

Threats of common method variance in student assessment of instruction instruments

John Garger
The Science Survival Academy, Binghamton, New York, USA
Paul H. Jacques
School of Business, Rhode Island College, Providence, Rhode Island, USA
Brian W. Gastle
*English Department, Western Carolina University,
Cullowhee, North Carolina, USA, and*
Christine M. Connolly
*Management and Marketing Department, Rhode Island College,
Providence, Rhode Island, USA*

Abstract

Purpose – The purpose of this paper is to demonstrate that common method variance, specifically single-source bias, threatens the validity of a university-created student assessment of instructor instrument, suggesting that decisions made from these assessments are inherently flawed or skewed. Single-source bias leads to generalizations about assessments that might influence the ability of raters to separate multiple behaviors of an instructor.

Design/methodology/approach – Exploratory factor analysis, nested confirmatory factor analysis and within-and-between analysis are used to assess a university-developed, proprietary student assessment of instructor instrument to determine whether a hypothesized factor structure is identifiable. The instrument was developed over a three-year period by a university-mandated committee.

Findings – Findings suggest that common method variance, specifically single-source bias, resulted in the inability to identify hypothesized constructs statistically. Additional information is needed to identify valid instruments and an effective collection method for assessment.

Practical implications – Institutions are not guaranteed valid or useful instruments even if they invest significant time and resources to produce one. Without accurate instrumentation, there is insufficient information to assess constructs for teaching excellence. More valid measurement criteria can result from using multiple methods, altering collection times and educating students to distinguish multiple traits and behaviors of individual instructors more accurately.

Originality/value – This paper documents the three-year development of a university-wide student assessment of instructor instrument and carries development through to examining the psychometric properties and appropriateness of using this instrument to evaluate instructors.

Keywords Common method variance, Instructor assessment, Single-source bias

Paper type Research paper

Introduction

Developing a method of assessing teaching that is valid and considers both disparate disciplines and various instructional methods is vexing to colleges and universities. Many institutions rely on student responses gathered by student assessment of instruction (SAI)



instruments to perform summative evaluations of instruction. However, teaching faculty members are often skeptical of the validity of data derived from such instruments and of the purposes to which the data are applied. This skepticism derives partially from lack of a common definition of excellence in teaching at institutions and general distrust of single-factor or numerical assessments of an activity as diverse as teaching. When assessment data are inaccurate while assessing teaching excellence, it influences an institution's ability to measure instructors' effectiveness as the institution defines it. Accurate data from students are essential to providing accurate feedback regarding teaching excellence to instructors. The method of application of data collected is also paramount.

A common response to this concern is to use extensive cafeteria-style SAI instruments that allow institutions to extract items believed to be relevant or appropriate to those institutions. It is believed that such diverse instruments yield data related to several facets of teaching, allowing institutions to focus on dimensions related to institutional definitions of teaching. Several prepackaged, cafeteria-style instruments exist (Richardson, 2005), but institutions and instructors often underuse data from such instruments, and consequently, institutions attempt to develop instruments that meet their own needs, which can lead to increased faculty investment in the instrument's utility. Developing an institutionally specific SAI instrument risks replication of validity problems discussed commonly in research on such instruments and offers no guarantee of either faculty buy-in or appropriate application of data collected. This study examines development of an SAI instrument at a university, highlighting the process and attendant difficulties related to developing an institution-specific SAI instrument.

Scholarship on SAI instruments is as prolific and varied as problems associated with such instruments, so much so that surveys of research itself have become common. McAlpin *et al.* (2014), for example, summarized SAI instrument research since the 1920s, and Wachtel (1998) analyzed arguments both for and against using student feedback for summative purposes. Hornstein (2017), Boring *et al.* (2016) and Spooren *et al.* (2013) challenged the appropriateness of these assessment tools when making decisions about tenure and promotion, questioning their ability to measure teaching competence. Most literature on instrument development and deployment focuses on the composition and related validity of an instrument. As institutions develop assessment methods, instrument items have undergone scrutiny as they relate to a number of variables, generally categorized as administrative, course, instructor and student aspects. A focus on implementation of an instrument itself, and how transferable the validity of the tool is to student understanding of what it intends to measure, has been minimal. Regardless of all issues pertaining to variability in SAIs, the overriding issue is the need for a valid instrument and collection method; without reliable and valid instrumentation, contextual issues cannot be assessed. This paper identifies a threat to obtaining useful information from the most commonly used instrument and collection method of SAIs. The SAI instrument, typically collected near the end of an institution's term, suffers from several limitations. At the forefront of these limitations are psychological limitations associated with a student assessing multiple attributes and behaviors of an instructor.

Common method variance

Thorndike (1920) discussed that in a study of corporate employees, multiple attributes (e.g. intelligence, skill and reliability) correlated highly in the same individual, concluding:

Those giving the ratings were unable to analyze out these different aspects of the person's nature and rate each [attribute] in independence of the others. Their ratings were apparently affected by a marked tendency to think of a person in general as rather good or rather inferior and to color the judgments of the qualities by this general feeling. (p. 25)

Thorndike later called this "the constant error of the 'halo'" (p. 28), which came to be known as the halo perceptual error. Common method variance is defined as overlapping variability that

is due to a collection method rather than from true relationships among constructs (Campbell and Fiske, 1959). Whereas overlap between the actual construct and measurement of that construct represents covariation supporting validity, covariation between two or more constructs due to the method used to measure the constructs represents common method variance (Podsakoff *et al.*, 2003). Although covariance can be both random and systematic (Bagozzi and Yi, 1991; Nunnally, 1978), covariance among measured constructs collected from one source is called single-source bias, representing a special case of common method variance. Single-source bias is often overlooked, and if not recognized, can confound findings (Baugh *et al.*, 2006). Students filling out a survey consisting of perceptions of multiple behaviors or attributes of instructors fall into this category.

At the core of single-source bias is the halo perceptual error, or within the context of this study, the inability of raters to distinguish multiple behaviors of an individual (Feeley, 2002; Spector, 1987; Taut *et al.*, 2018). Raters use cognitive structures or schemata to generalize evaluations (Dipboye and Flanagan, 1979; Mitchell, 1985), resulting in more efficient cognitive processing but less accurate perceptions. In the case of SAIs through a survey, it is difficult to measure multiple behaviors of instructors that go beyond generalized effect of an instructor's aggregate behaviors, attributes and traits. If students are unable to differentiate behaviors in which instructors engage, perceptions of teaching effectiveness are then evaluated generally, preventing differentiation of attributes that would otherwise provide greater accuracy in assessment data.

