

2011

Constructivist and Behaviorist Approaches: Development and Initial Evaluation of a Teaching Practice Scale for Introductory Statistics at the College Level

Rossi A. Hassad
Mercy College, New York, USA, Rhassad@mercy.edu

Follow this and additional works at: <http://scholarcommons.usf.edu/numeracy>

 Part of the [Mathematics Commons](#), and the [Science and Mathematics Education Commons](#)

Recommended Citation

Hassad, Rossi A. (2011) "Constructivist and Behaviorist Approaches: Development and Initial Evaluation of a Teaching Practice Scale for Introductory Statistics at the College Level," *Numeracy*: Vol. 4 : Iss. 2 , Article 7.

DOI: <http://dx.doi.org/10.5038/1936-4660.4.2.7>

Available at: <http://scholarcommons.usf.edu/numeracy/vol4/iss2/art7>

Constructivist and Behaviorist Approaches: Development and Initial Evaluation of a Teaching Practice Scale for Introductory Statistics at the College Level

Abstract

This study examined the teaching practices of 227 college instructors of introductory statistics from the health and behavioral sciences. Using primarily multidimensional scaling (MDS) techniques, a two-dimensional, 10-item teaching-practice scale, TISS (Teaching of Introductory Statistics Scale), was developed. The two dimensions (subscales) are characterized as constructivist and behaviorist; they are orthogonal. Criterion validity of the TISS was established in relation to instructors' attitude toward teaching, and acceptable levels of reliability were obtained. A significantly higher level of behaviorist practice (less reform-oriented) was reported by instructors from the U.S., as well as instructors with academic degrees in mathematics and engineering, whereas those with membership in professional organizations, tended to be more reform-oriented (or constructivist). The TISS, thought to be the first of its kind, will allow the statistics education community to empirically assess and describe the pedagogical approach (teaching practice) of instructors of introductory statistics in the health and behavioral sciences, at the college level, and determine what learning outcomes result from the different teaching-practice orientations. Further research is required in order to be conclusive about the structural and psychometric properties of this scale, including its stability over time.

Keywords

Statistics, Introductory, Literacy, Scale, Constructivist, Behaviorist, Teaching

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

Cover Page Footnote

Rossi A. Hassad is an associate professor in the School of Social & Behavioral Sciences, Mercy College, New York, where he teaches at both the undergraduate and graduate levels, and serves as a co-chair of the College-wide committee on student learning assessment. His teaching expertise and research interests include statistical literacy and evidence-based practice in the health and behavioral sciences. He also serves as an adjunct associate professor at Hunter College, Department of Psychology (City University of New York), and was the recipient of the American Statistical Association (Joint Statistical Meetings 2003) award for best contributed paper (Section on the Teaching of Statistics in the Health Sciences).

Introduction

The ability to critically evaluate research findings expressed in statistical language is an essential skill for practitioners and students in evidence-based disciplines such as the health and behavioral sciences (Belar 2003; Garfield and Ben-Zvi 2007a, 2009b; Hassad 2010). Undergraduate students in these disciplines, therefore, are generally required to take introductory statistics as a core course. Consequently, there is a consensus among educators that the goal of this course should be to facilitate statistical literacy through active learning strategies emphasizing concepts and applications rather than mathematical procedures (Franklin and Garfield 2006; Hassad and Coxon 2007; Froelich et al. 2008). Also motivating this pedagogical approach is the realization that, for the majority of these students, the introductory course will be their only formal exposure to statistics (Moore 1998).

For over a decade there has been emphasis on reform-oriented teaching at the college level, fueled by a consensus among educators that traditional curricular material and pedagogical strategies have not been effective in promoting statistical literacy (Cobb 1992; delMas et al. 2006; Garfield et al. 2002; Hassad 2009), an essential component of quantitative literacy (Steen 2004). In spite of these reform efforts focused on course content, pedagogy, assessment, and integration of technology, research continues to show that students are emerging with a lack of understanding of core concepts (delMas et al. 2006; Green et al. 2009). Such evidence has raised concerns about instructors' level of awareness, understanding, and appropriate use of active learning strategies (Hassad 2007). Also, empirical information on what core strategies underlie reform-oriented teaching of introductory statistics is lacking (Garfield et al. 2002). This is a major impediment to characterizing teaching practice and assessing the effectiveness of reform-oriented teaching compared to the traditional pedagogical approach.

Objective

The objective of the study reported here was to develop and validate a scale (instrument) to empirically assess and describe the teaching practice of instructors of introductory statistics in the health and behavioral sciences at the college level. Such a scale can be used to characterize teaching practice, toward identifying individual strengths and weaknesses regarding reform-oriented (constructivist or concept-based) teaching of introductory statistics. More importantly, this teaching-practice scale can help to determine what learning outcomes result from the different teaching-practice orientations considered, namely constructivist and behaviorist (traditional).

Theoretical and Conceptual Basis of Reform-Oriented Pedagogy

The conceptual underpinning of reform-oriented pedagogy is typically described with reference to theories of learning based on constructivism, which is considered a family of concepts and principles about the construction of knowledge and meaning (von Glasersfeld 1987; Cobb 1994; delMas et al. 1999; Trigwell and Prosser 2004; Fosnot 2005). Strictly speaking, constructivism is not a philosophy of learning; it is “a model of knowing that is pedagogically useful” (Thompson 2000, p. 423) and supports multiple teaching approaches and strategies. Indeed, constructivism is neither limited to reform-based teaching of introductory statistics nor the broader education context. However, there is a large and increasing body of scientific research that recognizes constructivism as a major theoretical influence in contemporary and reform-based science, mathematics and statistics education (Garfield 1993; Steinhurst and Keeler 1995; Mills 2002; Mvududu 2003; Fosnot et al. 2007; Froelich et al. 2008; Garfield and Ben-Zvi 2009).

Moreover, in the education literature (particularly mathematics and statistics) the expression, “constructivist pedagogy” is widely used interchangeably and synonymously with “reform-based” and “reform-oriented” teaching (Mvududu 2003; Fosnot 2005; Luft and Roehrig 2007; Pecore 2009; Jenkins 2010). Notably, the rationale for the use of the constructivist label in introductory statistics education is premised on the understanding that reform-based pedagogy, in this context, emphasizes “learners’ active participation” and “the social nature of learning,” which are the core principles of constructivism (Liu and Matthews 2005).

Another perspective of the constructivist paradigm is that it serves as an epistemological model, which defines knowledge as “temporary, developmental, socially and culturally mediated, and thus, non-objective” (Brooks and Brooks 1993, p. vii). There are two recognized forms of constructivism: cognitive or Piagetian constructivism (Piaget 1950, 1967, 1977), and social or Vygotskian constructivism (Vygotsky 1962, 1978). Cognitive constructivism emphasizes the mind of the individual, and views learning as simply the assimilation and accommodation of new knowledge by learners, in other words, merely a process of adjusting our mental models to accommodate new experiences. On the other hand, social or Vygotskian constructivism is aimed at social transformation and underscores the socio-cultural context in which the individual or student is situated. It holds that the construction of individual meaning and understanding results from mutually beneficial social or group interactions (primarily through collaboration and negotiation). Vygotsky posited that all learning takes place in the “zone of proximal development,” which he defined as the difference between

what a learner can do alone, and what he or she can do with assistance. Indeed, these two forms of constructivism are not mutually exclusive, as social constructivism is an extension of cognitive constructivism. When used in the teaching-learning context (in particular, reform), constructivism is understood to mean social constructivism.

