g., McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001) and yet have moderate correlations with measures of cognitive ability, personality, work experience, and job knowledge (McDaniel et al., 2001; McDaniel & Nguyen, 2001). They have demonstrable incremental validity over personality and cognitive ability predictors (Weekly & Ployhart, 2005). Finally, SJTs may be developed to reflect more subtle and complex judgment processes than are possible with conventional tests.

Teacher ratings. Reports by supervisors or teachers are the traditional methods used to assess employee or student performance and skills in organizations and schools. However, teachers' reports are often collected in the form of recommendation letters or annual reports. This method is qualitative and non-standardized, making it difficult to quantify the levels of students' teamwork skills. Moreover, such approaches do not allow a reliable comparison between students with different teachers. For these reasons, a standardized scale was developed for teacher reports of teamwork, with each item giving descriptions of students' teamwork behaviors at three different levels.

The Nomological Net for Establishing Validity Evidence for Teamwork

Several constructs conceptually similar and dissimilar to teamwork defined the nomological network (Cronbach & Meehl, 1955), which served as the basis for evaluating the construct validity of the teamwork assessment system. The first component, evidence for convergent validity, was tested by the relationships among the three teamwork measures, the four-factor structure of the self-report teamwork assessment, and by the convergence of different scoring protocols for the SJT (i.e., the different scoring rubrics are valid if they appear to assess the same thing). The second component, divergent validity, was represented by the distinctiveness of the teamwork measures from the five superfactors of personality: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience (Schulze & Roberts, 2006).[1] Developmental and learning trends formed the third component of the nomological network, with teamwork scores expected to be higher among older students or students who had completed more teamwork-related curriculum units. Criterion-related validity evidence was also tested in the form of relationships between the teamwork assessment system and students' school grades. Further theoretical justification for selecting these relationships as sources of validity evidence is outlined below.

Convergent validity evidence: Relations among different teamwork measures. One obvious form of validity evidence was that the self-report measure of teamwork assessed the intended content dimensions of cooperation, influence, conflict resolution, and guiding others. To this end, structural analyses of the self-report scale were expected to support the proposed content dimensions of teamwork. In addition, an important source of convergent validity evidence was the positive relationship between teamwork assessed by self-report, other report, and SJT. These techniques should measure similar things, with the caveat that the teacher report may only measure observable behaviors, whereas the self-report instruments may assess a wider range of characteristics.

Convergent validity evidence: Expert scoring of the situational judgment test (SJT). Scoring of SJTs by subject matter experts is an integral part of SJT development (e.g., McDaniel & Nguyen, 2001), and we used expert opinion to score the teamwork SJT developed here (i.e., experts identified the best of the four possible response options to each situation). Nevertheless, without additional validity evidence that such scores distinguished between test-takers with high and low levels of the construct measured, some problems remained unresolved. For teamwork, it

was not intuitively obvious who experts might be, or that *one* particular response to a particular situation represented better teamwork than all other responses. For this reason, we used latent class analysis (LCA) as a validity check on the expert scoring. Because this use of LCA was unusual, we describe the logic of our analyses in the following paragraph.

Latent class analysis is a type of statistical analysis used to identify qualitatively different groups of cases within a data set, based on consistencies in their response patterns. Key responses on the teamwork SJT were analyzed using LCA to determine whether discrete variables representing test-takers high in teamwork (who agreed with the teamwork experts) and low in teamwork (who disagreed with the teamwork experts) were present. If discrete groups could be characterized by whether they agreed with the teamwork experts, this would be evidence that using the expert scoring key was reasonable.

Divergent validity evidence: Relations between personality and teamwork. The teamwork assessments should be distinguishable from personality traits to provide evidence of divergent validity (i.e., most of the variation in teamwork scores should be independent of the five personality factors). However, although teamwork assessments should operate independently of personality, scores might be expected to show small relationships with personality, based on prior research findings: In prior research, four of the five superfactors of personality (all except openness) have demonstrated small to moderate relationships with team performance (e.g., Driskell et al., 2006; Morgeson et al., 2005). Thus, we expected the current teamwork assessment system to be substantially independent of personality (i.e., the proportion of shared variance less than .50) but to demonstrate small to moderate correlations (r = .30 to .50) with extraversion, agreeableness, conscientiousness, and emotional stability.

Establishing validity evidence: Relations between teamwork and academic achievement. Students' learning and achievement may relate to social as well as cognitive demands of academic life. With more and more teachers stressing a collaborative approach to learning and assessment, teamwork skills are becoming a more important tool for students' academic success (Ahles & Bosworth, 2004; Cordes et al., 1995). Criterion validity evidence for the three measures of teamwork in high school students might thus be judged by the relationship of these measures to students' grades.

Aims of the Current Study

The aims of the current study were threefold:

1. To develop a suitable multiple method assessment system to evaluate the teamwork skills of high school students
2. To provide preliminary reliability and validity evidence for each of the constructs comprising these assessments by examining each test in isolation (i.e., thus determining structural validity) and then the relationship between the tests (i.e., convergent validity evidence)
3. To provide additional validity evidence for the assessments by examining relationships between teamwork skills and personality and the relationship between the various teamwork measures and academic outcomes such as grade point average (GPA)

Method

Participants

All participants were high school students undertaking Ford PAS courses in workforce readiness (N = 159, 82 female). The mean age for these participants was 16.10 years (SD = 1.03). The ethnic composition for this sample was 18.9% White non-Hispanic, 64.2% African American, 3.1% Hispanic, 3.1% multiethnic, with the remaining 10.7% reporting their ethnicity as American Indian, Asian, or Other. While this sample was not representative of U.S. high school students, it was consistent with the student population that is generally targeted by the Ford PAS program.

Measures

Teamwork skills were measured by multiple evaluation methods: (a) student self-report rating, (b) SJT, and (c) teacher-report rating.

Student self-report teamwork assessment. Students completed 57 items on a 6-point scale from 1. *never* to 6. *always.* Fourteen items were reverse-coded. (See Table 1 for the items that survived further analysis. All items comprising the original scale are given in Appendix B, section 5.) These 57 items were designed to assess the four dimensions identified as central in our review: cooperation with others (15 items), advocate and influence (12 items), resolve conflict/negotiate (17 items), and guiding others (13 items).

Table 1

Factor Loadings of the Revised Student Self-Report Teamwork Scale

Items	F1	F2	F3	Mean *(SD)*
I enjoy bringing team members together	.75			4.45 (1.29)
I share ideas with others to accomplish a task	.74			4.58 (1.20)
I acknowledge the accomplishments of my peers	.71			4.71 (1.19)
I enjoy helping team members	.67			4.77 (1.19)
I value different perspectives to help me strengthen my understandings of issues or problems	.67			4.40 (1.26)
I provide appropriate feedback to team members	.60			4.16 (1.18)
I think that exchange of ideas among team members can lead to creative solutions	.60			4.84 (1.26)
I cooperate with other students	.53			4.72 (1.12)
I like team activities	.46			4.43 (1.33)
I am inspired by others' ideas and thoughts	.46			4.09 (1.25)
I contribute to the definition of a team's goals	.43			4.50 (1.22)
I respect the opinions of my peers	.41		.34	4.76 (1.11)
I like to be in charge of groups or projects		.70		3.75 (1.42)
I know how to make other students see things my way		.65		4.03 (1.12)
I can convince my peers about anything		.57		3.80 (1.19)
I believe I am a good leader		.56		4.68 (1.36)
I am comfortable providing constructive criticism		.44		3.80 (1.46)
I carefully consider the facts to persuade my peers		.43		4.30 (1.29)
I seek to influence my peers		.41		3.61 (1.44)
I suggest alternative solutions to problems		.30		3.99 (1.23)
My arguments are constructive		.23		4.15 (1.27)
I am a good listener			.72	4.91 (1.11)
I am open to varying opinions			.65	4.64 (1.23)
I take other students' interests into account			.55	4.32 (1.28)
I adapt to change well			.53	4.16 (1.15)
I am flexible in team situations			.51	4.68 (1.16)
I believe that there is only one "best" solution			.49	4.38 (1.37)
I dislike it when people challenge my views [a]			.38	3.92 (1.29)
I understand that each team member is different	.30		.31	5.33 (1.04)
I consider team members first			.29	3.98 (1.36)

Note. F1 = cooperation; F2 = advocating/influence; F3 = negotiation.

[a] Reverse-scored item.

Situational Judgment Teamwork Assessment. This assessment presented eight scenarios describing various situations and required participants to evaluate the effectiveness of four reactions to each situation. Effectiveness was evaluated on a 5-point scale, from 1. *very ineffective* to 5. *very effective.* (See Appendix B, section 6.) For each item, a three-person team of item developers decided which reaction was most effective, and the test-taker's rating of that reaction was the score awarded for each option (e.g., if Option C was the most effective reaction for Scenario 3 and a test-taker awarded C a 4 out of 5, he or she would score 4 for Scenario 3. The scoring key in the figure title of Figure 4 shows which response option was most effective for each scenario.) We also tried the standard distance-based scoring algorithm (calculating the distance between a student's responses and the experts' ratings of the four reactions for each scenario). Although we did not report the standard method, for reasons of space, the results indicated that rating of the most effective response was an empirically stronger method (i.e., had higher reliability and more reasonable correlations among different scenarios). In addition, the method we chose was also more appropriate for the sample size and for addressing issues of experimental dependence in the SJT format.

Teacher-report behaviorally anchored rating scale. Teachers evaluated each student's level of teamwork against 10 behaviorally anchored five-point Likert scale items, with descriptors at points 1, 3, and 5. (See Appendix C, part 2.)

Ten Item Personality Inventory. The Ten Item Personality Measure (TIPI; Gosling, Rentfrow, & Swann, 2003) was used to measure conscientiousness, extraversion, agreeableness, emotional stability, and openness. (See Appendix B, section 3.)