SAIs are used to identify both instructor and administration areas in which an instructor meets or fails to meet standards deemed important by the authors of an assessment instrument. With the effects of single-source bias, multiple constructs measured with an assessment instrument are not necessarily identifiable statistically, and without testing, are assumed to exist. This assumption has consequences in which significant covariation among constructs due to single-source bias indicates a situation for the instructor that would have been different under other methods of collection. This study empirically demonstrates an unidentifiable hypothesized factor structure in a proprietary SAI instrument and discusses the ramifications of unidentifiable factor structures in such measurements.

Development of a student assessment of instructor instrument

Every institution has its own method for developing and adopting an SAI instrument, collection of data and use of data collected, and it would be difficult to agree on what constitutes a typical process for developing or adopting such instruments. What follows is a description of the methods used to develop an SAI instrument at a mid-sized, regional, comprehensive, public university in the USA that approximates typical instrument development and implementation. Prior to the study, the institution's governing body requested development of a common, summative SAI instrument. At the time, teaching assessment instruments varied widely across colleges and departments, and included simple Likert-type scale response forms, exclusively discursive forms and hybrids of numeric and discursive responses. Lack of uniformity made it difficult to make equitable personnel decisions across the various colleges, departments and disciplines. The institution's administrators desired a common instrument to assess instruction better and more uniformly, but faculty members were concerned that such an instrument might not consider varied delivery methods and disciplinary topics offered at the institution. Faculty reservations about such evaluation instruments appear to be common. Nasser and Fresko (2002) and Shelton and Hayne (2017) addressed faculty perceptions of summative evaluation instruments, suggesting that few instructors change their teaching methods based on such evaluations, and many are leery of the uses to which the data are applied. There is little evidence that the use of these evaluation instruments enhances teaching quality (Spooren and Mortelmans, 2006). Academic institutions offer a variety of majors and courses at many

levels using disparate pedagogies. Concerns about evaluation instruments and their ability to define teaching excellence are valid and should be addressed to allow assessments to be perceived as credible and useful to enhancing instructor performance.

The institution's faculty senate formed a teaching evaluation committee to research current assessment instruments, review current scholarship of teaching assessment and develop an instrument appropriate to the institution. Over the next three years, the committee collected, sorted and analyzed teaching evaluation instruments from all colleges in the institution, compiled and analyzed teaching evaluation instruments from peer institutions in the region and across the country, reviewed published research in the field of teaching evaluation, especially research devoted to the validity of the kinds of questions asked, organized on-campus workshops with experts in the field and held open-campus forums to allow input from faculty, administrators and students. Since quality teaching is a primary aspect of the institution's mission, administrators and faculty agreed that an appropriate SAI instrument must relate to the institution's definition of good teaching. The institution subscribes to seven dimensions of teaching derived from Arreola (1995), Centra and Diamond (1987), Darling-Hammond and Hyler (2013) and Zhang *et al.* (1997), which include content expertise, instructional delivery skills, instructional design skills, course management skills, evaluation of students, faculty/student relationships and facilitation of student learning.

The committee's initial move to define and systematize an institution-specific teaching paradigm proved important during subsequent deliberations on instrument development since faculty had established this paradigm prior to work on the SAI instrument, and since the institution had already adopted the paradigm. It also clarified and delineated later faculty members' objections to specific SAI instrument items. For example, during open forums, faculty reactions were mixed regarding the extent to which students were able to assess an instructor's content expertise in a field. Pedagogy acknowledges that educational practices and faculty selection of teaching materials are not neutral, influencing perceptions of student assessment of instructor effectiveness. Attempts were made to broaden learning constructs; materials presented are shaped by books, technology, classroom activities, history, politics, economics, language used by the instructor and other factors that empower and disempower students. These choices affect instructor assessment. Teaching excellence is a contested concept, and definitions shifted based on social, economic and political contexts (Skelton, 2004), increasing the complexity of instructor assessment and identification of tools that assess excellence.

Analysis of SAI instruments across colleges and from other institutions foregrounded various methods and purposes of SAI instruments. Data from SAI instruments are used to make personnel decisions and improve instruction, and many schools used variations of the same form for both purposes. The committee concluded that SAI instruments aimed at improving instruction appear to contain more items than those whose purpose is informing personnel decisions. As Seldin (1993) suggested, student ratings are generally statistically reliable and provide important input into the consideration of teaching effectiveness. However, without exception, the study emphasizes that only multiple sources of data – and no single source – provides sufficient information to make a valid judgment about overall teaching effectiveness. Algozzine *et al.* (2004) reported that 30 empirical studies “refute the use of a global or overall rating for the evaluation of teaching effectiveness: A single score cannot represent the multidimensionality of teaching” (p. 136). Teacher effectiveness should instead be measured in a variety of ways including but not limited to peer evaluations, student evaluations, portfolio development, administrative ratings, self-evaluations and achievement of outcome objectives (Kalra *et al.*, 2015). During open-campus forums at the institution, most faculty members agreed that SAI was valid and useful, but they also expressed concerns regarding how administrators would use the data collected. Given the variety of course subjects offered, delivery methods practiced and teaching styles

employed at the institution, faculty desired an instrument that would account for these variables or not be affected by them, and that relates to the specific approach to teaching the institution adopted. A single global score should not be used as the sole deciding factor when evaluating teaching effectiveness (Marsh and Roche, 1997; Neumann, 2000; Stark-Wroblewski *et al.*, 2007).