Instructional design based on constructivism is generally contrasted with instruction based on behaviorism, which is typically described as a rigid procedural approach, aimed at using fixed stimuli and reinforcements to promote a fixed world of objective knowledge, measured primarily in terms of observable behavior (Skinner 1974; Caprio 1994). Instructional design based on behaviorism focuses on discrete and compartmentalized knowledge and skills rather than integration of knowledge, and conceptual understanding. The key difference between these two approaches is that behaviorism is centered around transmission of knowledge from the instructor to the student (passive student and a top-down or instructor-centered approach) whereas constructivism is focused on the construction of knowledge by the student (active student and a bottom-up or student-centered approach). According to Askew et al. (1997) highly effective teachers possess a constructivist (or connectionist) orientation rather than a behaviorist (or transmission) orientation.

In the constructivist context, the instructor utilizes active learning strategies to scaffold activities and tasks (so that students can progress from the simple to the complex), explore information, discover concepts, and construct knowledge and meaning. According to Fosnot (2005, p. 13), in this context, instructors become “facilitators, provocateurs and questioners.” This allows for the development of deep and conceptual understanding, that is, the ability to know “what to do and why” (Skemp 1987, p. 9) rather than surface knowledge (from rote learning associated with behaviorist pedagogy). A key goal in selecting active learning strategies is to facilitate cognitive apprenticeship (Singer and Willett 1993; Dennen 2004) through authentic activities (Leont’ev 1972), encompassing projects, group work (including discussions), problem-solving situations, oral and written presentations, as well as other tasks which model discipline-specific real-world activities, through expert demonstration and guidance (coaching). These activities should be structured and administered so as to provide stimuli for cognitive dissonance or conflict (Liu and Matthews 2005) which serves to promote inquiry, and challenges the individual to think critically and reason, resulting in learning that is deep and conceptual, and hence a meaning-making experience (Dennen, 2004).

Relevant Operational Definitions

The reform-oriented (concept-based or constructivist) approach to teaching introductory statistics is generally operationalized as a set of active learning strategies intended to facilitate statistical literacy. Such active learning strategies include projects, group discussions, data collection, hands-on computer data analysis, critiquing of research articles, report writing, oral presentations, and the use of real-world data. Statistical literacy (thinking and reasoning) refers to the ability to understand, critically evaluate, and use statistical information and data-based arguments (Gal 2000; Garfield et al. 2002). The GAISE (Guidelines for Assessment and Instruction in Statistics Education) report (Franklin and Garfield 2006) which serves as a blueprint for reform-oriented teaching of introductory statistics, recommends the following:

1. Emphasize statistical literacy and develop statistical thinking.
2. Use real data.
3. Stress conceptual understanding rather than mere knowledge of procedures.
4. Foster active learning in the classroom.
5. Use technology to develop conceptual understanding and analyze data.
6. Use assessments to improve and evaluate student learning.

Methodology

Study Design and Participants

The development of this teaching-practice scale was one component of an initial exploratory cross-sectional study which concurrently developed and validated an attitude scale for instructors of undergraduate introductory statistics (see Appendix). Qualitative methods (in-depth interviews and focus group discussions) were also employed, especially for item generation, item analysis, and in general, for establishing content validity. The study participants were 227 volunteer instructors of introductory statistics courses in the health and behavioral sciences at four-year regionally accredited, degree-granting institutions in the USA (and the equivalent in foreign countries¹). Both full-time and adjunct (part-time) instructors who had full responsibility for an introductory statistics course were eligible to participate. The ASA/MAA (American Statistical Association/Mathematical Association of America) Joint Committee on

¹ This includes accreditation by a Ministry of Education or other (Governmental) Higher Education Quality Assurance Body.

Undergraduate Statistics was consulted during the initial phase of the study regarding development of the study methodology².

Sampling

A purposive sample ($n = 227$) was used to reflect the broad range of instructors that the final measure is intended to be used on (Patton 1990), and to allow for meaningful statistical analysis (Sackett et al. 2000). Also, the generally recommended sample size of at least 200, deemed acceptable for scale development (including stability and replicability of structural analyses) was met (Gorsuch 1983; Floyd and Widaman 1995; Clark and Watson 1995). Purposive sampling has been widely used in major studies to explore teachers' beliefs, attitudes and practices in school reform situations (Goertz et al. 1996; Ravitz et al. 2000; Tschannen-Moran et al. 2000). Specifically, this sampling approach helps to guard against a restricted range in measurement, which can result in attenuated correlations among items (Gorsuch 1983; Comrey and Lee 1992; Tucker and MacCallum 1997; Fabrigar et al. 1999). Furthermore, it must be recognized that this was an initial exploratory study, and therefore, purposive sampling was desirable in order to "maximize discovery of the heterogeneous patterns and problems that occur in the particular context under study" (Erlandson et al. 1993, p. 82). Moreover, according to Viswanathan (2005, p. 70) "convenience sampling is suited for these studies rather than probabilistic sampling because the aim is not to establish population estimates, but rather to use correlational analysis to examine relationships between items and measures."

Respondents' Background Characteristics

Of the 227³ participants, 222 provided country information: 165 (74%) were from the U.S., and 57 (26%) were from international locations (primarily the UK, Netherlands, Canada, and Australia). In all, the participants represented 24 countries and 133 academic institutions. The median age category was 41–50 years, and median duration of teaching was 10 years. The majority (139 or 61%) of participants were male, and from the U.S. sub-sample, 135 (82%) identified as Caucasian. There were 94 (41%) instructors from the health sciences, 102 (45%) from the behavioral sciences, and 31(14%) who taught both in the health and behavioral sciences. The modal category for academic degree concentration was statistics, 92 (41%), followed by psychology/social/behavioral sciences, 71(31%), health sciences/public health/epidemiology/biostatistics, 28 (12%), education/business/operations research, 19(8%), and mathematics/engineering, 17 (8%).

²<http://www.austincc.edu/statcomm/m010804.pdf> (Minutes of the Committee Meeting) accessed December 15, 2010.

³ n varies between 219 and 227 due to missing data (item non-response).

Development of the Teaching-Practice (Scale) Items

Teaching practice was conceptualized as a continuum, that is, high-reform (concept-based or constructivist) practice to low-reform (traditional or behaviorist) practice. Development of the scale content was guided by the seminal recommendations of the ASA/MAA Joint Committee on Undergraduate Statistics (Cobb, 1992) and the GAISE report on introductory statistics (Franklin and Garfield 2006). Active learning strategies with reference to course content, pedagogy, assessment, and integration of technology are emphasized in these reports.

Toward content validity, the initial pool of teaching-practice items was formulated following an extensive literature review. Also, a teaching-practice inventory (Handal 2003) which profiles teachers of mathematics as either constructivist or behaviorist was used as a guide. In keeping with other best practices in scale development (Nunnally and Bernstein 1994; Haynes et al. 1995), the initial list of 14 teaching-practice items (with iterations of up to 30 items) was evaluated for breadth and relevance of content in a mini-survey of pioneer and expert statistics educators, researchers, and practitioners. Two qualitative item-analysis exercises followed, one in-person, and the other via email; these involved instructors from the disciplines of health sciences, psychology, education and statistics (including psychometrics), as well as language and communications. In general, items were added, rephrased, or removed based on consensus. All items were assessed for face and content validity, salience, readability (including double-barreled or ambiguous items), theoretical relevance, and redundancy.