Student self-report grades. Student self-report course grades in reading/language arts, math, science, social science, art, and music subjects were also collected as criterion information. (See Appendix B, section 2.) One hundred fifty-one grades were reported for reading, 150 for math, 146 for social science, 135 for social studies, 49 for art, and 41 for music.

Teacher brief report of students' academic skills. Teachers rated students' academic skills (e.g., reading, writing, speaking effectively) on five 5-point rating-scale items. (See the first five items in Appendix C, part 1). Ratings ranged from 1 *below average* to 5 *truly exceptional*.

Teacher brief report of students' teamwork skills. Teachers rated students' teamwork skills (e.g., cooperation, leadership, conflict resolution) on four 5-point rating-scale items. (See Items 6–10 in Appendix C, part 1). Ratings ranged from 1 *below average* to 5 *truly exceptional*.

Teacher brief report of students' social skills. Teachers rated students' social skills (e.g., understanding oneself, cultural sensitivity) on four 5-point rating-scale items. (See Items 11–14 in Appendix C, part 1). Ratings ranged from 1 *below average* to 5 *truly exceptional*.

Procedure

Participants were tested in class with teachers reporting on participants' performance during this time. The test protocol was approved by ETS's institutional review board (IRB). All measures also passed fairness review.

Data Analysis Steps

Separate exploratory factor analyses (EFAs) using principal factor analysis with promax rotation were conducted for the student self-report scale, SJT (eight key items selected by experts), and teacher-report scale to detect some problematic items (negative loading or low loading items) and to identify the factor structure of each scale. Scree plots (Cattell, 1966) and parallel analyses (e.g., Liu, Rijmen, & Kong, 2007; O'Connor, 2000) were used to determine the number of factors for each scale. Parallel analysis is a simulation-based approach and has been proven to be one of the most effective methods for determining the number of factors to retain. The underlying rationale for parallel analysis is that the eigenvalues of the salient factors from real data with a valid latent factor structure should be larger than the eigenvalues of the corresponding factors generated from random data with the same sample size and number of variables. A SAS macro program developed by Liu et al. (2007) was used to conduct the parallel analysis. Confirmatory factor analyses were then employed to test the factor structures and compare the goodness of fit for alternative factor models.

The SJT was analyzed with latent class analysis (LCA; e.g., Bartholomew, 1987; McCutcheon, 1987) as well as EFA. LCA was applied as an exploratory technique to detect some potential classes for the eight key items selected by experts. Only the eight key items could be used in the LCA analysis because the small sample size and nested item-scenario data structure meant that a model based on all 32 responses would be not reliable. Often referred to as

a categorical data analogue to factor analysis, LCA is a powerful and flexible technique to reduce the dimensionality of categorical data.

Convergent validity evidence was then investigated by looking at the relationship among student self-report factor(s), teacher-report factor(s), and SJT factor(s) or the latent classes from SJT items. Divergent validity evidence was assessed via the correlation between teamwork and the five factors of personality; evidence of criterion-related validity was assessed via the correlation between teamwork measures and both course grades and teacher brief reports of academic skills, teamwork, and social skills.

Results

Dimensionality of Each Scale

Student self-report scale. The initial EFA was conducted on a matrix of Pearson correlations. Visual inspection of the scree plot and parallel analysis both showed a four-factor solution. Figure 1 graphically displays the real eigenvalues, mean of parallel eigenvalues, and 95th percentile eigenvalues, illustrating how a four-factor solution was selected. A four-factor EFA was then conducted. However, results indicated some potentially problematic items due to negative factor loadings, items that did not load saliently (> .30) on any factor, or items that showed multiple loadings (> .40) on different factors. In addition, these four factors did not correspond to the four subscales originally designed. A sequential deletion method was then used to eliminate problematic items. Items with negative loadings were first eliminated and then items with low loadings and multiple loadings were eliminated. On the basis of item analyses, 27 items (13 of which were reverse-coded) were dropped.[2] An EFA was conducted on these discarded items and no clear factor pattern was identified.

The remaining items were then reanalyzed using EFA.[3] Results from the scree plot and parallel analysis are shown in Figure 2 and indicate a three-factor solution. Factor loadings from this analysis are shown in Table 1.

The reduced self-report scale consisted of 30 items and captured three factors: (a) cooperation (12 items), (b) advocating/influence (9 items), and (c) negotiation (9 items). The cooperation factor included ability and skills to bring ideas together, seek solutions, and provide feedback to team members. The advocating/influence factor referred to ability and skills to direct others, provide appropriate suggestion and criticism, and persuade others. The negotiation factor included students' skills at listening, ability to adapt to change during conflicts, and ability to

solve conflicts. A separate guiding others factor did not emerge, with items developed for this construct instead loading on the advocating/influence factor. Cronbach alphas for the items measuring the three factors were .88, .80, and .78, which were acceptable. Table 1 also provides the mean and standard deviation of each item.

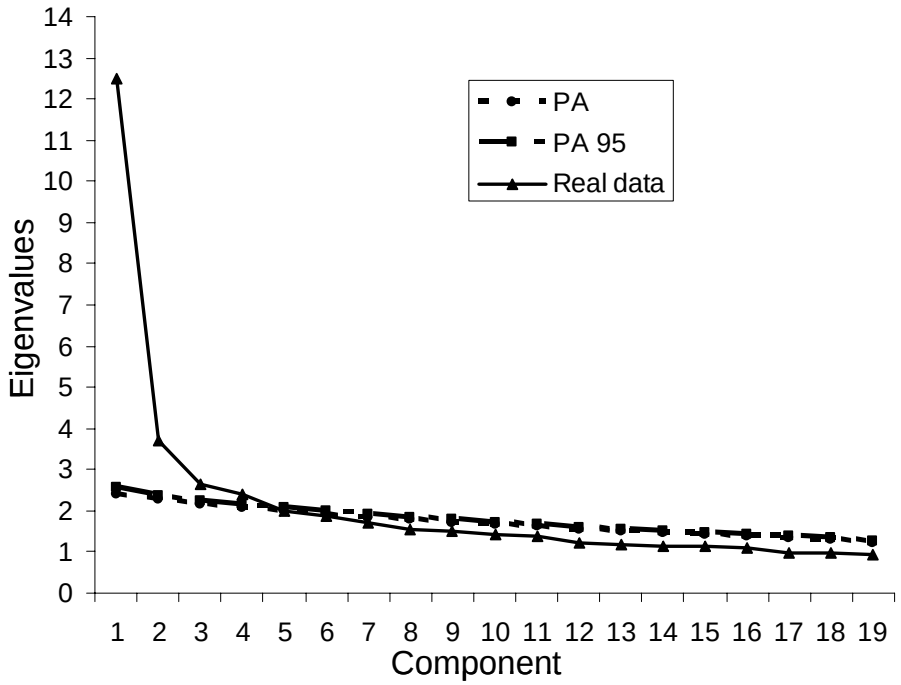

Figure 1. **Eigenvalue plot of parallel analysis results for the student self-report scale using four factors.**

Note. Parallel analysis was conducted with 1,000 replications. Both the mean (PA) and 95th percentile (PA95) of the eigenvalues from the generated random data were included. The values are very close as evidenced by the overlaps in the graph. Four factors from the real data had larger eigenvalues than the factors generated from random data.

A confirmatory factor analysis (CFA) using Mplus (Muthén & Muthén, 1998–2007) was used to investigate the adequacy of the factor structure. The items were assigned to one of the three dimensions based on the factor loading matrix from the EFA shown in Table 1. The results indicated that the model was moderately supported by the fit indices: comparative fit index (CFI; Bentler, 1990) = .85 and root mean square error of approximation (RMSEA; Browne & Cudeck, 1993) = .06. The correlations among the latent variables were .66 (cooperation, advocating/

influence), .79 (cooperation, negotiation), and .59 (advocating/influence, negotiation). Note that these correlations between latent variables were higher than correlations between summed scores, since measurement error had been partialled out of the latent variable estimates using structural equation modeling. However, correlations between factors were still quite high, so we investigated whether a two-factor model might be a better fit to the data. Three alternative two-factor models were fitted to test if the three factors were statistically distinct: 1. cooperation and advocating/influence combined to form one factor (i.e., the correlation between them was set to 1), 2. cooperation and negotiation combined to form one factor, and 3. advocating/influence and negotiation combined to form one factor. Results of the likelihood ratio tests are shown in Table 2. Results indicated that combining any two of the factors into one factor would significantly lower the model fit. Therefore, the three factors identified from the EFA and CFA were reasonably reliable, moderately correlated, and statistically distinct.

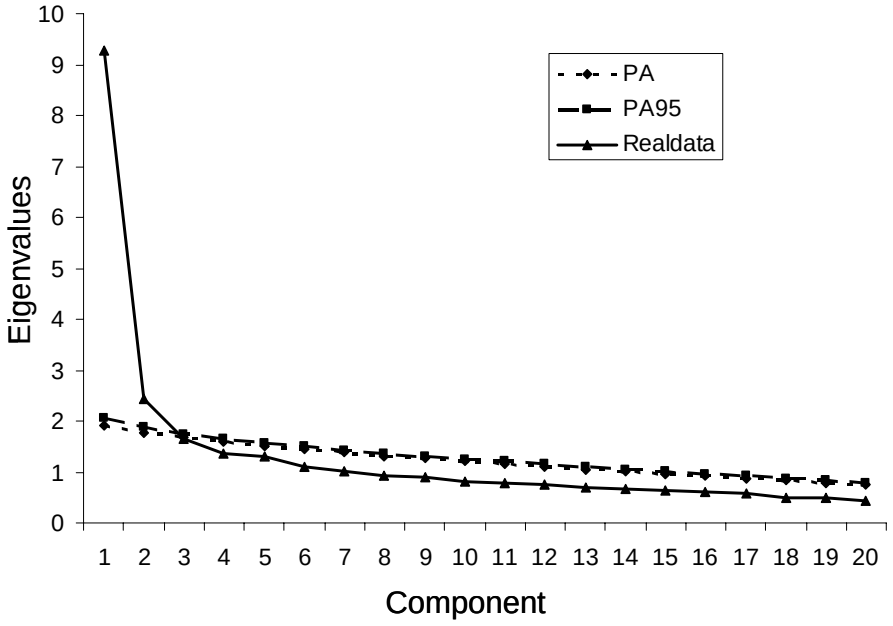

Figure 2. **Eigenvalue plot of parallel analysis results for the student self-report scale using three factors.**

Note. Parallel analysis was conducted with 1,000 replications. Both the mean (PA) and 95th percentile (PA95) of the eigenvalues from the generated random data were included. The values are very close as evidenced by the overlaps in the graph. Three factors from the real data had larger eigenvalues than the factors generated from random data.