An initial instrument developed attempted to consider the faculty's desire for a robust instrument, current research on correlative factors in SAI instruments and administrators' demands for uniformity. The initial proposal described an instrument with three sections: a seven-item instrument that measured seven dimensions of teaching common across the institution; a brief discursive section that could also be used for formative input; and a section devoted to items that were department/discipline specific, a kind of cafeteria supplement for individual programs. The proposal also included a list of suggested items for Parts 2 and 3. The campus community expressed few concerns about the instrument, but the only significant opposition contended that the seven questions represented single-item responses. Based on response validity – a topic common in research (Heller, 1984) – the opposition was so strong that the instrument was sent back to the committee for revision. A subsequent iteration of the instrument, developed in conjunction with faculty versed in psychometrics, included multiple SAI instruments based on course delivery method and multiple items for each of the seven dimensions defined by the institution. The number of constructs captured was reduced from 7 to 5, representing the final version of the instrument used university-wide immediately following adjournment of the committee. Measured with 20 items, the constructs in the final version of the SAI instrument were organization, enthusiasm, rapport, feedback and learning, collected using a five-point Likert-type scale that ranged from strongly disagree to strongly agree, with values 2–4 representing intermediate responses. The constructs and their corresponding items appear in the Appendix. Although the creators of the instrument did not formally hypothesize five factors, since the five constructs were to be used to assess teaching at the university, the creators of the instrument intended for the constructs to represent distinct areas of teaching that could be scrutinized and examined. They wanted to be able to assess an instructor's organization, enthusiasm, etc., separately without simultaneously assessing other extraneous variables. Consequently, the creators of the SAI instrument hypothesized a five-factor target model in which the component constructs are distinguishable both in practice and from a statistical analysis viewpoint. Analyses in this paper test whether the five-factor model is justifiable against the null hypothesis that five factors are not statistically distinguishable in data collected.

Methods

Participants

In total, 357 students in 29 courses offered at the institution participated in the study. In all, 40 percent of the students were female, with an average of 12.3 respondents per course. Participants completed and returned the SAI instrument near the end of the course before final grading was assessed and distributed, the period during which the committee intended to collect such data. To assess the constructs in the instrument, it was necessary to assess the data with several methods so results did not suffer from the limitations of one analysis. Consequently, three analyses were conducted to triangulate findings. The three analyses conducted were exploratory factor analysis (EFA), nested confirmatory factor analysis (NCFA) and within-and-between analysis (WABA).

Exploratory factor analysis

EFA identifies previously un-hypothesized factors present in a large data set and allows verification of hypothesized latent constructs (Conway and Huffcut, 2003). Unlike confirmatory

or criterion analyses that test models and identify the strength of relationships among variables (Kline, 2010), EFAs describe data, allowing researchers to assess whether raw data, such as items on a survey, correlate meaningfully and as expected (Browne, 2001). An example is when a researcher uses an EFA to determine whether items on a survey load on factors they should and do not load on factors that they should not, the factor structure of the data (Schmitt, 2011). Analyses can be conducted under a variety of assumptions, including strict discrimination, as with the use of orthogonal rotations, or with relaxed assumptions, as with oblique rotations (Thompson, 2004). Results of an EFA include factor loadings that identify the clustering of variables on known or unknown factors. Use of this multivariate statistical method assists with identification of relationships that exist between variables.

Nested confirmatory factor analysis

NCFA is a method of using confirmatory factor analysis to test competing models. Unlike a strictly confirmatory method of testing model fit (Brown, 2015), NCFA examines a variety of assumptions about the convergent and discriminant validity of hypothesized, underlying factors by comparing the measurement of several rather than just one model. We recognize that covariation might exist between or among some hypothesized constructs. Consequently, an investigation into the items and the constructs measured by those items might suggest that larger, more general constructs are present in the data (Schermelleh-Engel *et al.*, 2003). An NCFA begins with the null hypothesis that only one factor is present in the data. Increasingly complex models are subsequently tested by drawing out more factors until the most complex model – in this case the hypothesized five-factor model – is tested. If a five-factor model is identifiable in the data, each increasingly complex model should provide better fit than previous, less complex models. It is not the intention to test all possible construct combinations, but testing competing models against a target model provides insights into interpretation of the constructs' factor structure (Asparouhov and Muthén, 2009). Antonakis *et al.* (2003) and Avolio *et al.* (2003) provided examples of this technique applied to instrument validation. A one-factor (i.e. null hypothesis) model was tested using structural equation modeling. A second, more complex two-factor model was then tested using rapport and enthusiasm to represent a relationship construct and organization, learning and feedback to represent a process construct. A third, more complex model was tested that was identical to model two but with enthusiasm and rapport tested as separate constructs. Model 4 was made more complex by making it identical to Model 3 except that learning, enthusiasm and rapport were tested as separate constructs. Model 5 – the hypothesized target model – was the most complex model, with the five constructs tested for model–data fit separately. Descriptions of each of the five models appear in Table I.

Model fit was assessed by comparing fit indices among a progression of increasingly complex models until more factors no longer added significant discrimination of variables. Traditional fit indices such as the χ^2 , root mean square residual and goodness-of-fit index are inadequate because they are influenced by sample size and other factors (Jackson *et al.*, 2009).

Model	Description
1. One factor	All items hypothesized to load onto a single factor
2. Two factors	Items for organization, learning and feedback represent process and items for enthusiasm and rapport represent relationship
3. Three factors	As Model 2 except enthusiasm and rapport hypothesized separately
4. Four factors	As Model 2 except learning, enthusiasm and rapport hypothesized separately
5. Five factors	Target model; items hypothesized to load on intended constructs

Table I.
Descriptions of the
five confirmatory
factor analyses

Since using a single index to assess model fit is inadequate, four fit statistics were chosen to assess the target five-factor model against the four competing models – the normed fit index (NFI2), the comparative fit index (CFI) the root mean square error of approximation (RMSEA) and the χ^2/df ratio (Hooper *et al.*, 2008). χ^2 difference tests were calculated between the target five-factor model and each of the four competing models to determine whether the competing models represented more parsimonious relationships among data.

Within-and-between analysis

WABA assesses variation and covariation among constructs within groups vs between groups (Yammarino and Dansereau, 2009). In the current study, groups are defined as students rating perceptions of instructor behaviors in their respective classrooms. WABA can be used to assess single-source bias when both between-group and within-group variability are assessed to conclude one of four possible outcomes (Martin *et al.*, 2010). Significant variation between but not within groups represents a wholes inference, where participants who rate instructors can be grouped according to instructor. Significant variation within but not between groups represents a parts inference, where participants are hierarchically ranked within groups. No significant variation either between or within groups represents the null case. The fourth possible conclusion suggests single-source bias, where there is significant variability both between and within groups – the equivocal inference. An equivocal inference suggests that response variation is due to individual differences rather than from inclusion in groups. In the case of students rating instructors on five behaviors by filling out a survey, an equivocal inference for the hypothesized constructs suggests that variation in data is due to individual raters, a result consistent with single-source bias. Although there might have been variation among instructors’ behaviors, the single-source collection method makes it impossible to distinguish these attributes both practically and statistically (McCrae, 2018).