A pilot survey was then conducted via email with a sample of 30 instructors,⁴ and the resulting data (including open-ended feedback) allowed for refining the item set. At this stage, responses were evaluated for variability, in particular, the potential discriminant value of each item. The revised item set was reviewed via email by a group of introductory statistics educators, before it was finalized with ten items on a frequency of use scale of 1 (never) through 5 (always). Also, in order to explore variability in teaching practice, the questionnaire ascertained information on gender, age, ethnicity, country, duration of teaching, teaching area, membership status in professional organizations, highest academic degree concentration, and employment status. In general, items intended to measure constructivist and behaviorist teaching orientations were alternated so as to reduce possible acquiescent-response bias. Supposedly, the level of education of the participants was also beneficial in this regard.

⁴ These instructors were not part of the final sample.

Recruitment of Subjects and Data Collection

The general goal of recruitment was to enlist a sample of instructors that represents the broad range of teaching practices. The questionnaire was programmed in Hypertext Markup Language (HTML), and three emails (an invitation to participate, a reminder, and a last call to participate) were sent one week apart with an online link to the questionnaire. Informed consent was obtained online, and data collection took place between August and October of 2005. The completed questionnaires were checked for redundant or duplicate submissions, and, as an incentive to participate, all subjects who completed the instrument were given a chance to win one of three \$100 awards toward conference registration, journal subscription, continuing education courses or other professional development activities. All data were self-reported.

Specifically, recruitment was aimed at maximizing similarities and differences with respect to teaching practice. It involved targeting instructors of introductory statistics in the health and behavioral sciences at four-year colleges where pioneer educators active in the reform movement were employed. This approach was assumed to increase the likelihood of recruiting instructors who have adopted or are moving toward reform recommendations (that is, constructivist pedagogy). Such strategy for identifying institutions was used by Riel and Becker (2000) in their study of teacher leaders' beliefs and practices regarding computer use. Pioneer educators were identified from the membership database of the ASA (Sections on Statistics Education, and Teaching of Statistics in the Health Sciences). Additionally, the ASA Directory of Minority Statisticians was consulted.

Instructors were also targeted based on their publications, research interests, and a review of their course outlines. Faculty who could be characterized as having a mathematical/traditional/behaviorist focus were equally targeted. Additional contact information was obtained from the following sources:

1. Articles in the *Journal of Statistics Education* (online).
2. The first USCOTS (United States Conference On Teaching Statistics) resource notebook.
3. JSM (Joint Statistical Meetings, American Statistical Association) conference proceedings.
4. Online institutional faculty lists.
5. ICOTS (International Conference On Teaching Statistics) conference proceedings.
6. ASA/MAA Joint Committee on Undergraduate Statistics.
7. The Stanford University online directory of statistics departments (USA and international).
8. The ASA list of schools offering degrees in statistics.

9. CAUSE (Consortium for the Advancement of Undergraduate Statistics Education).
10. The RSS (Royal Statistical Society, UK), Center for Statistics Education.
11. IASE (The International Association for Statistical Education).

Additionally, Departmental Chairs were contacted and asked to circulate the questionnaire link to relevant faculty. Departments included statistics, mathematics, health sciences, biostatistics, public health, epidemiology, psychology, behavioral sciences, social sciences, and sociology. To further supplement the sample, participation was solicited from instructors of introductory statistics who participated in a preliminary mini email survey (Hassad 2003) and self-characterized their teaching as either concept-based or calculation-based. Participation was also sought via online discussion forums, and listservs such as:

1. ALLSTAT@JISCMail.AC.UK - A UK-based worldwide email broadcast system for the statistical community.
2. TEACHING-STATISTICS@JISCMail.AC.UK - A UK-based worldwide email broadcast system, concerned with the initial learning and teaching of statistics.
3. SRMSNET@LISTSERV.UMD.EDU - A mailing list of the Survey Research Methods Section of the ASA (American Statistical Association).
4. EDSTAT-L@LISTS.PSU.EDU - An email forum devoted to discussion of topics related to the teaching and learning of statistics at the college level.

Data Analysis

Multidimensional Scaling (MDS)

The underlying dimensionality of the teaching-practice data was examined using primarily multidimensional scaling (MDS) techniques. MDS is a set of exploratory multivariate statistical techniques aimed at reducing and organizing data so as to elucidate how and why the measured variables or items are related (Kruskal and Wish 1978; Coxon 1982). The aim of MDS is to achieve a low-dimensional spatial representation (geometric map or configuration) of the latent or hidden structure that underlies and explains the relationships among the measured variables or items (Kruskal and Wish 1978; Coxon 1982; Fitzgerald and Hubert 1987). Such graphical representations or spatial maps (not obtained with factor analysis) can be formative and intuitive toward identifying possible dimensions underpinning reported perceptions and behaviors (Kruskal and Wish

1978; Coxon 1982; Jaworska and Chupetlovska-Anastasova 2009). Therefore, compared to factor analysis, MDS can result in a more interpretable, plausible and parsimonious latent structure or model (Kruskal and Wish 1978; Fitzgerald and Hubert 1987; Jaworska and Chupetlovska-Anastasova 2009).

Furthermore, unlike factor analysis which requires the assumptions of metric data, linear relationships, and multivariate normality, “MDS procedures can be used on a wide variety of data, using different models and allowing different assumptions about the level of measurement” (Coxon 2004, p. 1). Specifically, there are metric (linear transformation) and non-metric (ordinal transformation) variants of MDS. The teaching-practice data of this study are ordinal (obtained on a Likert-type scale) and hence non-metric; therefore, MDS is suitable for this study. The input information for MDS is a numerical measure of distance indicating how similar (or dissimilar) each item is to every other item. Both metric (MRSCAL⁵) and non-metric (MINISSA⁶) MDS were carried out using the NewMDSX program (Coxon et al. 2010), and for each model both Pearson’s product-moment correlation coefficient (based on the interval properties of the data) and Kendall’s tau (based on the rank order of the data) were used as measures of similarity. By accommodating other measures of similarity or association (including Spearman’s rho and Kendall’s tau), and allowing for the modeling of non-linear relationships, MDS can facilitate a more comprehensive exploration of the data, compared to factor analysis (Coxon 1982).

It is worth emphasizing that this is an initial exploratory study, and hence there was no firm *a priori* model specification (for teaching practice, in this context) to test; therefore, methods such as confirmatory factor analysis (CFA) were not warranted at this stage. As Byrne (2001, p. 99) stated, “the application of CFA procedures to assessment instruments that are still in the initial stages of development represents a serious misuse of this analytic strategy.”