Situational judgment test (SJT). Scree test and parallel analysis with EFA showed that the SJT was unidimensional, with the first factor explaining 26% of the variance of the eight key items. The Cronbach alpha of these eight items was .71. The results of the CFA indicated that the model was adequately supported by the goodness of fit indices (CFI = .96, RMSEA = .049).

Table 2

Fit Indices From the Confirmatory Factor Models for the Student Self-Report Teamwork Scale

Fit indices	3-factor model	Correlation between factors fixed to 1.00		
		F1, F2	F1, F3	F2, F3
Chi-square	651.5	743.8	716.1	741.6
df	402	403	403	403
CFI	.84	.79	.81	.80
RMSEA	.06	.07	.07	.07
Likelihood ratio test		92.3/1	64.6/1	90.1/1

Note. F1 = cooperation; F2= advocating/influence; F3 = negotiation; CFI=comparative fit index (Bentler, 1990); RMSEA= root mean square error of approximation (Browne & Cudeck, 1993).

Unrestricted LCA models with one to four classes were fitted to the eight key SJT items, with raw data used as input. Table 3 displays the goodness of fit indices for these LCA models. Based on the Akaike information criterion (AIC) results, a three-class LCA model fit the data best. However, the Bayesian information criterion (BIC) results suggested a two-class solution. Profile plots (means ratings of the 32 item responses for each class) were generated for the two-class LCA model (see Figure 3) and the three-class LCA model (see Figure 4).

Table 3

Goodness of Fit Indices for One- to Four-Class Latent Class Analysis (LCA) Models

Latent class	Log likelihood	Number of parameters	AIC	BIC
1	-1,541.1	31	3,144.2	3,238.5
2	-1,428.4	63	2,982.9	3,174.6
3	-1,382.9	95	2,955.9	3,245.0
4	-1,361.4	127	2,976.7	3,363.2

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion.

The plots show that Class 1 and Class 2 in the three-class model were not visually distinguishable from each other. However, the two classes from the two-class model discriminated very clearly. Combining the results from AIC, BIC, and profile discrimination abilities, a two-class model was selected to represent the dimensionality of the SJT scale (one dimension, two classes). From Figures 3 and 4, we can also see that students in Class 1 rated the eight key items that experts thought were best reactions more effectively than those in Class 2 did. This result provided an empirical form of validity evidence for the expert keying. Class 1 showed more variations across reactions within each scenario than Class 2, which indicated that Class 1 students could better differentiate between different reactions within each scenario than Class 2 students. Therefore, we labeled Class 1 as the high teamwork skill group and Class 2 as the low teamwork skill group.

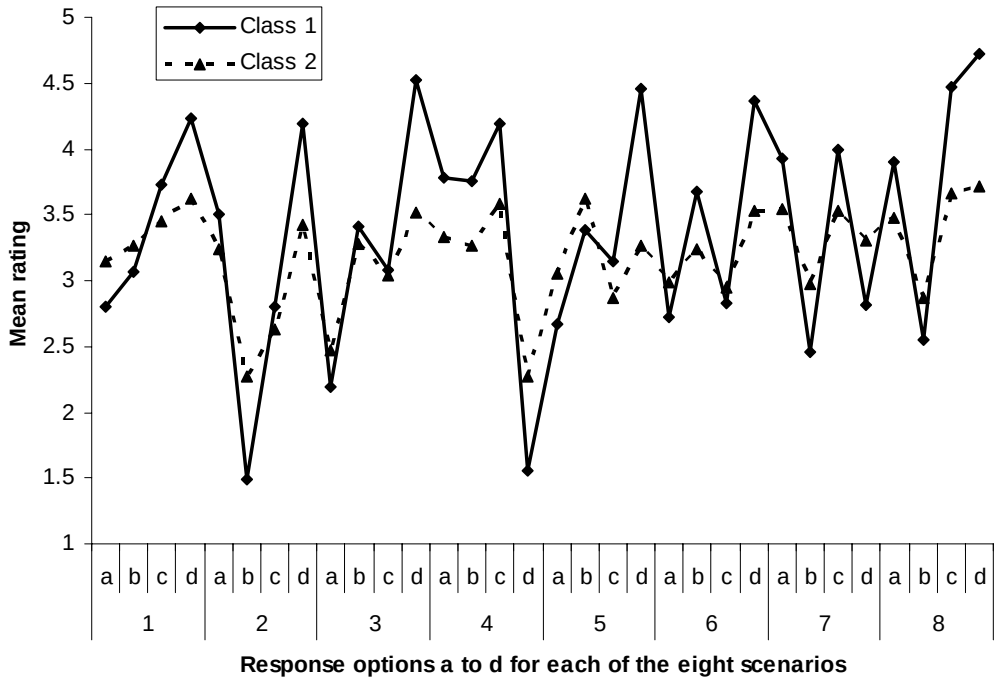

***Figure 3*. Mean profile plots from the latent class analyses (LCA; two classes).**

Note. The y-axis represented the rated means of effectiveness of each reaction across latent classes. The x-axis represented the reaction numbers (a, b, c, and d) and scenario numbers (1–8). For Scenarios 1–8, the most effective reactions rated by experts (key items) are c, d, d, c, d, d, c, and d.

Teacher-report scale. The scree test and parallel analysis with EFA showed that the teacher evaluation scale was unidimensional, with the first factor explaining 83% of the variance of the 10 items. The Cronbach alpha of the scale was .98. The mean of the composite scores of these 10 items was 3.35 (*SD* = 1.14, range = 1–5). Because 35 cases showed zero variance across 10 items, another EFA was conducted without these 35 cases. Similarly, a one-factor solution was detected explaining 72% of the variance, with a Cronbach alpha of .96. Therefore, we concluded that the teacher-report scale was unidimensional.

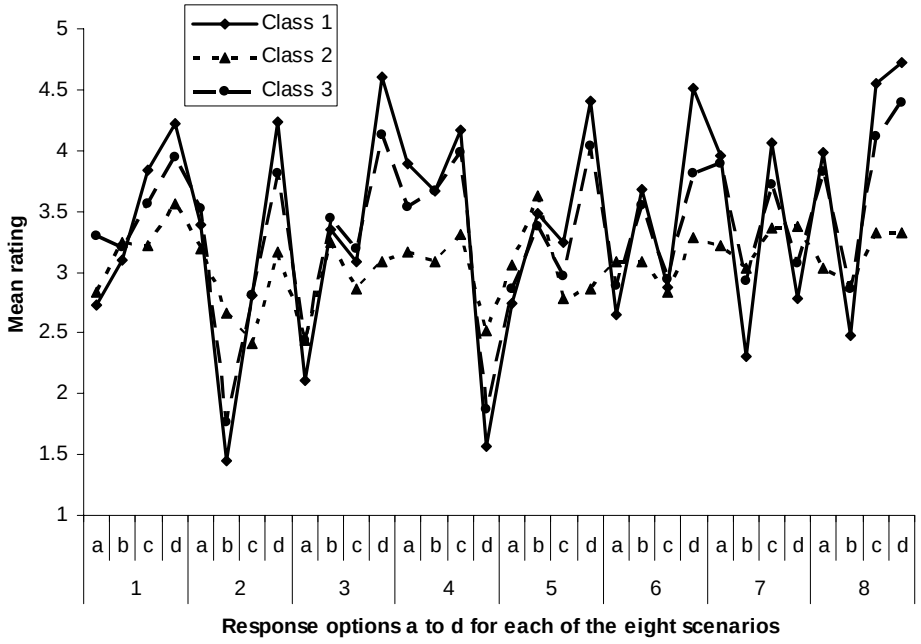

Figure 4. **Mean profile plots from the latent class analyses (LCA; three classes).**

Note. The y-axis represented the rated means of effectiveness of each reaction across latent classes. The x-axis represented the reaction numbers (a, b, c, and d) and scenario numbers (1–8). For Scenarios 1–8, the most effective reactions rated by experts (key items) are c, d, d, c, d, d, c, and d.

Relationships Among the Three Evaluation Methods

Table 4 displays the correlations of the five teamwork scales (cooperation, advocating/influence, negotiation, SJT factor, and teacher-report factor). The SJT assessment was significantly correlated with all of the three student self-report assessments and the teacher report. Teacher-report scores were correlated with two of the student self-report scores: cooperation and advocating/influence.

Table 4

Correlations Among Three Teamwork Evaluation Methods

	Student self-report			SJT	Teacher report
	Cooperation	Advocating/ influence	Negotiation		
Cooperation (12)	**.88**				
Advocating/influence (9)	.66**	**.80**			
Negotiation (9)	.79**	.59**	**.78**		
SJT (8)	.52**	.47**	.60**	**.71**	
Teacher report (10)	.19*	.32**	.14	.33**	**.98**

Note. The numbers in the parentheses are the numbers of items for each scale/subscale. The bolded numbers on the diagonal are the alpha reliabilities. SJT = situational judgment test.