Results

Descriptive statistics from the data, including means, standard deviations, reliabilities and inter-correlations, are shown in Table II. Results suggest internal reliability for each of the constructs. Significant un-hypothesized inter-correlations of a similar magnitude suggest that the individual items are not measuring the constructs discriminately. These correlations represent an indication of a problem, not a means by which to make conclusions about the structure present in the data or suggest that the structure is congruent with the hypothesized five-factor structure. However, they do warrant further investigation.

Method I – exploratory factor analysis

An EFA was conducted to assess the loadings of each item on the five-factor model. In an initial analysis (principal components extraction, eigenvalues equal to 1 and varimax orthogonal rotation), the data failed to converge, so no rotation or output was possible. This failure is synonymous with a one-factor structure. However, researchers too commonly use

Table II.
Means, standard deviations, inter-correlations and reliabilities

Variable	Mean	SD	1	2	3	4	5
1. Organization	4.13	0.77	(0.88)				
2. Learning	3.99	0.79	0.87*	(0.89)			
3. Feedback	3.85	0.87	0.84*	0.86*	(0.85)		
4. Enthusiasm	4.01	0.84	0.87*	0.87*	0.84*	(0.89)	
5. Rapport	3.98	0.80	0.82*	0.85*	0.83*	0.82*	(0.84)

Notes: Cronbach’s α coefficients shown on the diagonal. * $p < 0.01$

default software package settings to conduct EFAs, and oblique rotations, in which statistical cut-off criteria are relaxed, are appropriate when constructs are hypothesized or suspected to correlate (Conway and Huffcut, 2003). Since there are five hypothesized constructs, setting the number of factors extracted to five is appropriate in lieu of allowing the statistical program to determine how many factors are present in the data. Results from a maximum likelihood extraction, five-factor and direct oblimin rotation EFA appear in Table III.

Results from the EFAs demonstrate that there is little evidence for the hypothesized five-factor structure. The oblique rotation suggests a one-factor structure with a few un-hypothesized cross-loadings, correlations that should not exist if the five constructs are unrelated. Factor loadings obtained from the EFA do not offer even cursory or preliminary evidence of any structure beyond a one-factor structure. Factors 2–5 obtained through the rotated matrix show no clustering of shared variance around identifiable factors; all loadings cluster on one factor. Evidence of one factor rather than five factors suggests that students were unable to distinguish multiple instructor behaviors while rating their instructors with the SAI instrument. This inability to identify disparate behaviors suggests that aggregate perceptions of these behaviors are the result of single-source bias.

Method II – nested confirmatory factor analysis

NCFA was conducted to validate the five-factor model against the null hypothesis that a five-factor structure is not present in the data. Four competing models, including a one-factor model as suggested by the EFA, were tested against the five-factor model. The results of the NCFA appear in Table IV.

Fit indices for the target five-factor model indicate adequate fit. Both the CFI (0.953) and NFI2 (0.954) were above the suggested 0.9 threshold for adequate fit, and the RMSEA was below the 0.8 value to suggest adequate fit. The χ^2/df ratio was between 2 and 3, suggesting that the five-factor model provided adequate fit. However, examination of the same statistics for the competing models shows that there is no evidence to support that the five-factor model provides better fit than the other models, including the one-factor model; the CFI and NFI2 fit indices for

Item	Factors				
	1	2	3	4	5
1	0.67	0.51	-0.03	-0.03	0.03
2	0.77	0.35	0.02	0.10	0.02
3	0.88	0.12	0.03	0.04	0.46
4	0.73	-0.02	-0.03	0.07	0.04
5	0.82	0.07	-0.03	0.04	-0.01
6	0.85	0.10	-0.05	0.10	-0.04
7	0.80	0.01	-0.05	0.12	0.02
8	0.75	-0.02	0.04	-0.11	0.09
9	0.81	-0.07	0.00	-0.16	0.03
10	0.82	0.03	0.03	-0.13	-0.05
11	0.84	0.13	0.07	0.01	-0.10
12	0.86	-0.06	-0.08	0.34	-0.04
13	0.55	-0.02	-0.06	-0.21	-0.07
14	0.76	-0.06	-0.10	-0.21	-0.04
15	0.83	-0.07	-0.02	0.03	-0.04
16	0.83	0.00	0.09	0.04	0.11
17	0.82	-0.14	-0.04	0.02	-0.03
18	0.82	-0.03	0.57	0.01	0.02
19	0.75	-0.16	0.01	0.04	-0.01
20	0.80	-0.10	0.05	0.03	-0.14

Table III.
Exploratory factor
analysis: rotated
factor matrix

each of the models show no significant difference in each of the models' ability to describe the data. The RMSEA and χ^2/df ratios were similar for all five models. Finally, the change in χ^2 statistic from each of the competing models to the five-factor model shows some significant differences among the models. However, this was due primarily to changes to degrees of freedom associated with each of the more restrictive models rather than an indication of better fit (Bollen and Long, 1993). These results suggest no justification for a five-factor model. Since none of the models demonstrated better fit in comparison to the one-factor model, the most parsimonious model, the one-factor model, cannot be rejected. These results demonstrate that students were unable to distinguish disparate instructor behaviors using the SAI instrument, resulting from aggregate perceptions regarding the instructor. In essence, single-source bias made it impossible to justify the presence of five constructs in the data.

Method III – within-and-between analysis

Since the number of raters per instructor varied, and in line with Avolio and Yammarino (1991), the number of raters per instructor was kept constant by choosing ten raters randomly for each of thirteen instructors. Results presented in Table V suggest that based on the *E*-test for practical significance, within-group η s were larger than between-group η 's, prompting a test for a parts inference. However, each construct collected with the SAI instrument failed the *F*-test for statistical significance, suggesting an equivocal inference. Results also indicate that seven of ten between-group correlations were larger than within-group correlations based on the *A*-test for practical significance and *Z*-test for statistical significance. Both between- and within-group correlations were significant based on the *R*-test and *t*-test, suggesting an equivocal inference. Raw-score correlations and component between- and within-group correlations are reported. Examining WABA analyses in aggregate, the equivocal inferences drawn from analysis of between and within η s, and between and within correlations, suggest that individual differences instead of an identifiable factor drove results. These results suggest that the inability of raters to distinguish disparate instructor behaviors resulted from single-source bias and that variation between and within groups of raters is due to individual differences rather than inclusion in groups.