Interpretation of the MDS Maps (Configurations)

Interpretation involved identifying and assigning meanings to patterns and regions (clusters of items or variables), and this is referred to as the neighborhood approach (as against the dimensional approach which is based on multiple regression). In particular, the neighborhood approach to interpretation of the maps can uncover other (and more meaningful) patterns in the data because its focus “is primarily on the small distances (large similarities), while a dimensional approach attends most to the large distances” (Kruskal and Wish 1978, p. 44). According to Kruskal and Wish (1978, p. 58), for the neighborhood approach “a two-dimensional configuration is far more useful than one involving three or more

⁵MRSCAL: MetRic SCALing (Roskam 1972)

⁶MINISSA: Michigan-Israel-Netherlands-Integrated Smallest Space Analysis (Roskam and Lingoes 1970)

dimensions.” However, other configurations were explored. For all MDS procedures, the behaviorist items were reverse-coded to obtain meaningful scores. Also, cluster analysis was used to guide the identification of patterns within the spatial maps (Coxon 1982).

All MDS maps were rotated to simple structure to achieve interpretability (Kruskal and Wish 1978). Following rotation,⁷ the projections of points on the axes change, but the distances between points or items (which the configuration is based on) do not change (Kruskal and Wish 1978). For ordinary MDS, the axes are arbitrary from the interpretation perspective (Coxon 1982). Rotation allows for delineating possible meaningful clusters representing separable structures or dimensions underlying the measured variables or items.

MDS plots the objects (each item) on a map, placing similar objects close to each other and dissimilar objects further apart (Coxon 1982; Kruskal and Wish 1978; Young 1987). In metric MDS (using the MRSCAL algorithm) the Euclidean distance between any pair of items is linearly related to the corresponding input proximity value (correlation coefficient). This involves a linear transformation of the data, and the relationship is inverse, as the correlations represent similarities. That is, higher correlation coefficients (Pearson’s and Kendall’s) would indicate greater similarity between the items, and hence they will be closer in the spatial map (that is, a shorter inter-item distance). For non-metric MDS (using the MINISSA algorithm) the Euclidean distance between any pair of items in the spatial map matches the rank-order of the corresponding input proximity data, and hence a monotonic, ordinal or non-metric transformation of the data is carried out (Borg and Groenen 1997).

The adequacy of the MDS solutions was evaluated using the level of Stress1 and the coefficient of determination (R^2). Both values are measures of goodness of fit between the input data and the MDS model (Coxon 1982). The stability of the solutions was assessed using the guideline of at least $4k + 1$ objects (items) for a k -dimensional solution, as well as consistency across the MDS configurations (Kruskal and Wish 1978). In general, all MDS solutions were evaluated with regard to adequacy, interpretability, parsimony, plausibility, and construct validity.

Reliability Analysis

Cronbach’s alpha (Cronbach 1951) which quantifies the degree of internal consistency (reliability) of a set of items, was calculated for each subscale, as well as the overall scale. In general, a Cronbach’s alpha of at least .7 is viewed as the minimum acceptable level of reliability (Nunnally 1978); however, a prior recommendation that “in the early stages of research ... reliabilities of .60 or .50

⁷ Rotation does not alter the solution in the Euclidean distance model.

will suffice” was also considered (Nunnally 1967, p. 226), as this is an initial exploratory study. Furthermore, based on its mathematical underpinning, Cronbach’s alpha varies directly with the number of items and the mean inter-item correlation (Streiner and Norman 1989), so any interpretation of alpha must take into account these two parameters. Loewenthal (1996) suggests that a reliability level of .6 may be considered acceptable for scales with less than ten items. Also, dimensionality must be given key consideration as Cronbach’s alpha is an underestimate of reliability if the scale is not unidimensional (Cronbach 1947, 1951; Schmitt 1996).

Consequently, the descriptive statistics (mean, standard deviation, and range) of the distribution of the inter-item correlations were examined to assess both internal consistency and homogeneity. For homogeneity (unidimensionality), a recommended optimal mean inter-item correlation between .2 and .4 (with almost all of the individual inter-item correlations being moderate and in the range of .15 to .5) was the standard used (Briggs and Cheek 1986; Clark and Watson 1995). Indeed, higher inter-item correlations would suggest redundancy in the contents of the scale. Also, in order to establish the contribution of individual items to each subscale, the change in Cronbach’s alpha was noted following the deletion of each item. Furthermore, according to Nunnally and Bernstein (1994), the corrected⁸ item-total correlations should be at least .3 in order for that item to be considered a meaningful contribution to the scale, and this criterion was followed. Additionally, inter-subscale correlation analysis was performed to give further insight into the interpretability of the subscales. Note that for all relevant analyses, the observed correlation coefficients are presented in light of the controversy surrounding correction for attenuation due to measurement error (Lord and Novick 1968), and given that this is an initial exploratory study.

Criterion (Concurrent) Validity Analysis

Validity is a multidimensional concept (more precisely termed construct validity), and refers to whether the scale measures the construct (teaching practice) as theorized (Cronbach and Meehl 1955). In other words, validity is the extent to which an instrument measures what it is supposed to measure (Carmines and Zeller 1979). Muldoon et al. (1998, p. 543) noted that “while validity can be examined in several ways, comparison with the best indicator available (criterion validity) is the preferred method.” Specifically, “criterion validity assesses the measure’s ability to distinguish between groups that it should theoretically be able to distinguish between” (MacInnes 2003). This is particularly relevant to psychological constructs (e.g., attitude) which are latent and hence not directly

⁸This refers to the correlation between an item and the remaining items in a scale.

observable. Therefore, in order to determine whether a hypothetical construct is being measured, we must show that it relates to a measure of another construct (the criterion) in a theoretically predictable way (Cronbach and Meehl, 1955). Note that criterion validity is also referred to as concurrent validity when both constructs are measured simultaneously.

Accordingly, criterion validity is reported herein. In this regard, the attitude-practice relationship was explored, consistent with the Theory of Reasoned Action (TRA; Fishbein and Ajzen 1975; Ajzen and Fishbein 1980), and the Theory of Planned Behavior (TPB; Ajzen 1991), and with the expectation of a meaningful and statistically significant relationship between attitude score and practice score (recognizing that this could be a bi-directional relationship). Instructors' attitude toward reform-based teaching was measured concurrently with teaching practice, and the Pearson's product-moment correlation coefficient (r) was used to assess this relationship. Additionally, standard multiple linear regression analysis was performed to determine the extent to which the five attitude components (the independent variables) can explain and predict teaching practice (the dependent variable) in this context (Meehl 1954; Asher 1997), and the relative contribution of each predictor to the model. An alpha level of .05 was used for all tests of significance. Also, where applicable, assumptions underlying statistical methods were checked, and post-hoc analyses (with Bonferroni correction) were performed. SPSS version 18.0 was also used for data entry and analysis.

Calculation of the Teaching-Practice and Attitude Scores

A composite teaching-practice score was obtained for each respondent by summing the values of the ten practice items (Table 1). The distribution has a mean ($SD = 5$) and median of 35, and a mode of 36, with a possible maximum score of 50. Note that higher scores reflect more favorable reform-based practice. The behaviorist items were reverse-coded for calculation of the composite teaching-practice score.

Instructors' attitude toward reform-oriented teaching was measured using the Faculty Attitudes Toward Statistics (FATS©) scale (see Appendix), which consists of 25 items. A composite attitude score was obtained for each respondent by summing the values of the 25 items. The distribution has a mean of 99 ($SD = 12.6$) and a median and mode of 100, with a possible maximum score of 125. Note that higher scores reflect a more favorable attitude toward reform-based practice. In calculating the composite scores for attitude and practice, all items were equally weighted (Russel 2002; Hogue et al. 2005).