*$p < .05$. **$p < .01$.

In order to further investigate the relationship among the three assessment methods, self-report and teacher-report composite scores (observed mean scores) were also compared between two latent classes. (See Table 5 for results.) The high teamwork skill class had significantly higher means for all of the three subscales of the student self-report scale than the low class. However, this difference was not observed for the teacher ratings. The results also provided some construct validity evidence for the student self-report and SJT scales.

Table 5

Teamwork Assessments Compared by Latent Classes

	High teamwork skills (N = 84)	Low teamwork skills (N = 71)	
	Mean *(SD)*	Mean *(SD)*	*t*
Cooperation (12)	4.74 (.74)	4.32 (.81)	3.38**
Advocating/influence (9)	4.14 (.81)	3.88 (.78)	2.00*
Negotiation (9)	4.59 (.68)	4.36 (.78)	2.01*
Teacher report (10)	3.43 (1.05)	3.28 (1.25)	0.81
SJT (8)	4.27 (.35)	3.50 (.44)	11.85**

Note. The numbers in the parentheses are the numbers of items for each scale/subscale. SJT = situational judgment test.

*$p < .05$. **$p < .01$.

Relationships Between the Teamwork Assessments and Demographic Variables

No significant gender differences were found for the three student self-report subscales, teacher-report scores, or SJT scores. In addition, no significant gender differences were found between the two latent classes. Age was positively correlated with the three student self-report subscales (r = .31 to .35, $p < .01$) and SJT scores (r = .32, $p < .01$) but not significantly with teacher-report scores. The mean age in high teamwork skills class (M = 16.31, SD = 0.96) was significantly larger (t = 2.42, $p < .05$) than the mean age in low teamwork skills class (M = 15.91, SD = 1.07). Age was a significant predictor of all teamwork measures when controlling for the number of Ford PAS modules a student had undertaken (r = .25 to .30 for self-report measures, .28 for the SJT, and .18 for the teacher report; $p < .05$ in all cases). However, the number of Ford PAS modules undertaken was not a significant predictor of any teamwork measure after controlling for age. No significant ethnic group differences were found for any of the measures.

Relationship Between the Teamwork Assessments and Personality

Correlations between TIPI personality scores and the teamwork assessments are shown in Table 6. No significant correlations with personality were found for either the teacher reports or the SJTs. Nor were there any significant differences in personality measures between the two latent classes. For the student self-report scale, the cooperation factor was positively correlated with agreeableness, conscientiousness, and openness. By contrast, the advocating/influence factor was positively correlated with extraversion, conscientiousness, and openness, while negotiation correlated positively with agreeableness, conscientiousness, emotional stability, and openness.

Relationship Between the Teamwork Assessments and Course Grades

The SJT scores did not correlate significantly with students' grades (although correlations were in the expected direction). The teacher-report scale correlated significantly with math, science, and social studies grades (r = .21, .30, and .27 respectively; $p < .01$). Cooperation correlated moderately with science and music grades (r = .18 and .38; $p < .05$), while advocating/influence correlated positively with science (r = .32; $p < .01$), social science (r = .19; $p < .05$), and music grades (r = .40; $p < .01$). Negotiation shared a positive correlation with music grades only (r = .50; $p < .01$). A grades composite was calculated by taking the means of the different course grades. Only advocating/influence scores of the student self-report scale and

teacher-report scores were significantly correlated with the grades composite scores (r = .25 and .25; $p < .01$).

Table 6

Correlations Between Teamwork Assessment and Big-Five Personality

	Extraversion	Agreeableness	Conscientiousness	Emotional stability	Openness
Cooperation	.15	.19*	.33**	.07	.23**
Advocating/influence	.18*	-.01	.20*	.01	.30**
Negotiation	-.07	.38**	.34**	.30**	.35**
Teacher report	.07	.02	.07	.14	-.07
SJT	.06	.10	.12	.13	.14

Note. SJT = situational judgment test.

*$p < .05$. **$p < .01$.

The predictive power of three teamwork skills assessments in determining courses' grades composite scores over and beyond the personality measures was also examined. The model with only personality measures (Big-Five variables) explained 12.8% of the variance of grades composite scores ($p < .01$). The cooperation, advocating/influence, negotiation subscales, SJT scores (eight key items), and teacher-report scores added 0.1% ($p > .05$), 3.2% ($p = .02$), 0.2% ($p > .05$), 0.3% ($p > .05$), and 4.2% ($p < .01$) to the explained variance, respectively.

Brief teacher reports of academic attributes and teamwork. Teacher brief reports of academic skills, teamwork skills, and social skills were all reliable (Cronbach alpha = .96, .96, and .96 respectively) but highly correlated among themselves (r = .87 for academic and teamwork, .92 for teamwork and social skills, and .88 for academic and social skills). As was the case for the teacher-report behaviorally anchored rating scale, the teacher-rated brief reports contained 29 cases with no variance (e.g., teachers selected all 2s or all 3s to describe a single student). Correlations between these brief reports and the teamwork assessments are shown in Table 7. All three brief assessments showed small correlations with the SJT and the advocating/influence factors as well as very high correlations with the teacher-reported behaviorally anchored rating scale.

Discussion

The three different assessment tools related to each other were relatively independent from personality and also predicted some variation in school grades. Generally, the aims of the study were met: A multiple-method assessment system was developed for high school students, and this assessment showed evidence of convergent, divergent, and criterion-related validity. Although all teamwork assessments were related, the different methods of assessment presented some distinguishing features with related implications for the uses and purposes of such assessments. These issues are discussed in the next section.

Table 7

Correlations Between Brief Teacher Reports and Teamwork Measures

	Cooperation	Advocating/ influence	Negotiation	Teacher behaviorally anchored rating scale	SJT
Brief report: academic skills	.10	.22**	.01	.79**	.19*
Brief report: teamwork skills	.08	.20**	.00	.87**	.27**
Brief report: social skills	.06	.16*	-.02	.83**	.19*

Note. SJT = situational judgment test.

*$p < .05$. **$p < .01$.

Convergent Validity Evidence: The Relationship Among the Teamwork Measures

The strongest relationship between self- and teacher reports was for the advocating/influence scale, perhaps because this factor was more obvious to external observers (cooperation and negotiation are less open to observation than leadership-like behaviors). An important corollary to this point was that if educators wish to measure constructs that are not obviously and frequently observable in a student's behaviors, they may need to avoid reports from teachers and use self-reports, reports from observers who have known the student very well for many years (e.g., family members), or reports from peers who have worked with the student in a team environment. Indeed, reports from peers who have worked in a team with the student might be an ideal expansion of the criterion space for future studies. The SJT also related less strongly to the teacher reports than to the three self-report scales, again indicating that teachers might not have been able to observe some important aspects of teamwork. Generally, results

supported the construct validity of the scales but indicated that different measurement methods may capture different aspects of the teamwork construct.

Divergent Validity Evidence: The Relationship Between Personality and Teamwork

Correlations between personality and teamwork were observed only for the self-report teamwork assessment, and these correlations were never more than moderate ($r < .40$). The teamwork assessments appeared to be tapping something outside of the five-factor personality space, although using the very brief personality measure (two items per personality dimension) made this conclusion preliminary.

Openness to experience was the personality trait most consistently related to teamwork, significantly correlating with all three self-report scales of teamwork. This finding contrasted with prior research, where openness was the only factor unrelated to teamwork (Driskell et al., 2006; Morgeson et al., 2005), perhaps because the TIPI five-factor measure stressed the creative, exploratory, unconventional aspects of the openness dimension rather than the intellectual aspects that openness-to-experience measures commonly assess (TIPI items were "open to new experiences/complex" and "conventional/uncreative" [reverse-keyed]). Further research examining the facets as well as the broad dimensions of personality may clarify the relationship between teamwork and the openness-to-experience construct.

Validity Evidence: The Developmental Trajectory of Teamwork Over Late Adolescence

This study also found significantly positive correlations between age and both self-report subscales and SJT scores. Because these students were involved in teamwork courses run by the Ford PAS, results could have been due to learning effects (as the students in higher grades had taken more courses than the students in lower grades). However, correlations remained significant when controlling for the number of Ford PAS modules taken (and indeed, increased in significance for the teacher-report assessment), indicating that increases were due to maturation rather than exposure to the course modules. This outcome has important implications for using such assessments to evaluate program effectiveness, since increasing teamwork would have to be compared to natural developmental gains. That said, it should be noted that this study was not a controlled intervention study. Therefore, future studies might be undertaken to investigate some intervention effects (e.g., designing some special curriculum on improving

students' teamwork skills) by using a pre–post and control group design and having the multiple-method assessment of teamwork serve as the dependent variable(s).

Validity Evidence: Does Teamwork Predict Academic Achievement?

We also found that self-report teamwork skills were correlated with different courses' grades to a disparate extent. Only the teacher-report scale and the advocating/influence self-report scale predicted the grades composite over and above the Big-Five personality measures. Prediction of grades from the teacher report was not surprising, as academic performance is usually evaluated or determined by teachers. In fact, teachers may use their knowledge of a student's academic standing as a cue when reporting that student's teamwork skills. This idea was supported by the high correlations between teachers' brief reports of students' academic skills and teamwork skills (r = .87).

Teamwork differentially predicted grades across subjects, with the strongest relationships found for music (r = .38, .40, and .50 for the self-report scales). Although music had the lowest number of cases (N = 41), this strong relationship with teamwork seemed to make conceptual sense. Of all the subjects measured, academic performance in music depends most on teamwork: Playing pieces as a group forms an essential part of the subject, with the negotiation of piece choice, solos, group tempo, and rhythm; even such things as group practice times and places play a role in the final performance and grade. Such a focus on team performance was not essential for other subjects reported in this study, although it might be necessary for other performing arts such as drama or dance, or group activities such as debating or team sports. In essence, overall GPA may be too broad a variable for teamwork to predict, but teamwork may be more useful in predicting carefully considered teamwork-related aspects of academic performance.