Discussion of results

The three analyses conducted in this study suggest that the hypothesized five-factor structure is not evident and that single-source bias resulted in a one-factor structure. Since the committee that created the instrument did not hypothesize a one-factor model, there is question regarding what the instrument measures and what conclusions can be drawn from data collected. Triangulation of results suggests that without testing, an SAI instrument collection method can have inherent flaws that skew perceptions of teaching quality and effectiveness. Since teaching effectiveness is used as a primary criterion for tenure, promotion and performance excellence, administrators must be cognizant of the validity and reliability of similar assessment instruments when measuring instructor success. A well-planned

Model	χ^2	df	χ^2/df	RMSEA	CFI	NFI2	$\Delta\chi^2$	Δdf
Independence	6,392.42	190	–	–	–	–	–	–
1	490.44	170	2.89	0.073	0.948	0.949	41.89*	10
2	489.66	169	2.90	0.073	0.948	0.948	41.11*	9
3	471.69	167	2.82	0.072	0.951	0.951	23.14*	7
4	464.67	164	2.83	0.072	0.952	0.952	16.12*	4
5 ^a	448.55	160	2.80	0.071	0.953	0.954	–	–

Notes: ^aThe target, five-factor model. **p* < 0.01

Table IV.
Results of the
nested confirmatory
factor analyses

Variable	Raw-score correlation ^c	η^2 's ^a		Correlations ^b		Component		Inference
		Between	Within	Between	Within	Between	Within	
Organization with Learning	0.88 ^{††**}	0.53	0.85 [†]	<i>0.95</i>	<i>0.86</i>	0.23	0.65	Equivocal/Individual
Feedback	0.80 ^{††**}	0.53	0.85 [†]	<i>0.93^{†*}</i>	<i>0.74</i>	0.26	0.53	Equivocal/Individual
Enthusiasm	0.89 ^{††**}	0.57	0.82 [†]	<i>0.97^{†*}</i>	<i>0.85</i>	0.29	0.59	Equivocal/Individual
Rapport	0.84 ^{††**}	0.49	0.87 ^{††}	<i>0.96^{†**}</i>	<i>0.80</i>	0.25	0.59	Equivocal/Individual
Learning with Feedback	0.86 ^{††**}	0.45	0.89 ^{††}	<i>0.94</i>	<i>0.84</i>	0.23	0.63	Equivocal/Individual
Enthusiasm	0.89 ^{††**}	0.57	0.82 [†]	<i>0.96[*]</i>	<i>0.87</i>	0.25	0.64	Equivocal/Individual
Rapport	0.89 ^{††**}	0.49	0.87 ^{††}	<i>0.97^{†*}</i>	<i>0.86</i>	0.22	0.67	Equivocal/Individual
Feedback with Enthusiasm	0.83 ^{††**}	0.53	0.85 [†]	<i>0.94^{†*}</i>	<i>0.78</i>	0.29	0.54	Equivocal/Individual
Rapport	0.81 ^{††**}	0.49	0.87 ^{††}	<i>0.93^{†*}</i>	<i>0.77</i>	0.25	0.56	Equivocal/Individual
Enthusiasm with Rapport	0.87 ^{††**}	0.57	0.82 [†]	<i>0.94</i>	<i>0.84</i>	0.27	0.60	Equivocal/Individual

Notes: $n = 130, J = 10$. ^aSignificant E -test (†) results of the difference between within and between η^2 's. No F -tests were significant. ^bSignificant between- and within-cell correlations based on R -test and t -test results are in italic. Significant A -test (†) and Z -test ($*$) results of the difference between within and between cell correlations. ^cSignificant raw-score correlations based on R -test (†) and t -test ($*$) results. † = 15 degrees; †† = 30 degrees. * $p < 0.05$; ** $p < 0.01$

Table V.
Results of within-and-between analysis

instrument might yield reliable information regarding student perceptions of instruction, but data obtained from such an instrument might not provide information institutions want or should use to make summative evaluations of instructors. Evaluation of teaching effectiveness should be conducted using multiple methods and a variety of techniques. Relying too heavily on data from one source such as SAIs can be inappropriate no matter how intently an institution develops an instrument to collect such data.

Practical implications

Sproule (2000) outlined potential statistical fallacies associated with student evaluations of teaching, such as cardinal measurement errors and ordinal measures of teaching effectiveness in the absence of statistical controls. Corroborating these findings, this study demonstrates one aspect of statistical invalidity in SAI instruments. We argue that single-source bias further undermines the validity of such instruments. These instruments might yield data that are efficient but inaccurate reflections of student perceptions of disparate instructor behaviors. Administrations should resist relying on single methods and tools for summative evaluations and should instead use multiple and alternative techniques subject to the same rigorous analyses to which traditional cafeteria-style SAI instruments have been subjected. Single-source bias might be obviated by using several methods of collecting information concerning instruction. This can be accomplished using peer visitation, self-assessments, student exit interviews, exit exams tied to learning outcomes and portfolio assessments, which

include materials such as student samples, syllabi and teaching philosophies (Algozzine *et al.*, 2004). Multi-method, multi-trait data collection can be applied to an SAI instrument directly to address single-source bias and other statistical problems. For example, SAI instruments keyed to traits might be administered at different times during a semester. Multiple collection would also allow correlation across time and people. Frey (1976) suggested no difference between SAI instrument data when collected at the end of the first week of classes from data collected at the end of the semester, so SAI instruments, focusing on different attributes, administered throughout a semester might yield more accurate data. Students asked to focus on a single trait might be better able to differentiate and evaluate those traits. The method might address part of the problem that single-source bias suggests – the inability of students to differentiate teaching traits from generalized emotional sympathy for an instructor.