Results

Multidimensional Scaling (MDS) of the Teaching-Practice Items

Of the two- and three-dimensional MDS configurations produced by the metric (MRSCAL) and non-metric (MINISSA) analyses, the two-dimensional maps were the most meaningful and interpretable (following rotation to simple structure), and the best fit was obtained with non-metric MDS using Pearson's correlation coefficient (r) as the input measure of similarity (Figure 1). This map separates the ten teaching-practice items (Table 1) into two distinct clusters (of five items each), which are labeled behaviorist and constructivist in the figure and table. Teaching practice was modeled and items generated, based on these two domains—behaviorist and constructivist—which structure was confirmed by this MDS analysis.

Table 1
Teaching of Introductory Statistics Scale

| Teaching-Practice Items | Never | Rarely | Some-times | Usually | Always | N | *Mean(SD) |
|--|-------|--------|------------|---------|--------|-----|------------|
| (1) I emphasize rules and formulas as a basis for subsequent learning. B | 5 | 4 | 3 | 2 | 1 | 227 | 2.73(0.82) |
| (2) I integrate statistics with other subjects. C | 1 | 2 | 3 | 4 | 5 | 226 | 3.77(0.95) |
| (3) Students use a computer program to explore and analyze data. C | 1 | 2 | 3 | 4 | 5 | 226 | 4.02(1.08) |
| (4) I assign homework primarily from the textbook. B | 5 | 4 | 3 | 2 | 1 | 226 | 2.87(1.13) |
| (5) Critiquing of research articles is a core learning activity. C | 1 | 2 | 3 | 4 | 5 | 227 | 2.98(1.15) |
| (6) The mathematical underpinning of each statistical test is emphasized. B | 5 | 4 | 3 | 2 | 1 | 227 | 2.78(0.95) |
| (7) I use real-life data for class demonstrations and assignments. C | 1 | 2 | 3 | 4 | 5 | 226 | 4.06(0.75) |
| (8) I require that students adhere to procedures in the textbook. B | 5 | 4 | 3 | 2 | 1 | 225 | 2.77(0.93) |
| (9) Assessment includes written reports of data analysis. C | 1 | 2 | 3 | 4 | 5 | 225 | 3.59(1.12) |
| (10) I assign drill and practice exercises (mathematical) for each topic. B | 5 | 4 | 3 | 2 | 1 | 227 | 2.67(1.07) |

(B) Behaviorist subscale items. These items must be reverse-coded (as shown here) for calculation of the overall teaching-practice score, so that higher values indicate more favorable levels of reform-oriented (concept-based or constructivist) practice. **(C) Constructivist subscale items.** *For the Mean and SD presented here, these items were not reverse-coded.

The solution configuration of Figure 1 fits the input data very well, with a Stress⁹ value (residual sum of squares from monotonic regression) of .07. Stress1 values closer to zero represent a better fit. In this case, the Stress1 value is less than one-half the size of Stress1 from simulation studies of random configurations of the same number of points and dimensions (Spence 1979). The proportion of variance in the input data that is accounted for by this two-dimensional solution is 89% (R^2), suggesting a very good fit. Additionally, the stability of the solution is in keeping with the empirical guideline of at least $4k + 1$ objects for a k -dimensional solution with non-metric scaling (Kruskal and Wish 1978). In this context, the objects are the ten teaching-practice items, and $k = 2$, as this is a two-dimensional solution.

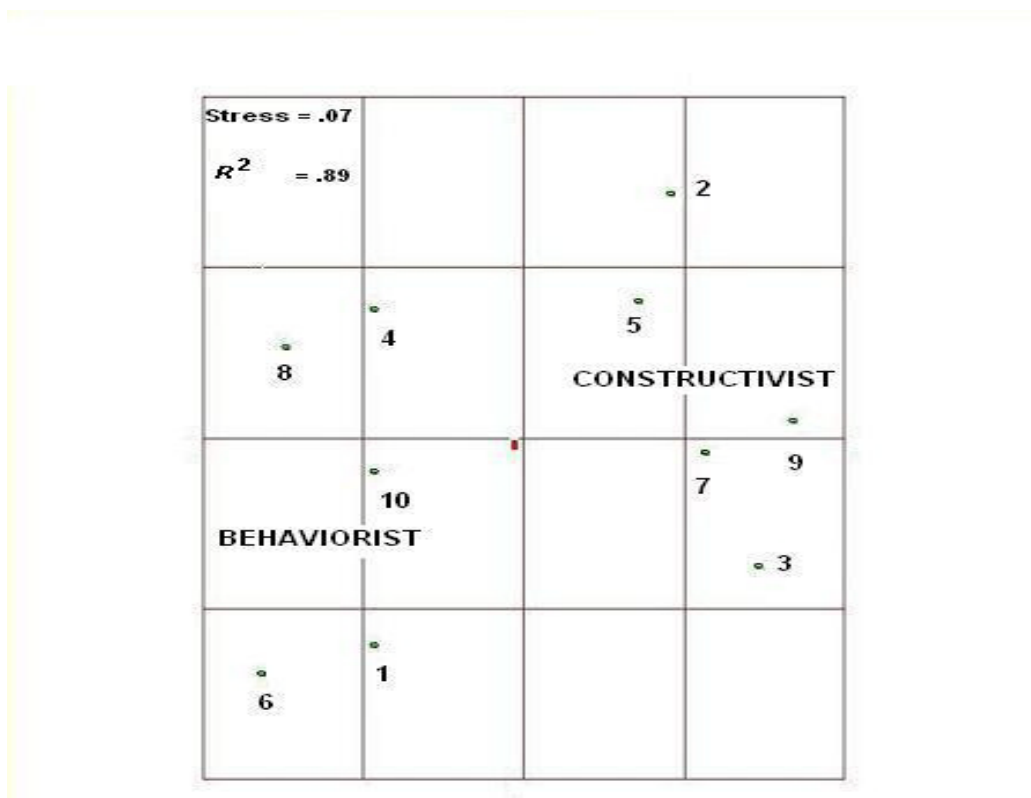


Figure 1. Two-dimensional MDS map (non-metric). Numbers refer to teaching-practice items (see Table 1).

⁹Stress1 is a form of raw Stress, normalized for symmetric data.

Reliability Analysis

The Cronbach's alpha of the overall scale is .6 (Table 2), which (albeit less than the standard minimum of .7) is considered an acceptable level for initial exploratory studies (Nunnally 1967; Robinson et al. 1991). Furthermore, each subscale (behaviorist: alpha = .61, constructivist: alpha = .66) is somewhat more internally consistent than the overall scale, and this could support the MDS finding that two dimensions underlie teaching practice (Yu 2001). Notably, each subscale has five items and according to Loewenthal (1996), a reliability level of .6 may be considered acceptable for scales with less than ten items.