Comparison of Self-Report, Situational Judgment Test (SJT), and Teacher-Report Methods of Measurement

There are advantages and disadvantages to each of the three methods of assessment used in this study, and an argument could be made that these complement and offset each other. The primary problem with self-reports (both the rating scale and the SJT) is that these may be susceptible to impression management—students may "fake good" or under some circumstances even "fake bad" on self-report scales. Teacher reports are relatively more objective and may reduce faking problems. However, it is unlikely that a teacher could observe all aspects of

students' teamwork skills, thus teacher ratings may not cover the entire content spectrum of the teamwork construct (Beitchman & Corradini, 1988; Youngstrom, Loeber, & Stouthamer-Loeber, 2000). In addition, teacher ratings of multiple students may be difficult to implement in practice, given the teacher's workload of such an activity in a typical classroom of 20–30 students. Indeed, the zero variability found for 35 teacher-report cases in this study might indicate that teachers were fatigued by this assessment procedure.

Another potential issue with the teacher ratings is the "halo effect" (Thorndike, 1920; Nisbett & Wilson, 1977), whereby one positive impression influences many ratings of many theoretically distinct constructs. Several sources of evidence indicated that halo might have occurred in the current study: (a) high correlations between the brief teacher reports and behaviorally anchored teacher reports, (b) high correlations among the three teacher-rated brief reports, and (c) substantial number of zero variability teacher reports of both the brief teacher ratings and behaviorally anchored teacher ratings.

Despite the problematic effects of halo, a possible advantage of teacher assessments is that ratings are not confounded by the verbal ability of the student. Students who have low verbal ability or are new to the English language may not understand some items and thus may provide answers indicating poor teamwork due to lack of language comprehension rather than poor teamwork. This potential confound may be particularly problematic for the SJT, where students must read through large amounts of text. However, presenting SJT items via video, audio, or still pictures, rather than text, may ameliorate this problem.

In addition, SJTs may detect subtle judgment processes by asking participants to provide intuitive or contextual judgments about ecologically valid scenarios. This ecological validity may also make SJT items more engaging than traditional self-report questionnaires, although such context-rich material makes objective standards for scoring difficult, such that expert scoring must instead be used. In this study, the LCA of students' responses provided independent confirmation of the expert judgments. Classes who agreed and disagreed with the expert scoring key were identified, providing some important evidence for the validity of the expert judgments. This novel approach to checking the validity of expert opinion could be usefully extended outside the teamwork sphere to SJTs assessing workplace competencies, tacit knowledge, or social and emotional intelligence. Although the small sample size in this study made a multilevel analysis of all 32 responses (four ratings for each of the eight scenarios) impossible, such

multilevel LCA models could be developed in future research to accommodate all responses in a ratings-based SJT.

Future Applications for Teamwork Research in High Schools

Subject to certain caveats, this teamwork assessment system might be applied in high schools in several useful ways. Firstly, this multiple-method assessment might be used for early identification and primary intervention, as deficits in teamwork could potentially harm students' higher education, career opportunities, and quality of life. Certainly, it would be beneficial to identify students with deficits in teamwork sooner rather than later and provide appropriate remediation. Students with high teamwork skills might be selected as mentors or role models for students with low teamwork skills, with study or project groups composed accordingly. In addition, feedback and suggestions for improvement might be given to students based on their own profiles of teamwork skills.

Secondly, the instrument might be used to gauge the effects of training. Teamwork training programs are already implemented in many schools (the obvious example being the Ford PAS program in which our sample was involved). The multiple-method assessment could help determine which particular aspects of teamwork see the most growth from training (i.e., negotiation, cooperation, or advocating/influence) and thus be used to fine-tune the emphasis of the training programs.

Thirdly, the instrument might be used to select team members for high-stakes activities where teamwork is particularly important.

Finally, the instrument might be used as a form of career guidance or advice in conjunction with cognitive tests and interest inventories. For example, students with very high negotiation skills might be directed toward courses or careers where these skills might prove useful.

Overall, this study suggests some promising new directions in teamwork research and its application in high schools. A reliable multiple-method teamwork assessment system was developed, with quite promising validity evidence. Such an instrument might profitably be used for identification, training evaluation, or selection purposes in high schools, with multiple-methods a useful technique for overcoming the practical limitations evident in giving a single assessment in isolation.

Epilogue

The approach adopted herein—where multiple methods were developed to assess a construct—and its apparent success have important implications for future directions that ETS might take in developing assessments for all noncognitive constructs. Seemingly, a combination of self-reports, other reports (peer, teacher, family member), situational judgment, and still further methodologies (including the implicit association, conditional reasoning, and a range of information processing paradigms) offer promise in increasing prediction of valued outcomes (e.g., GPA, retention) while simultaneously getting around noteworthy problems often associated with any assessment given in isolation (e.g., faking, lack of insight). Indeed, multiple methods may (and probably should) feed into a number of products currently in development (e.g., LearnerSnapshot™ and Ready Edge™ platforms) as well as in basic research programs (e.g., an Army Research Institute contract to develop assessments of emotional competence).

The current project was also a catalyst in the development of an upward extension of the teamwork SJT for community colleges. That instrument has since gone through a number of revisions and data collection efforts, which will be the subject of a separate report. Of note, feedback and action plans that are tailored to individual users are now developed in concert with each new noncognitive assessment we create; conceivably this could be done for the current teamwork assessment suite.

Relevance to the Ford Partnership for Advanced Studies (Ford PAS) Program

The Ford PAS program is an innovative curriculum designed to prepare high school students for success in their transition to the workforce. The program comprises 15 curriculum modules teaching business, data analysis, communication skills, statistics, engineering, economics, and other topics, with titles such as Ensuring Quality, Reverse Engineering, and Markets Without Borders. Each module is defined by content, which is essentially instructional objectives, such as (know the) difference between data and information and (know how to do) product failure analysis and materials testing. In addition, each module is associated with a set of skills taught (e.g., conduct Internet research, debate using evidence to support a position), teacher and student prerequisites (e.g., PowerPoint, Algebra I), student projects (e.g., redesign a product to increase safety), and a set of six activities (e.g., use market research strategies).

The program content is aligned with national and state curriculum standards in English language arts, social studies, mathematics, science, business education, and educational use of

technology. However, the focus of the program and the feature that sets it apart from other curricula is its emphasis on inculcating a set of general or applied skills that support the acquisition of content knowledge and skills. These four applied skills, identified as *learning pillars*, are critical thinking, problem solving, teamwork, and communication. By acquiring these skills, students will be prepared to learn not only during the school years but throughout their lifespan.

Ford PAS has planned to conduct ongoing evaluation studies of its program beginning in 2008. The assessments developed for this study could well become an integral part of evaluation studies, at least as they apply to one of the learning pillars—teamwork. In particular, they could be used to answer the questions: Does the Ford PAS program lead to an increase in applying teamwork and collaboration skills, and does this learning transfer to later success in finishing school and in the workforce?

References

Ahles, C. B., & Bosworth, C. C. (2004, Spring). The perception and reality of student and workplace teams. *Journalism and Mass Communication Educator*, 42–59.

Bartholomew, D. J. (1987). *Latent variable models and factor analysis*. New York: Oxford.

Barton, P. E. (2007). What about those who don't go? *Educational Leadership, 64*, 26–27.

Beitchman, J. H., & Corradini, A. (1988). Self-report measures for use with children: A review and a comment. *Journal of Clinical Psychology, 44*, 477–490.

Brannick, M. T., & Prince, C. (1997). An overview of team performance measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 3–16). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by a multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

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.

Carpenter, M. A. (2002). The implications of strategy and social context for the relationship between top management team heterogeneity and firm performance. *Strategic Management Journal, 23*, 275–284.

Casner-Lotto, J., & Barrington, L. (2006). *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to 21st century U.S. workforce.* Retrieved March 10, 2008, from the Conference Board Web site: http://www.conference-board.org/pdf_free/BED-06-Workforce.pdf

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276.

Chen, G., Kirkman, B. L., & Kanfer, R. (2007). A multilevel study of leadership, empowerment, and performance in teams. *Journal of Applied Psychology, 92*, 331–346.

Chow, S. C. (Ed.). (2003). *Encyclopedia of biopharmaceutical statistics* (2nd ed.). New York: Marcel Dekker.

Cordes, D., Parker, J. K., Hopenwasser, A., Laurie, C., Izatt, J. R., & Nikles, D. (1995). *Teaming in technical courses.* Retrieved from the Frontiers in Education Web site, March 10, 2008, from http://fie.engrng.pitt.edu/fie95/4c3/4c34/4c34.htm

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity of psychological testing. *Psychological Bulletin, 52,* 281–302.

Driskell, J. E., Goodwin, G. F., Salas, E., & O'Shea, P. G. (2006). What makes a good team player? Personality and team effectiveness. *Group Dynamics: Theory, Research, and Practice, 10,* 249–271.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37,* 504–528.

Guzzo, R. A., & Dickson, M. W. (1996). Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology, 47,* 307–338.

Hackman, J. R. (1987). The design of work teams. In J. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315–342). New York: Prentice Hall.

Hoegl, M., & Gemuenden, H. G. (2001). Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. *Organization Science, 12,* 435–449.

Janz, B. D., Colquitt, J. A., & Noe, R. A. (1997). Knowledge worker team effectiveness: The role of autonomy, interdependence, team development, and contextual support variables. *Personnel Psychology, 50,* 877–904.

Kichuk, S. L., & Wiesner, W. H. (1998). Work teams: Selecting members for optimal performance. *Canadian Psychology, 39,* 1–2.

Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest, 7,* 77–124.

Kyllonen, P. C., & Lee, S. (2005). Assessing problem solving in context. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 11–25). Thousand Oaks, CA: Sage.