The inability of students to differentiate traits of their instructors suggests a more significant course of action; students could be trained to be better assessors. The reason students do not differentiate among instructor traits is limited to either an inability or unwillingness to do so. Both assumptions have been addressed obliquely by SAI instrument research that focuses on how to administer SAI instruments, but such research focuses on the method of data collection, such as timing, the purpose of evaluation, anonymity of responses and the presence of an instructor during evaluation (Wachtel, 1998). Focusing solely on the instrument – the method of data collection – neglects or tangentially addresses a basic issue with data collection – the source of data. Many SAI instruments are administered at the beginning of a class period rather than at the end, in part so that students do not rush to complete the assessment to leave class sooner. This method is employed to address a student's unwillingness to take the time to assess an instructor meaningfully. Another approach is to institutionalize training of students to assess instructors, including the ability to differentiate positive and negative attributes. Such training could take place *ad hoc* within each course prior to SAI instrument deployment. It could also be integrated into institutional curricula as part of a general education program. Such instruction should focus on methods for assessing and evaluating peers, subordinates and superiors, and it should draw from research on educational theory, psychology and management, yielding valid results of differentiation of instructor attributes when students are asked to assess instruction. Doing so would also prepare students for careers systematically that little coursework already does unless they happen to be, for example, business majors who take managerial or organizational behavior classes, or education majors who are trained in assessment of student development and learning outcomes. The ability to use clear, valid criteria to differentiate and appropriately assess other people's attributes, characteristics and capabilities is itself a learning outcome desirable and marketable in a diverse global community.

Research suggests that students can function at a similar level of effectiveness as trained observers when rating instructor behaviors. Murray (1983) examined eight clusters of behaviors associated with Marsh's (1982) study that identifies global factors associated with student perceptions of quality instruction. These clusters define behavioral characteristics that observers of teaching could identify objectively, such as various speaking behaviors, non-verbal behaviors (e.g. gestures), the nature of an instructor's explanations in support of material being presented, perceptions of organization of presented material, interest in the subject matter that the instructor is perceived to display, an instructor's task orientation, the rapport an instructor has with students and the degree to which an instructor facilitates student participation in the classroom. Training students to assess disparate traits reliably begins with communicating to students the attributes that describe such traits. The first such communication would most appropriately take place during the first semester of enrollment, with a brief summary/reminder at the beginning of each course and one immediately prior to students' evaluation of an instructor. Abrami *et al.* (2007) advocated this idea, developing a model to aid with objectifying elements of desired instructor behaviors.

The focus of this paper is on student evaluations of instruction, but research spanning nearly 30 years suggests that student ratings are only one measure of faculty teaching effectiveness. Marsh (2007) argued that student evaluations should be one of many elements that factor into assessment of faculty performance. Abrami *et al.* (2007) suggested that student ratings measure only a narrow element of the teaching process – student satisfaction with teaching – and that a distinction must be made among items that measure instructional products as opposed to those that measure instructional processes, such as instructor friendliness. Kolitch and Dean (1999) found that student evaluations can be based on flawed models, whereby the design of student rating instruments is biased toward outmoded models such as teacher-centered behaviors (vs engaged-critical teaching), which tend to reinforce suboptimal instruction. Marsh (2007) argued that “practitioners and researchers alike give lip-service to the adage that teaching effectiveness should be evaluated with multiple indicators of teaching, not just [student evaluations]” (p. 343). He cautioned researchers that there is a dearth of research that relates student ratings of instruction to other constructs of interest within education when examining student development, including self-motivation, learning quality and character development.

A final implication is that individuals who charge university committees with recommending an instrument to be used by students to rate faculty performance experience a unique challenge. The literature is filled with examples of instruments that are unvalidated and that have poor psychometric properties. Researchers are typically exceptionally competent in their chosen fields, but might not possess the depth of understanding in the narrow domain of student assessment of faculty teaching necessary to create such an instrument. Since these tools are used to assess instructor performance, instruments that lack validity lead to flawed assumptions about teaching competence and effectiveness. These flawed assumptions should be considerations while an instrument is being created and before administrators rely on results from such instruments during faculty reviews. Committee members might infer from university administrators that an expectation exists to create an SAI instrument that is unique to the university. Committees might find the challenge of developing an instrument sufficiently compelling to overpower members’ knowledge of the process of designing and validating instruments. In contrast, there are instruments that can be implemented readily that measure SAI well and that have been shown to exhibit excellent psychometric properties. Examples of such instruments that focus on evaluation of students’ experiences in a course include the Students’ Evaluation of Educational Quality (Richardson, 2005) and the Course Experience Questionnaire (Richardson and Woodley, 2001).

Conclusion

An instructor’s role during learning is a significant element of student success both in a course and for lifelong learning. Readers are likely able to recall instructor behaviors that influenced their life’s trajectory, even those that had lasting effects decades later. Responsible selection and use of SAI are critical to identifying and rewarding instructors who influence positive learning experiences and whose behaviors result in shaping student trajectories in careers and life. Broad categories of stakeholders have an interest in selection and use of a particular SAI instrument that make for compelling motives to design those processes with balanced consideration of the needs represented by each stakeholder group. These groups and their interests include faculty members, whose primary interests include course feedback and career influence; administrators, who leverage SAI outcomes to determine pay, promotion and tenure decisions; current students, who earn a voice in course and instructor feedback, and future students, who benefit from improvements implemented because of SAI outcomes; and the university, which enjoys enhanced reputation and ranking because of positive SAIs, using SAIs to identify areas of organization-wide accountability and opportunities so overall quality of instruction improves. New SAIs or

modification of existing ones must be introduced in a way that builds support for the instrument. Meaningful input from representatives of each stakeholder group and a needs assessment should precede changes to SAIs. Implementation of changes to a university's SAI process should include a rationale regarding how the new process addresses issues identified during needs assessment. Replacing one flawed instrument for another is not only counterproductive, it risks unintended consequences concerning behaviors of individuals in stakeholder groups. Using SAIs represents the possibility of enhancing rapport between instructor and student, instructor and direct supervisor, and instructor and administrator. A quality SAI program strengthens a university's culture, student engagement and collaboration between students and other members of the university community.