Table 2
Cronbach's Alpha and Mean Correlation Coefficients for the Overall Teaching-Practice Scale and Subscales

| Scale/subscale | N | Number of items | Cronbach's alpha | Mean Correlation (SD) | |
|--|-----|-----------------|------------------|-----------------------|------------|
| | | | | Inter-item | Item-scale |
| Overall Teaching-Practice Scale (TISS) | 219 | 10 | .60 | .13 (.15) | .25 (.11) |
| Behaviorist Subscale | 224 | 5 | .61 | .25 (.10)* | .37 (.06) |
| Constructivist Subscale | 222 | 5 | .66 | .29 (.10)** | .42 (.08) |

Inter-subscale correlation: Pearson's $r = -.06$, $df = 217$, ns

*Ranged from .06 to .39 (with 9 (90%) of the correlations being between .14 and .39, and 70% between .23 and .39)

**Ranged from .12 to .50 (with 9 (90%) of the correlations being between .19 and .50)

More importantly, the mean inter-item correlation for the overall scale (.13) is substantially less than the mean inter-item correlations for the behaviorist (.25) and constructivist (.29) subscales. In particular, almost all of the inter-item correlations for both subscales are moderate in magnitude and cluster narrowly around the mean value, and together this can be considered strong evidence of two separate and distinct subscales with acceptable internal consistency (Green 1978; Clark and Watson 1995). Additionally, deletion of any item did not appreciably improve the reliability of the subscales, and the corrected item-total correlations are .3 or higher, indicating that each item is meaningful to its subscale (Nunnally and Bernstein 1994). Furthermore, the two subscales (behaviorist and constructivist) are almost orthogonal to each other (Pearson's product-moment correlation coefficient, $r = -.06$, $df = 217$, ns), and hence can be considered independent dimensions of teaching practice.

Criterion (Concurrent) Validity

According to attitude theory, and the attitude-behavior relationship (Ajzen and Fishbein 2004; Wallace et al. 2005; Schwartz 2007), high-reform instructors (higher practice scores) should possess more favorable attitudes (higher attitude scores) toward reform-based pedagogy. As shown in Table 3, the relationships are

in the expected (positive) direction, and all but “perceived difficulty” (ease of use) were moderate, meaningful, and statistically significant. Indeed, these relationships could be viewed as bi-directional.

Table 3
Bivariate Correlation between Total Attitude (and Subscale Scores) and Teaching-Practice Score

| Attitude Subscales | N | Teaching- Practice Score |
|------------------------------------|-----|-----------------------------|
| | | Pearson's <i>r</i> |
| Perceived Usefulness | 219 | .364* |
| Intention | 218 | .452* |
| Personal Teaching Efficacy | 217 | .421* |
| Avoidance-Approach | 218 | .387* |
| Perceived Difficulty (Ease of Use) | 218 | .073 |
| Total Attitude Score | 214 | .498* |

* $p < .001$

Multiple regression analysis (Meehl 1954; Asher 1997) was next conducted. Although perceived difficulty was not statistically significant in the bivariate analyses (Table 3), it was entered into the multiple regression model because of its noted conceptual (and empirical) relevance to both intention and behavior (Albarracin et al. 2001). Teaching-practice score (the dependent variable) was regressed on the five attitude subscale scores (the independent variables). The standard assumptions for multiple regression were met: normality, linearity, non-multicollinearity, homoscedasticity, and independence of residuals. The strongest correlation among the predictors was noted for perceived usefulness and intention ($r = .7$, $df = 223$, $p < .001$). Also, perceived difficulty (ease of use) was significantly correlated with personal teaching efficacy ($r = .4$, $df = 222$, $p < .001$) only.

The overall model (Table 4) was statistically significant, and explained 28% of the variance in teaching practice (see adjusted R^2), which is consistent with major attitude-behavior research (Armitage and Conner 1999). Intention (one component of attitude) was the strongest predictor of practice ($\beta = 0.264$, $p = .003$), a finding that is both theoretically and empirically well-supported (Armitage and Conner 1999; Ajzen and Fishbein 2004; Wallace et al. 2005). Notably, perceived usefulness was not statistically significant in the multiple regression model, and as previously mentioned, this construct had a strong and statistically significant relationship with intention ($r = .7$), but did not meet the generally acceptable statistical criterion (of $r = .9$) for redundancy (Kline 2005). Also, from a theoretical and empirical perspective, these two constructs (perceived usefulness and intention) are known to be strongly related but different

(Ajzen and Fishbein 1977; Davis et al. 1989; Taylor and Todd 1995; Kloeblen and Batish 1999; Venkatesh and Davis 2000).

Table 4
Multiple Regression Analysis of Overall Teaching-Practice Score on Attitude Subscale Scores

| Predictors (Attitude Subscales) | Unstandardized Coefficients | | Standardized Coefficients | | |
|------------------------------------|-----------------------------|------------|---------------------------|----------|------|
| | <i>B</i> | Std. Error | Beta (β) | <i>t</i> | Sig. |
| (Constant) | 17.05 | 2.30 | | 7.42 | .001 |
| Perceived Usefulness | 0.04 | 0.61 | 0.01 | 0.07 | .946 |
| Intention | 1.58 | 0.52 | 0.26 | 3.01 | .003 |
| Personal Teaching Efficacy | 1.62 | 0.54 | 0.24 | 3.01 | .003 |
| Avoidance-Approach | 1.46 | 0.48 | 0.20 | 3.07 | .002 |
| Perceived Difficulty | -0.42 | 0.35 | -0.08 | -1.22 | .225 |

Model Significance: $F(5, 208) = 17.3$, $p < .001$, Adjusted $R^2 = .28$ (28%)

Correlates of Teaching Practice

Overall teaching-practice, as well as the constructivist and behaviorist subscale scores did not vary significantly with respect to gender, age, ethnicity, duration of teaching, teaching area, and employment status. However, statistically significant differences were noted as follows. For these analyses, the independent samples *t*-test and one-way analysis of variance (ANOVA) were used. Note that these differences and any suggested trend are tentative (given the non-probability nature of the sample), and should not be generalized, but used to inform hypothesis generation.

- (1) **Location:** Instructors from international locations¹⁰ ($M = 36.80$, $SD = 4.44$) reported a higher level of reform-based teaching compared to those from the U.S. ($M = 34.03$, $SD = 4.57$), $t(212) = 3.88$, $p = .001$. In particular, instructors from international locations ($M = 12.13$, $SD = 3.27$), reported a lower level of behaviorist practice than those from the U.S. ($M = 14.32$, $SD = 2.85$), $t(217) = 4.76$, $p = .001$.
- (2) **Highest Academic Degree Concentration:** In general, instructors with mathematics and engineering degrees ($M = 32.81$, $SD = 4.28$) reported the lowest level of reform-based teaching compared to those with health sciences degrees ($M = 36.00$, $SD = 4.96$), who reported the highest level, $t(41) = 2.14$, $p = .038$. Specifically, instructors with mathematics and engineering degrees ($M = 15.23$, $SD = 2.88$) reported the highest level of behaviorist practice compared to those with health sciences degrees ($M = 12.39$, $SD = 3.22$) who reported the lowest level, $t(43) = 2.98$, $p = .005$.

¹⁰ These instructors were primarily from the UK, Netherlands, Canada, and Australia.

- (3) **Membership Status in Professional Organizations:** In general, instructors with membership ($M = 35.33$, $SD = 4.92$) reported a marginally higher level of reform-based teaching compared to non-members ($M = 34.07$, $SD = 4.43$), $t(214) = 1.99$, $p = .048$.

Discussion

The objective of this study was to develop and initially validate a scale (instrument) to characterize the teaching-practice orientation of instructors of introductory statistics in the health and behavioral sciences at the college level. Given the totality of the evidence, in accordance with scientific standards for scale development (Clark and Watson 1995; DeVillis 2003), a scale with acceptable levels of reliability and validity has been developed and will be referred to as TISS (Teaching of Introductory Statistics Scale).