Liu, O. L., Rijmen, F., & Kong, N. (2007). *An initial investigation of a modified procedure for parallel analysis* (ETS Research Rep. No. RR-07-41). Princeton, NJ: ETS.

Loughry, M. L., Ohland, M. W., & Moore, D. D. (2007). Development of a theory-based assessment of team member effectiveness. *Educational and Psychological Measurement, 67*, 505–524.

Magazine, S. L., Williams, L. J., & Williams, M. L. (1996). A confirmatory factor analysis examination of reverse coding effects in Meyer and Allen's Affective and Continuance Commitment Scales. *Educational and Psychological Measurement, 56*, 241–250.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review, 26*, 356–376.

Maydeu-Olivares, A., Hernández, A., & McDonald, R. P. (2006). A multidimensional ideal point item response theory model for binary data. *Multivariate Behavioral Research, 41*, 445–471.

McCutcheon, A. L. (1987). *Latent class analysis*. Thousand Oaks, CA: Sage.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*, 730–740.

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*, 103–113.

Morgeson, F. P., Reider, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology, 58*, 583–611.

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (4th ed.). Los Angeles, CA: Author.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35*, 250–256.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers, 32*, 396–402.

O'Neil, H. F., Jr., Chung, G. K. W. K., & Brown, R. S. (1997). Use of network simulations as a context to measure team competencies. In H. F. O'Neil Jr. (Ed.), *Workforce readiness: competencies and assessment.* Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr., Wang, S., Jr., & Lee, C. (2003). Assessment of teamwork skills via a teamwork questionnaire. In H. F. O'Neil Jr. & R. S. Perez (Eds.), *Technology*

applications in education: A learning view (pp. 283–303). Mahwah, NJ: Lawrence Erlbaum Associates.

Rodebaugh, T. L., Woods, C. M., & Heimberg, R. G. (2006). The factor structure and screening utility of the Social Interaction Anxiety Scale. *Psychological Assessment, 18,* 231–237.

Schriesheim, C. A., & Eisenbach, R. J. (1995). An exploratory and confirmatory factor-analytic investigation of item wording effects on the obtained factor structures of survey questionnaire measures. *Journal of Management, 21*, 1177–1193.

Schulze, R., & Roberts, R. D. (2006). Assessing the Big-Five: Development and validation of the Openness Conscientiousness Extraversion Agreeableness Neuroticism Index Condensed (OCEANIC). *Zeitschrift für Psychologie, 214,* 133-149.

Stevens, M., & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management, 25*, 207–228.

Thorndike, E. L. (1920). A constant error on psychological rating. *Journal of Applied Psychology, 4*, 25–29.

Weekly, J. A., & Ployhart, R. E. (Eds.). (2005). *Situational judgment tests: Theory, measurement, and application*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Youngstrom, E. A., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male youth behavior ratings. *Journal of Consulting and Clinical Psychology*, *68*, 1–13.

Notes

[1] While the phrase *discriminant validity* rather than *divergent validity* was used in the seminal articles by Cronbach and Meehl (1955) and Campbell and Fiske (1959), discriminant validity has a different meaning in biomedical statistics (where a test has discriminant validity if it can discriminate between groups: e.g., the Zung depression scale should discriminate between people with and without a diagnosis of clinical depression; see Chow, 2003, p. 483). In order to avoid confusion with the biomedical definition of discriminant validity, psychologists in clinical or personality psychology often use divergent validity instead of discriminant validity to refer to a test having low observed correlations with conceptually unrelated constructs (particularly when studies examine both types of validity evidence in their research, e.g., Monga et al., 2000). For this reason, we prefer divergent validity to discriminant validity in this instance.

[2] The phenomenon where reverse-keyed items do not load on the same factors as nonreverse-keyed items (and hence are removed from the item pool in reliability or structural analyses) is not isolated to the current dataset but is a measurement issue that has been commented on frequently in literature. It is as yet unresolved how to concurrently deal with this issue and also control for acquiescence or other response sets. Thus, Barnette (2000) states: "The controversy with regard to using reverse or negatively worded survey stems has been around for several decades; it is a practice of questionable utility intended to guard against acquiescence or response set behaviors" (p. 361; see also, e.g., Magazine, Williams, & Williams, 1996; Maydeu-Olivares, Hernández, & McDonald, 2006; Rodebaugh, Woods, & Heimberg, 2006; Schriesheim & Eisenbach, 1995).

[3] An earlier reviewer of this article wondered "whether more items could be retained if the three-factor solution is tried on the whole item set." Initially, we tried one to six factor solutions on the whole item set. None of the solutions made substantive sense in the beginning. So we decided to use the four-factor solution based on the results from the scree plot and the parallel analysis and start exploring the problematic items from the four-factor solution. The issue of why so many items were deleted relates strongly to the reverse-keying issue, mentioned in a previous footnote. We began with 14 reverse-keyed items, discovered that these did not work, and removed virtually all of them.

Appendix A

ETS's Proposal for Measuring Employability Skills Pilot

Submitted to Ford in November 2005

It is well recognized that workers need a set of knowledge, skills, and abilities to be successful. Growing attention is being given to a set of "soft" skills or employability skills that workers also need to be effective. These skills include making ethical decisions, teamwork, motivation, advocating and influencing others, resolving conflict and negotiation, guiding others, and cultural and global awareness.

The Research Division of the Educational Testing Service (ETS) has a New Constructs group that concentrates on the study of attitudes, beliefs, and dispositions known as the soft skills. The group consists of local, national, and international researchers who are experts in this area of study, and are widely recognized by the field. Their work includes completing comprehensive reviews of the research literature, designing instruments to measure soft skills, creating innovative score and feedback reports, primary and secondary research about the constructs and their relationships to other measures, and disseminating their findings.

The Higher Education Division at ETS has led the development of groundbreaking projects including the Information and Communication Technology (ICT) Literacy Assessment and English Language Learning tools for teachers and students. The latest efforts include projects focused at transitions from high school to college and transitions to the workforce.

ETS proposes to work with the Ford Partnership for Advanced Studies (PAS) staff and selected high schools to conduct a pilot of instruments designed to measure a subset of the employability skills addressed in the Ford PAS curriculum. The constructs to be measured match the core employability skills classified as *Make Ethical Decisions*, *Cooperate with Others*, *Advocate and Influence*, *Resolve Conflict and Negotiate*, and *Guide Others*.

The proposed pilot includes two components: a student self-assessment survey and a teacher ratings survey. The student self-assessment survey requires students to complete self-ratings of

their perceptions of their skills in the five constructs and provide relevant biographical information. The teacher rating survey requires teachers to rate their participating students on the five constructs.

The proposed pilot will include the following steps:

1) ETS researchers will draft items for the student self-assessment survey and the teacher rating survey based on their previous research and a review of current literature.

2) ETS will work with Ford PAS staff and selected high school teachers to review the items as verification of criteria for each construct and to determine the biographical data to be collected on the pilot version. ETS researchers will revise the surveys based on the discussions.

3) ETS will develop a pilot specification plan. Ford, with ETS's guidance, will recruit a total of 500 students (Juniors or Seniors) who have completed at least two PAS courses in four to six sites. ETS will collaborate with Ford staff to create the necessary communication pieces for the teachers and students.

4) ETS will assemble and mount a pilot version of the student self-assessment survey and the teacher rating survey on an online delivery platform. The student survey will consist of biographical data, items to measure the 5 core skills, and an established set of questions measuring personality traits. There will also be a few evaluation questions about their pilot experience. ETS cannot provide scores or feedback on the 5 new construct measures, but can include the measure of personality traits and provide individual, confidential reports. Both surveys will be delivered online and be able to be completed in one class period (50 minutes maximum).

5) The pilot version will be completed by students and teachers in the late spring. Students, if they elect to, can receive information about their personality traits. These reports will be sent via email directly to the students.

6) ETS researchers will complete the data analysis of the surveys and create a technical report of findings to be shared with the Ford staff.

This project should take about 6 months to complete, starting January 1, 2006. The proposed timeline for this pilot project is as follows:

Activity	**Timeframe**
Draft items	January - early February
Review items	Mid - late February
Pilot plan and recruitment	February
Finalize items and prepare delivery	March
Pilot the surveys	April
Analyses and report writing	May and June

ETS will contribute the researchers' time, but is requesting a contribution from Ford PAS for $20,000 to cover the costs of assembling, mounting and delivering the pilot online and other administration tasks specific to this pilot.

Appendix B

Student Survey

Leadership and Teamwork Scale --- Student Survey

Student ID: _____

Which Ford PAS module (s) have you taken? (Circle the modules you have studied)

	Course 1	Course 2	Course 3	Course 4	Course 5	
Modules	**1 - 2 - 3**	**4 - 5 - 6**	**7 - 8 - 9**	**10 - 11 - 12**	**13 – 14 - 15**	**Unsure**

Which Ford PAS module (s) are your currently in the process of taking? (Circle the modules you are presently studying)

	Course 1	Course 2	Course 3	Course 4	Course 5	
Modules	**1 - 2 - 3**	**4 - 5 - 6**	**7 - 8 - 9**	**10 - 11 - 12**	**13 – 14 - 15**	**Unsure**

Use No. 2 Pencil Only Right Mark ● Wrong Marks ⊙ ⊗ ⊘ ☑

SECTION 1: BACKGROUND INFORMATION

DIRECTIONS:

The purpose of this section is to be able to describe the geographic, economic, and ethnic backgrounds of the individuals who have participated in this pilot study.

	Male	Female
1. What is your gender?	○	○

	African American	American Indian or Alaskan Native	Asian and Pacific Islander	White, not Hispanic	Hispanic	Other
2. What is your race/ethnicity?	○	○	○	○	○	○

	14	15	16	17	18	19	20
3. How old are you?	○	○	○	○	○	○	○

	English	English and another Language	Another language only
4. What is the primary language spoken in the home?	○	○	○

4b. If you selected another language, which language(s) do you speak?