References

- Abrami, P.C., d'Appolonia, S. and Rosenfield, S. (2007), "The dimensionality of student ratings of instruction: what we know and what we do not", in Perry, R.P. and Smart, J.C. (Eds), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-based Perspective*, Springer, New York, NY, pp. 385-456.
- Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., Mohanty, G. and Spooner, F. (2004), "Student evaluation of college teaching: a practice in search of principles", *College Teaching*, Vol. 52 No. 4, pp. 134-141.
- Antonakis, J., Avolio, B.J. and Sivasubramaniam, N. (2003), "Context and leadership: an examination of the nine-factor full-range leadership theory using the multifactor leadership questionnaire", *Leadership Quarterly*, Vol. 14 No. 3, pp. 261-295.
- Arreola, R. (1995), *Developing a Comprehensive Faculty Evaluation System*, Anker, Bolton, MA.
- Asparouhov, T. and Muthén, B. (2009), "Exploratory structural equation modeling", *Structural Equation Modeling*, Vol. 16 No. 3, pp. 397-438.
- Avolio, B.J. and Yammarino, F.J. (1991), "Identifying common methods variance with data collected from a single source: an unresolved sticky issue", *Journal of Management*, Vol. 17 No. 3, pp. 571-587.
- Avolio, B.J., Sivasubramaniam, N., Murry, W.D., Jung, D. and Garger, J.W. (2003), "Assessing shared leadership: development and preliminary validation of a team multifactor leadership questionnaire", in Pearce, C.L. and Conger, J.A. (Eds), *Shared Leadership: Reframing the Hows and Whys of Leadership*, Sage, Thousand Oaks, CA, pp. 143-172.
- Bagozzi, R.P. and Yi, Y. (1991), "Multitrait-multimethod matrices in consumer research", *Journal of Consumer Research*, Vol. 17 No. 4, pp. 426-439.
- Baugh, S.G., Hunt, J.G. and Scandura, T.A. (2006), "Reviewing by the numbers: evaluating quantitative research", in Baruch, Y., Sullivan, S.E. and Schepmyer, H.N. (Eds), *Winning Reviews: A Guide for Evaluating Scholarly Writing*, Palgrave Macmillan, New York, NY, pp. 156-172.
- Bollen, K.A. and Long, J.S. (Eds) (1993), *Testing Structural Equation Models*, Sage, Newbury Park, CA.
- Boring, A., Ottoboni, K. and Stark, P.B. (2016), "Student evaluations of teaching (mostly) do not measure teaching effectiveness", *Science Open Research*, pp. 1-11, available at: www.scienceopen.com/document_file/e211a751-3e15-4b54-b79a-45c7fac2d97a/PubMedCentral/e211a751-3e15-4b54-b79a-45c7fac2d97a.pdf
- Brown, T. (2015), *Confirmatory Factor Analysis for Applied Research*, Guilford Press, New York, NY.
- Browne, M.W. (2001), "An overview of analytic rotation in exploratory factor analysis", *Multivariate Behavioral Research*, Vol. 36 No. 1, pp. 111-150.
- Campbell, D.T. and Fiske, D. (1959), "Convergent and discriminant validation by the multitrait-multimethod matrix", *Psychological Bulletin*, Vol. 56 No. 2, pp. 81-105.
- Centra, J. and Diamond, R. (1987), *A Guide to Evaluating Teaching for Promotion and Tenure*, Copley Publications Group, Littleton, CO.

-
- Conway, J. M. and Huffcut, A.I. (2003), "A review and evaluation of exploratory factor analysis practices in organizational research", *Organizational Research Methods*, Vol. 6 No. 2, pp. 147-168.
- Darling-Hammond, L. and Hyler, M.E. (2013), "The role of performance assessment in developing teaching as a profession", *Rethinking Schools*, Vol. 27 No. 4, pp. 10-15.
- Dipboye, R.L. and Flanagan, M.F. (1979), "Research settings in industrial and organizational psychology: are findings in the field more generalizable than in the laboratory?", *American Psychologist*, Vol. 34 No. 2, pp. 141-150.
- Feeley, T.H. (2002), "Evidence of halo effects in student evaluations of communication instruction", *Communication Education*, Vol. 51 No. 3, pp. 225-236.
- Frey, P.W. (1976), "Validity of student instructional ratings: does timing matter?", *Journal of Higher Education*, Vol. 47 No. 3, pp. 327-336.
- Heller, S. (1984), "Math department balks at officials' effort to require forms for student evaluation", *Chronicle of Higher Education*, Vol. 28 No. 8, p. 24.
- Hooper, D., Coughlan, J. and Mullen, M.R. (2008), "Structural equation modelling: guidelines for determining model fit", *Journal of Business Research Methods*, Vol. 6 No. 1, pp. 53-60.
- Hornstein, H. (2017), "Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance", *Journal of Cogent Education*, Vol. 4 No. 1, available at: www.tandfonline.com/doi/full/10.1080/2331186X.2017.1304016
- Jackson, D.L., Gillaspay, J.A. and Purc-Stephenson, R. (2009), "Reporting practices in confirmatory factor analysis: an overview and some recommendations", *Psychological Methods*, Vol. 14 No. 1, pp. 6-23.
- Kalra, V., Negi, G. and Chauhan, V. (2015), "Comparison of two formats for student evaluation of teacher effectiveness", *International Journal of Applied and Basic Medical Research*, Vol. 5 No. 1, pp. 1-4.
- Kline, R.B. (2010), *Principles and Practice of Structural Equation Modeling*, 3rd ed., Guilford Press, New York, NY.
- Kolitch, E. and Dean, A.V. (1999), "Student ratings of instruction in the USA: hidden assumptions and missing conceptions about 'good' teaching", *Studies in Higher Education*, Vol. 24 No. 1, pp. 27-42.
- McAlpin, V., Algozzine, M. and Lee, N. (2014), "A comparison of web-based and paper-based course evaluations", *Journal of Academic Administration in Higher Education*, Vol. 10 No. 2, pp. 49-58.
- McCrae, R.R. (2018), "Method biases in single-source personality assessments", *Psychological Assessment*, Vol. 30 No. 9, pp. 1160-1173.
- Marsh, H.W. (1982), "SEEQ: A reliable, valid, and useful instrument for collecting students' evaluation of university teaching", *British Journal of Educational Psychology*, Vol. 52 No. 1, pp. 77-95.
- Marsh, H.W. (2007), "Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and usefulness", in Perry, R.P. and Smart, J.C. (Eds), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-based Perspective*, Springer, New York, NY, pp. 319-384.
- Marsh, H.W. and Roche, L. (1997), "Making students' evaluations of teaching effectiveness effective: the central issues of validity, bias, and utility", *American Psychologist*, Vol. 52 No. 1, pp. 1187-1197.
- Martin, R., Epitropaki, O., Thomas, G. and Topakas, A. (2010), "A review of leader-member exchange research: future prospects and directions", in Hodgkinson, G.P. and Ford, J.K. (Eds), *International Review of Industrial and Organizational Psychology*, Wiley-Blackwell, West Sussex, pp. 35-88.
- Mitchell, T.R. (1985), "An evaluation of the validity of correlational research conducted in organizations", *Academy of Management Review*, Vol. 10 No. 2, pp. 192-205.
- Murray, H.G. (1983), "Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness", *Journal of Educational Psychology*, Vol. 75 No. 1, pp. 138-149.
- Nasser, F. and Fresko, B. (2002), "Faculty views of student evaluation of college teaching", *Assessment and Evaluation in Higher Education*, Vol. 27 No. 2, pp. 187-198.