The Structure of the Teaching-Practice Scale

The TISS is two-dimensional, and the two components or subscales (constructivist and behaviorist) are orthogonal to each other, indicating that two independent or separable (but complementary) dimensions underlie teaching practice, in this context. In other words, contrary to the initial conceptualization and general view, teaching practice appears not to be bipolar or reciprocal. That is, a higher level of constructivist practice does not result in a lower level of behaviorist practice, and vice versa. And knowing one dimension does not meaningfully inform us about the other. Therefore, both subscales must be used in order to meaningfully measure and address teaching practice. Similar findings of an orthogonal relationship between behaviorist and constructivist dimensions with regard to teachers' beliefs and practices (elementary and secondary school) were reported by Handal (2003) and Woolley et al. (2004), with correlation coefficients of $-.232$ and $-.011$, respectively.

This orthogonal two-dimensional construct or scale (TISS) is substantive, and can be plausibly explained in terms of two possible underlying motivational (and perceptual) processes regarding constructivist and behaviorist teaching practices, each with different antecedents and likely complex interactions (Cacioppo et al. 1997). This could mean, that in addition to pedagogical content knowledge (Shulman 1986; Hassad; 2006), the decision to utilize a particular set of teaching strategies is influenced by instructors' perceptions about contextual factors, such as the level of preparedness of students, duration of the session, availability and accessibility of teaching resources, class size and heterogeneity (including variability in students' ability and academic major), as well as administrative support. Logically and empirically, beliefs relating to such contextual factors can facilitate or inhibit the implementation of reform-based practices, and hence an

instructor may vary in his/her pedagogical approach. In particular, these beliefs can influence an instructor's perception of self-efficacy and affect decision-making regarding the use of particular teaching strategies. Another important factor in this process is perceived usefulness (of pedagogical strategies), especially given its strong and significant relationship with intention, which emerged in this study, as the strongest predictor of teaching practice. Perceptions (or beliefs) about self-efficacy and usefulness are modifiable; they should be a focus of professional development programs aimed at facilitating the adoption and use of reform-based pedagogical practices.

The Use of MDS (Multidimensional Scaling)

The plausible two-dimensional structure underlying this teaching-practice scale emerged from non-metric (monotonic or ordinal) multidimensional scaling (MDS), with Pearson's correlation coefficient (r) as the input measure of proximity. The non-metric MDS algorithm transforms the proximity data (r) into quasi-Euclidean distances, which preserve the rank-order of the proximity data (Coxon 1982; Borg and Groenen 1997). Hence an ordinal transformation of the data is carried out. These distances are then represented in a low-dimensional geometric space as a map or configuration, resulting in solutions that tend to be more parsimonious, interpretable, and meaningful than those obtained from standard Factor Analysis (FA). Also, if Factor Analysis proceeds from measures such as Pearson's r rather than the original scores, then a non-metric analysis cannot be performed due to the problem of communalities, and Kruskal and Shepard (1974) have moreover shown that the non-linearity permitted by the monotone function can actually give a better solution.

Indeed, given the ordinal nature of these data (obtained from Likert-type scales), as well as our understanding of psychological processes and behavioral phenomena (as being largely non-linear), a statistical model other than a non-metric variant could be deemed tenuous. Accordingly, the use of non-metric MDS (with Pearson's r) is a more conservative data analytic approach than any of the standard forms of FA. This statistical methodology is therefore a key strength of this study, and the size ($n = 227$) and type of sample (purposive) can help to assure replicability and stability of this scale structure. Further research could utilize INDSCAL (Individual Differences Scaling; Carroll and Chang 1970), a three-way weighted MDS model to determine subgroup differences with regard to the salience (weights) attributed to the dimensions underlying teaching practice (Coxon 1982). Subgroup variability in terms of underlying dimensionality of teaching practice should also be explored.

Psychometric Properties of the Teaching-Practice Scale

Reliability. Critical to the use of any instrument or scale are its psychometric properties, namely reliability and validity, and based on the evidence from this study, acceptable levels of both properties have been obtained. While the Cronbach's alphas for the overall scale (.6) and the subscales (behaviorist = .61, constructivist = .66) are less than the usual minimum (albeit sometimes mechanical) standard of .7, other criteria are more pertinent to the evaluation of internal consistency (reliability), in this context. This is particularly so, given that the (overall) scale is not unidimensional, that both subscales consist of just five items each, and that this is an initial exploratory study. These additional criteria, which are based on the mean inter-item correlations, as well as the range of the individual inter-item correlations (Briggs and Cheek 1986; Clark and Watson 1995), were adequately met, and are presented and explained above (see Table 2). Furthermore, the utility of this scale is strengthened by the comprehensive and scientific methodology that was used for content coverage, and the resulting meaningful items. In this regard, national and international experts (in statistics education and psychometrics) served as content specialists. It is also worth noting that it is not uncommon to encounter major studies (Clarke and Watson 1995; Leonhard et al. 2000; Cruise et al. 2006; Kahveci 2009) that have invoked and used Nunnally's (1967, p. 226) suggestion that reliability values of .5 or .6 are acceptable in the early stages of research, as well as Loewenthal's (1996) recommendation that a reliability level of .6 may be considered acceptable for scales with fewer than 10 items.

Validity. Undoubtedly, the greatest strength of this study is in demonstrating acceptable construct (specifically, criterion) validity of the newly developed scale (TISS). Toward this end, the attitude-practice relationship was explored based on theoretical understanding (particularly, the TRA and TPB), and empirical data that attitudes are predispositions to act favorably or unfavorably, and hence are necessary to predict and explain behavior (practice). As shown in Table 3, the correlation coefficients for the bivariate relationships between total attitude (and subscale) scores and teaching-practice score are moderate to large (Cohen 1988), and consistent with major attitude-behavior research (Ajzen and Fishbein 2004; Schwartz 2007). Moreover, the five attitude subscales together accounted for 28% (R^2) of the variance in teaching practice (see Table 4) which is among the higher values of R^2 that have been reported for attitude-behavior research premised on the TRA and TPB (Armitage and Conner 1999; Ajzen and Fishbein 2004). Additionally, the R^2 value of 28%, obtained in this study, parallels the average R^2 of 27% from a major authoritative meta-analysis of 185 independent attitude-behavior studies guided by the TPB model (Armitage and Conner 1999; Armitage and Christian 2003; Ajzen and Fishbein 2004).

Another finding that enhances the construct validity of this scale is the emergence of intention as the strongest predictor of teaching practice, which is theoretically and empirically well established in attitude-behavior research (Fishbein and Ajzen 1975; Ajzen 1991; Armitage and Conner 1999; Ajzen and Fishbein 2004). In general, the validity coefficients (r and R^2) obtained in this study are relatively substantial in the attitude-behavior context, and in this regard, one aspect of this study methodology warrants emphasis. According to Ajzen and Fishbein (1977) attitude is a stronger predictor of behavior when the measures of attitude and behavior correspond in specificity, regarding target, action, context and time. This principle was a key design feature of this study, in that attitude and practice items were concurrently measured, and framed with reference to the teaching of introductory statistics in the health and behavioral sciences at the tertiary level, using the concept-based or reform-oriented pedagogy (considered an innovative teaching approach).