_____

	Yes	No
5. Do you participate in any volunteer activities?	○	○

SECTION 2: EDUCATIONAL BACKGROUND INFORMATION

DIRECTIONS:
The purpose of this section is to obtain information about your educational background.

1. What grades did you receive in the following subjects on your most recent report card?	A	B	C	D	F
Reading/Language Arts	○	○	○	○	○
Math	○	○	○	○	○
Science	○	○	○	○	○
Social Studies	○	○	○	○	○
Art	○	○	○	○	○
Music	○	○	○	○	○
Other (describe below)	○	○	○	○	○

2. List the course described as "Other" above

_____

SECTION 3: PERSONAL CHARACTERISTICS

DIRECTIONS:
In this section, we have listed a number of characteristics that may or may not apply to *YOU*. Fill in the corresponding answer for each statement to indicate the extent to which you agree or disagree with the statement. Please rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

I see myself as:	Disagree Strongly	Disagree Moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree Moderately	Agree Strongly
1. Extraverted, enthusiastic	○	○	○	○	○	○	○
2. Critical, quarrelsome	○	○	○	○	○	○	○
3. Dependable, self-disciplined	○	○	○	○	○	○	○
4. Anxious, easily upset	○	○	○	○	○	○	○
5. Open to new experiences, complex	○	○	○	○	○	○	○
6. Reserved, quiet	○	○	○	○	○	○	○
7. Sympathetic, warm	○	○	○	○	○	○	○

8. Disorganized, careless	○	○	○	○	○	○	○
9. Calm, emotionally stable	○	○	○	○	○	○	○
10. Conventional, uncreative	○	○	○	○	○	○	○

*SECTION 4: WHAT ARE YOUR VIEWS ABOUT **ABILITY** AND **EFFORT?***

DIRECTIONS:
In this section, we have listed several ideas about ability and effort. Fill in the corresponding answer to indicate the extent to which you agree or disagree with the statement.

What is *your* opinion of the following:	Strongly Disagree	Disagree	Agree	Strongly Agree
1. You have a certain amount of intelligence and no matter how hard you study nothing will change it.	○	○	○	○
2. If a student gets a bad grade in a subject, it's because he/she isn't very good in that subject.	○	○	○	○
3. If students get bad grades in a subject, it's because they didn't try hard enough.	○	○	○	○
4. Intelligence changes throughout your life.	○	○	○	○
5. Every student could do really well if he/she tried hard enough.	○	○	○	○
6. Some people can only learn so much, no matter how hard they try.	○	○	○	○

SECTION 5: WHAT ARE YOUR VIEWS ABOUT **TEAMWORK** AND **LEADERSHIP**?

DIRECTIONS:
Below, you will find a number of statements that you should read and decide how well each one of them describes the way you think or feel. Here is an example:

	Never	Rarely	Sometimes	Often	Usually	Always
I love participating in class discussion.	○	○	○	○	○	○

You might respond:

	Never	Rarely	Sometimes	Often	Usually	Always
I love participating in class discussion.	○	○	○	●	○	○

You should reply to all statements. Give your first impression of whether each statement describes the way you think and feel about it. Don't spend too long on deciding what your answer should be. Answer all statements even if you're not entirely sure of your answer. There are NO right or wrong answers.

	Never	Rarely	Sometimes	Often	Usually	Always
1. I act without consulting my peers.	○	○	○	○	○	○
2. I prefer for other students to take responsibility.	○	○	○	○	○	○
3. I respect the opinions of my peers.	○	○	○	○	○	○
4. I reserve my opinions until all the facts are at hand.	○	○	○	○	○	○
5. I know how to make other students see things my way.	○	○	○	○	○	○
6. I take other students' interests into account.	○	○	○	○	○	○
7. My arguments are constructive.	○	○	○	○	○	○
8. I lose my temper.	○	○	○	○	○	○
9. I seek to influence my peers.	○	○	○	○	○	○
10. I give in when arguing.	○	○	○	○	○	○

	Never	Rarely	Sometimes	Often	Usually	Always
11. I understand that each team member is different.	○	○	○	○	○	○
12. I am flexible in team situations.	○	○	○	○	○	○
13. I find it difficult keeping team members on task.	○	○	○	○	○	○
14. I am a good listener.	○	○	○	○	○	○
15. I am open to varying opinions.	○	○	○	○	○	○
16. I carefully consider the facts to persuade my peers.	○	○	○	○	○	○
17. I value competition over cooperation.	○	○	○	○	○	○
18. I dislike being in a position of responsibility for others.	○	○	○	○	○	○
19. I adapt to change well.	○	○	○	○	○	○
20. I like to be in charge of groups or projects.	○	○	○	○	○	○
21. I like to have my opinions heard.	○	○	○	○	○	○
22. I acknowledge the accomplishments of my peers.	○	○	○	○	○	○
23. I enjoy helping team members.	○	○	○	○	○	○
24. I put down other students' ideas or suggestions.	○	○	○	○	○	○
25. I prefer working alone over teamwork.	○	○	○	○	○	○
26. I cooperate with other students.	○	○	○	○	○	○
27. I can make bargains in any situation.	○	○	○	○	○	○
28. I believe that there is only one “best” solution.	○	○	○	○	○	○
29. I am comfortable providing constructive criticism.	○	○	○	○	○	○

	Never	Rarely	Sometimes	Often	Usually	Always
30. I dislike it when people challenge my views.	○	○	○	○	○	○
31. I am not interested in classmate's project issues.	○	○	○	○	○	○
32. I change my opinions.	○	○	○	○	○	○
33. Feedback is important to me.	○	○	○	○	○	○
34. I like to solve problems using different tactics.	○	○	○	○	○	○
35. During group assignments, I make demands on other students.	○	○	○	○	○	○
36. I am good at taking input from other students.	○	○	○	○	○	○
37. I like team activities.	○	○	○	○	○	○
38. I act as a mediator (i.e., a go-between).	○	○	○	○	○	○
39. I suggest alternative solutions to problems.	○	○	○	○	○	○
40. I find it difficult to approach classmates.	○	○	○	○	○	○
41. I contribute to the definition of a team's goals.	○	○	○	○	○	○
42. I enjoy bringing team members together.	○	○	○	○	○	○
43. I try to understand why others disagree.	○	○	○	○	○	○
44. I can sense if people are "conflicted."	○	○	○	○	○	○
45. I consider team members first.	○	○	○	○	○	○
46. I share ideas with others to accomplish a task.	○	○	○	○	○	○
47. I believe I am a good leader.	○	○	○	○	○	○
48. I can convince my peers about anything.	○	○	○	○	○	○
49. I can fight for a cause that I believe in.	○	○	○	○	○	○

	Never	Rarely	Sometimes	Often	Usually	Always
50. I know when to step in when a disagreement starts getting out of control.	○	○	○	○	○	○
51. I am influenced by other students' opinions.	○	○	○	○	○	○
52. I provide appropriate feedback to team members.	○	○	○	○	○	○
53. I value different perspectives to help me strengthen my understanding of issues or problems.	○	○	○	○	○	○
54. I learn from other students.	○	○	○	○	○	○
55. I think that exchange of ideas among team members can lead to creative solutions.	○	○	○	○	○	○
56. I am inspired by others' ideas and thoughts.	○	○	○	○	○	○
57. My ideas are enhanced in a team situation.	○	○	○	○	○	○

Subjective Judgment Test

DIRECTIONS

In this section, you will find a number of paragraphs describing various situations. You will be asked to evaluate how effective you think several reactions to the situation are. Please read each scenario carefully before choosing your response. Don't spend too long deciding on each answer. Please answer all of the statements even if you're not entirely sure of your answer. There are no right or wrong answers.

SCENARIO 1:

You are part of a study group that has been assigned a large presentation for class. As you are all dividing up the workload, it becomes clear that both you and another member of the group are interested in researching the same aspect of the topic. Your friend already has a great deal of experience in this area, but you have been extremely excited about working on this part of the assignment ever since your teacher first mentioned it.

Rate the following courses of action as to how effective you think these are:	Very Ineffective	Ineffective	Neither ineffective nor effective	Effective	Very Effective
A. Flip a coin to determine who gets to work on that particular aspect of the assignment.	○	○	○	○	○
B. Insist that, for the good of the group, you should work on that aspect of the assignment because your interest in the area means you will do a particularly good job.	○	○	○	○	○
C. Compromise your preferences for the good of the group and allow your friend to work on that aspect of the assignment.	○	○	○	○	○
D. Suggest to the other group member that you both share the research for that aspect of the assignment, and also share the research on another less desirable topic.	○	○	○	○	○

SCENARIO 2:

You are on the Homecoming Committee at school, helping to plan all of the activities for the upcoming Homecoming Week. The committee members are having a problem trying to decide on a theme for the big Homecoming Dance. Everybody keeps objecting to the various suggestions because they have been used for so many dances that everyone thinks they are boring. You have a new idea, a unique approach, but you think that some of the more popular kids on the committee might not like it.

Rate the following courses of action as to how effective you think these are:	Very Ineffective	Ineffective	Neither ineffective nor effective	Effective	Very Effective
A. Tell a couple of the group members about your idea privately, and see if they will jointly present the idea to the rest of the group.	○	○	○	○	○
B. Keep your idea to yourself because you would rather see the group as a whole come up with a silly theme than have them laugh at your suggestion.	○	○	○	○	○
C. Throw out your idea as a joke and hope that someone picks up on it and turns it into a real suggestion.	○	○	○	○	○
D. Present your idea in spite of potential backlash because it is more important to solve the problem than to protect yourself.	○	○	○	○	○

SCENARIO 3:

You are in a large class. After breaking out into small groups to discuss a scenario, your class reconvenes to go over the ideas generated in each small group. When it's your group's turn, the person who volunteered to serve as spokesperson begins to present his or her own ideas, some of which had not been discussed in the small group, rather than the ideas generated in the group discussion.