- Neumann, R. (2000), "Communicating student evaluation of teaching results: rating interpretation guides (RIGs)", *Assessment and Evaluation in Higher Education*, Vol. 25 No. 2, pp. 121-134.
- Nunnally, J.C. (1978), *Psychometric Theory*, 2nd ed., McGraw-Hill, New York, NY.
- Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y. and Podsakoff, N.P. (2003), "Common method biases in behavioral research: a critical review of the literature and recommended remedies", *Journal of Applied Psychology*, Vol. 88 No. 5, pp. 879-903.
- Richardson, J.T.E. (2005), "Instruments for obtaining student feedback: a review of the literature", *Assessment and Evaluation in Higher Education*, Vol. 30 No. 4, pp. 387-415.
- Richardson, J.T.E. and Woodley, A. (2001), "Perceptions of academic quality among students with a hearing loss in distance education", *Journal of Educational Psychology*, Vol. 93 No. 3, pp. 563-570.
- Schermelleh-Engel, K., Moosbrugger, H. and Müller, H. (2003), "Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures", *Methods of Psychological Research Online*, Vol. 8 No. 2, pp. 23-74.
- Schmitt, T.A. (2011), "Current methodological considerations in exploratory and confirmatory factor analysis", *Journal of Psychoeducational Assessment*, Vol. 29 No. 4, pp. 304-321.
- Seldin, P. (1993), "The use and abuse of student ratings of professors", *Chronicle of Higher Education*, Vol. 39 No. 46, p. A40.
- Shelton, L.R. and Hayne, A.N. (2017), "Developing an instrument for evidence-based peer review of faculty online teaching", *Nursing Education Perspectives*, Vol. 38 No. 3, pp. 157-158.
- Skelton, A. (2004), "Understanding teaching excellence in higher education: a critical evaluation of the national teaching fellowship scheme", *Studies in Higher Education*, Vol. 2 No. 4, pp. 451-468.
- Spector, P.E. (1987), "Method variance as an artifact in self-report affect and perceptions at work: myth or significant problem?", *Journal of Applied Psychology*, Vol. 72 No. 3, pp. 438-444.
- Spooren, P. and Mortelmans, D. (2006), "Teacher professionalism and student evaluation of teaching: will better teachers receive higher ratings and will better students give higher ratings?", *Educational Studies*, Vol. 32 No. 2, pp. 201-214.
- Spooren, P., Brockx, B. and Mortelmans, D. (2013), "On the validity of student evaluation of teaching: the state of the art", *Review of Educational Research*, Vol. 83 No. 4, pp. 598-642.
- Sproule, R. (2000), "Student evaluation of teaching: a methodological critique of evaluation practices", *Education Policy Analysis Archives*, Vol. 8 No. 50, available at: <http://epaa.asu.edu/epaa/v8n50.html> (accessed February 19, 2019).
- Stark-Wroblewski, K., Ahlering, R.F. and Brill, F.M. (2007), "Toward a more comprehensive approach to evaluating teaching effectiveness: supplementing student evaluations of teaching with pre-post learning measures", *Assessment & Evaluation in Higher Education*, Vol. 32 No. 4, pp. 403-415.
- Taut, S., Jiménez, D., Puente-Duran, S., Palacios, D., Inés Godoy, M. and Manzi, M. (2018), "Evaluating the quality of teaching: can there be valid differentiation in the middle of the performance distribution?", *School Effectiveness and School Improvement*, doi: 10.1080/09243453.2018.1510842.
- Thompson, B. (2004), *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*, American Psychological Association, Washington, DC.
- Thorndike, E.L. (1920), "A constant error in psychological ratings", *Journal of Applied Psychology*, Vol. 4 No. 1, pp. 25-29.
- Wachtel, H.K. (1998), "Student evaluation of college teaching effectiveness: a brief review", *Assessment and Evaluation in Higher Education*, Vol. 23 No. 2, pp. 191-211.
- Yammarino, F.J. and Dansereau, F. (2009), "A new kind of organizational behavior", in Yammarino, F.J. and Dansereau, F. (Eds), *Multi-level Issues in Organizational Behavior and Leadership*, Emerald Group Publishing, Bingley, UK, pp. 13-60.
- Zhang, J., Jensen, B.E. and Mann, B.L. (1997), "Modification and revision of the leadership scale for sport", *Journal of Sport Behavior*, Vol. 20 No. 1, pp. 105-122.

Appendix

Organization

- My instructor is well prepared for class meetings.
- My instructor explains the subject matter clearly.
- My instructor clearly communicates course goals and objectives.
- My instructor answers questions appropriately.

Enthusiasm

- My instructor is enthusiastic about teaching this course.
- My instructor presents the subject in an interesting manner.
- My instructor stimulates my thinking.
- My instructor motivates me to do my best work.

Rapport

- My instructor helps students sufficiently with course-related issues.
- My instructor is regularly available for consultation.
- My instructor is impartial in dealing with students.
- My instructor respects opinions different from his or her own.

Feedback

- Assessment methods accurately assess what I have learned in this course.
- Grades are assigned fairly.
- The basis for assigning grades is clearly explained.
- My instructor provides feedback on my progress in the course on a regular basis.

Learning

- My instructor advances my knowledge of course content.
- My instructor promotes my understanding of important conceptual themes.
- My instructor enhances my capacity to communicate effectively about the course subject matter.
- My instructor encourages me to value new viewpoints related to the course.

Corresponding author

John Garger can be contacted at: john@johngarger.com

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com