Rate the following courses of action as to how effective you think these are:	Very Ineffective	Ineffective	Neither ineffective nor effective	Effective	Very Effective
A. Do nothing, in order to preserve the impression that your group worked well together.	○	○	○	○	○

B. Challenge the spokesperson, because it is important that only the ideas agreed upon by your group be presented.	○	○	○	○	○
C. Keep silent during class, but afterwards speak with the instructor privately about the spokesperson's behavior.	○	○	○	○	○
D. Politely add additional comments, making sure that the ideas the group discussed are presented.	○	○	○	○	○

SCENARIO 4:

You are the president of your school's drama club. You are starting to plan the big spring musical, and you are meeting with the other members of the club to decide who will take on the various jobs (building sets, painting, getting costumes, serving as stage manager, etc.) required for the production.

Rate the following courses of action as to how effective you think these are:	Very Ineffective	Ineffective	Neither ineffective nor effective	Effective	Very Effective
A. Ask each club member which job they would like to take on, and then resolve any conflicts based on who asked first.	○	○	○	○	○
B. Ask each club member which job they would like to take on, and then resolve any conflicts by drawing names out of a hat.	○	○	○	○	○
C. Make a list of the club members' strengths and preferences and assign jobs that match their strengths most closely.	○	○	○	○	○
D. Assign the best jobs to your friends first, and then give the remaining jobs to the rest of the club members randomly.	○	○	○	○	○

SCENARIO 5:

You have recently formed a study group with several of your classmates in order to prepare for a difficult final exam. Unfortunately, the various members of the group have very different schedules, so you all meet right after school one day to try to work out a final schedule for your group review sessions.

Rate the following courses of action as to how effective you think these are:	Very Ineffective	Ineffective	Neither ineffective nor effective	Effective	Very Effective
A. Making sure that the schedule will allow the smartest students to attend, so that the study group will cover more material.	○	○	○	○	○
B. Making sure the proposed meeting times do not conflict with the extracurricular activities you are involved in.	○	○	○	○	○
C. Yielding to the majority of the group even if it means some members will not be able to participate.	○	○	○	○	○
D. Trying to find sub-groups with compatible schedules so that you can create new groups everyone can take part in.	○	○	○	○	○

SCENARIO 6:

You are a member of a club that is sponsoring a job fair at your school tomorrow afternoon. Your club has been planning this event for weeks and you are scheduled to work at the event, but your boss from your after-school job has just called and insists that because several other employees have become ill, you must come work a shift tomorrow during the hours that the job fair is scheduled.

Rate the following courses of action as to how effective you think these are:	Very Ineffective	Ineffective	Neither ineffective nor effective	Effective	Very Effective
A. Say that you cannot work the shift, explaining to your boss that you haven't been feeling well yourself.	○	○	○	○	○
B. Say that you cannot work the shift, explaining to your boss that you simply cannot break a prior commitment to work at the job fair.	○	○	○	○	○
C. Agree to work the shift, and leave a message for the job fair's organizer explaining that your paid job is more important and you won't be able to attend.	○	○	○	○	○
D. Agree to work the shift, but contact the job fair's organizer and offer to arrive early to set up or come late to clean up afterwards.	○	○	○	○	○

SCENARIO 7:

You are taking part in a study group with a number of your classmates in preparation for a particularly difficult upcoming exam. As the first review session gets underway, it becomes clear that the other members of the group have not taken good notes and are not as familiar with the material as you are.

Rate the following courses of action as to how effective you think these are:	Very Ineffective	Ineffective	Neither ineffective nor effective	Effective	Very Effective
A. Suggest that everyone read over the textbook in preparation for the next review session.	○	○	○	○	○
B. Leave the group because you will be better off studying on your own.	○	○	○	○	○
C. Offer to use your notes as the basis for the remaining review sessions.	○	○	○	○	○
D. Ask the teacher to postpone the exam because many of the students are clearly not ready.	○	○	○	○	○

SCENARIO 8:

You have been assigned to train the new members of the yearbook staff in how to use software to crop and lay out photos. During the first day of training, it becomes apparent that while some of the new staff members have previous experience in this area and are grasping the material quickly, a few are really struggling.

Rate the following courses of action as to how effective you think these are:	Very Ineffective	Ineffective	Neither ineffective nor effective	Effective	Very Effective
A. Adjust your training to focus on the needs of the struggling students, since it is important that everyone master the material.	○	○	○	○	○
B. Proceed with the training as planned, in order to keep the best students from becoming bored.	○	○	○	○	○
C. Offer to mentor the struggling students individually after school.	○	○	○	○	○
D. Pair the struggling students with more experienced counterparts who can help mentor them during the training.	○	○	○	○	○

Appendix C

Teacher Questionnaire

Teamwork Scale --- Teacher Questionnaire

Student's ID: _____

Student's GRADE (please circle): 9 10 11 12

Background Questions

1. How many months or years have you been teaching Ford PAS? _____ years _____months

2. Which of the following Ford PAS modules have you taught: (Circle all the modules you have taught)

	Course 1	Course 2	Course 3	Course 4	Course 5
Modules	1 - 2 - 3	4 - 5 - 6	7 - 8 - 9	10 - 11 - 12	13 - 14 - 15

3. What type of Ford PAS teacher training have you participated in? (Check off all relevant options)

_____ Annual Conference _____1-2 day regional training _____ Weeklong summer institute _____School-based training _____None

5. Are you comfortable implementing a student-centered/teamwork approach to teaching and learning in your classroom? (circle the word that most closely describes how comfortable you are)

Never Rarely Sometimes Often Usually Always

6. Which Ford PAS module(s) has this student taken? (Circle the modules you know the student has taken)

	Course 1	Course 2	Course 3	Course 4	Course 5	
Modules	1 - 2 - 3	4 - 5 - 6	7 - 8 - 9	10 - 11 - 12	13 - 14 - 15	Not sure

Evaluation of Students

We would like to know your evaluations of **each student** in a number of areas
Please read each question carefully and answer all questions as honestly as possible.
Remember to fill out a separate survey for **each student.**

Use No. 2 Pencil Only Right Mark ● Wrong Marks ⊙ ⊗ ⊘ ☑

Part 1

This student:	Below Average	Average	Above Average	Outstanding	Truly Exceptional
1. Writes clearly and effectively	①	②	③	④	⑤
2. Speaks clearly and effectively	①	②	③	④	⑤
3. Thinks critically and analytically	①	②	③	④	⑤
4. Learns effectively on their own.	①	②	③	④	⑤

5. Organizes work and time effectively.	①	②	③	④	⑤
6. Works effectively with others.	①	②	③	④	⑤
7. Co-operates well with others	①	②	③	④	⑤
This student:	**Below Average**	**Average**	**Above Average**	**Outstanding**	**Truly Exceptional**
8. Leads others effectively	①	②	③	④	⑤
9. Resolves conflicts effectively	①	②	③	④	⑤
10. Teaches others effectively	①	②	③	④	⑤
11. Understands himself/herself	①	②	③	④	⑤
12. Understands people of other cultural backgrounds	①	②	③	④	⑤
13. Contributes to the welfare of the community.	①	②	③	④	⑤
14. Maintains high ethical standards.	①	②	③	④	⑤

Part 2

(Q1) When facing adverse or difficult circumstances this student:

① puts little effort into improving things, and may either avoid the problem or lose control.	②	③ faces the problem and takes some steps to change things, but may appear stressed.	④	⑤ takes appropriate actions to change the situation, and remains in control of himself/herself.

(Q2) In their relationships and interactions with classmates, this student:

① irritates or angers others and is consequently not well-liked.	②	③ gets along reasonably well with most other students, but can make occasional social gaffes.	④	⑤ deals well with almost everyone, and is consequently very well-liked

(Q3) When working on a group goal or project, this student:

① ignores or does not notice others' ideas or suggestions.	②	③ listens to others' contributions.	④	⑤ always listens to others and respects their contributions.

(Q4) When working on a group goal or project, this student:

① acts without regard to others' interests or suggestions.	②	③ attempts to adjust his/her actions to achieve the group goal.	④	⑤ takes the relevant and appropriate actions to achieve the group goal.

(Q5) When planning a group task, this student:

①	②	③	④	⑤
has no clear plan, or a plan that does not involve the rest of the group.		creates adequate plans that usually involve others' interests and/or resources.		creates clear and careful plans that take others' interests and resources into account.

(Q6) When planning a group task, this student:

①	②	③	④	⑤
ignores or does not notice the needs of other group members.		revises plans if other group members suggest the need.		carefully revises plans so they are fair and relevant to all group members.

(Q7) When there is conflict or disagreement within a group, this student:

①	②	③	④	⑤
ignores or does not notice that conflict occurs.		notices the conflict and tries to work out why it is occurring.		clearly acknowledges the conflict and correctly analyzes why it has occurred.

(Q8) When there is conflict or disagreement within a group, this student:

①	②	③	④	⑤
ignores or actively worsens the conflict.		creates a strategy to resolve the conflict, but may not execute or revise this strategy successfully.		creates strategies to resolve the conflict, and executes and revises these strategies successfully.

(Q9) When other students need help, this student:

①	②	③	④	⑤
does not notice that they need help and/or does not provide help.		tries to help, but may help inappropriately, or be seen as condescending.		accurately assesses others' needs and skills to provide appropriate learning opportunities without condescension.

(Q10) When helping other students, this student:

①	②	③	④	⑤
provides little useful help.		notes how the others are doing, and gives accurate feedback, but may not change helping strategies if others are not doing well.		changes feedback according to how others are doing and feeling, and also revises her/his helping strategies in line with this.