Worthy of note, is the seemingly counter-intuitive observation that two of the attitude subscales, *perceived usefulness* and *perceived difficulty* (ease of use), were not statistically significant independent predictors of teaching practice (Table 4). Similar observations regarding perceived difficulty have been reported by Wallace et al. (2005) and Kaiser and Schultz (2009). Specifically, Wallace et al. (2005) concluded from a major meta-analysis that “one intriguing and unexpected finding” was that perceived difficulty did not moderate the attitude-behavior relationship regarding both self-reported and directly observed behaviors. Nonetheless, both perceived usefulness and perceived difficulty are generally considered conceptually relevant to attitude formation and the prediction of behavior (practice). Therefore, it is prudent to make a distinction between predicting and explaining behavior or practice (Cook and Campbell 1979; Asher 1997), especially recognizing that “a multiple regression equation, although succinct, may suppress useful explanatory relationships” (Asher 1997, p. 7). Accordingly, while these two attitude subscales are not significant predictors of practice (in this context), the results of this analysis should not be interpreted as indicating that perceived usefulness and perceived difficulty are not important in explaining and understanding the theory underlying the overall model (Asher, 1997; Ajzen 2002). The relationship that each of these two subscales has with other variables (which emerged as significant predictors of practice) must be considered, particularly, perceived usefulness and intention ($r = .7$), and perceived difficulty and personal teaching efficacy ($r = .4$). This could suggest the presence of nested, higher-order or hierarchical factors, which should be hypothesized and explored.

While attitude is generally conceptualized as necessary for meaningful explanation and prediction of behavior (or practice), it is clearly not sufficient in this regard. In this study (and based on multiple regression analysis), attitude

toward teaching accounted for 28% (R^2) of the variance in teaching practice, leaving a large proportion of the variance unexplained or unaccounted for. However, such model performance is not uncommon in the behavioral sciences, and this level of R^2 is considered acceptable, particularly for a single predictor construct. For example, Ajzen and Fishbein (2004, p. 433), in commenting on two attitude-behavior models with R^2 values of 21% and 31% (Armitage and Conner 1999), noted that they accounted for “considerable variance.” Nonetheless, other factors, including possible moderating and mediating variables (Armitage and Christian 2003; Schwartz 2007), should be explored toward improving the predictive value of this attitude-practice model.

Correlates of Teaching Practice

Finally, as noted above, the observed variability in teaching practice with regard to location, highest academic degree concentration, and membership status in professional organizations, should be viewed as tentative, given the non-probability nature of the sample, which could have compromised representativeness, and therefore limit external validity or generalizability. Also, the practical significance of these findings must be carefully evaluated with reference to the size of the overall sample and subsamples. Notwithstanding these concerns, these results are interesting and plausible; they hold much potential for informing the statistics education community regarding facilitating and inhibiting factors for reform-based teaching of introductory statistics, and should be further explored with an appropriate sample. Note that the issue of representativeness regarding the correlates of teaching practice does not apply to the teaching-practice scale *per se*. Representativeness in the context of scale development research does not follow conventional wisdom; that is, the goal is not to closely represent any defined population but to ensure that those who are likely to score high and those who are likely to score low are well represented. This was achieved with the purposive sample used in this study (Gorsuch 1997).

Conclusions

This initial exploratory study examined the teaching practices of college instructors of introductory statistics from the health and behavioral sciences. Using primarily multidimensional scaling (MDS) techniques, a two-dimensional, ten-item teaching-practice scale was developed, and acceptable levels of reliability and validity were demonstrated. This scale will be referred to as TISS (Teaching of Introductory Statistics Scale). The two teaching dimensions (subscales) are characterized as constructivist (reform-oriented, student-centered, and active learning), and behaviorist (traditional, instructor-centered, and passive learning); they are orthogonal. Further research will be required in order to be

conclusive about the structural and psychometric properties of this scale. Also, this study examined internal consistency (reliability), and not test-retest reliability, which should be assessed in order to determine the stability of the scale over time. Indeed, this new scale (TISS), will allow us to empirically assess and describe the pedagogical approach (teaching practice) of instructors of introductory statistics in the health and behavioral sciences, at the college level, and determine what learning outcomes result from the different teaching-practice orientations.

Acknowledgments

This study had widespread support from the statistics education community (nationally and internationally). Special thanks to all the respondents from 24 countries, the Administrations of Mercy College and Hunter College, members of the ASA/MAA Joint Committee on Undergraduate Statistics, and to the following, who served as either content experts and/or provided feedback at various stages of this study: Martin Bland (UK), Paul J. Fields (USA), Joan Garfield (USA), Cyndy Long (USA), Anne Porter (Australia), Candace Schau (USA), W. Robert Stephenson (USA), Dirk Tempelaar (the Netherlands), and Jeff Witmer (USA). To Dr. Edith Neumann and Dr. Frank Gomez, thanks for your guidance and support over the years. Also, I extend my sincere appreciation to Gerald Iacullo for his help. I am immensely grateful to Dr. Anthony Coxon for his expert advice on Multidimensional Scaling, and continued mentorship. Finally, I wish to thank the *Numeracy* reviewers and editor assigned to this paper for the helpful feedback and guidance that made possible the publication of this study.

Appendix: The FATS©

The Faculty Attitudes Toward Statistics scale (FATS©) was developed and initially validated concurrently with the teaching-practice scale (Hassad and Coxon 2007; Hassad 2007). The FATS© was developed to measure instructors' attitudes toward concept-based (reform-oriented or constructivist) teaching of introductory statistics in the health and behavioral sciences at the college level. It consists of five subscales with a total of 25 items, and an overall Cronbach's alpha (internal consistency) of .89. Further evidence of the psychometric worthiness of the FATS© is presented by McGowan (2009).

Attitude was conceptualized and defined as an evaluative disposition toward some object based upon cognitions, affective reactions, and behavioral intentions. In other words, attitude is an informed predisposition to respond. According to the tripartite attitude theory, attitude is composed of three dimensions: beliefs, affect (feelings), and a readiness or intent for action (Smith 1947; Rosenberg and

Hovland 1960; Oppenheim 1982; Zimbardo and Leippe 1991; Jaccard and Blanton 2005). The five attitude components or subscales encompass cognition, affect and intentionality, consistent with the tripartite attitude theory, and are detailed in Table A1. Additionally, the FATS© is grounded in the Theory of Reasoned Action (Fishbein and Ajzen 1975; Ajzen and Fishbein 1980), the Theory of Planned Behavior (Ajzen 1991), and the “Stages of Concern” component of the Concerns Based Adoption Model (Hall and Hord 1987), with attention to change and innovation.

Responses are obtained on a five-point Likert-type scale: 1 = strongly disagree, 2 = disagree, 3 = undecided, 4 = agree, and 5 = strongly agree. A composite attitude score is calculated by summing the values of the 25 items, with a possible maximum score of 125. All “negatively” worded items (indicated with an asterisk in the table) must be reverse-coded so that for all items, higher values indicate a more positive attitude toward reform-oriented pedagogy. Note that either *reform-oriented* or *constructivist* can be substituted for concept-based.

| Table A1 |
|---|
| The Faculty Attitudes Toward Statistics (FATS©) Scale |
| <p>Perceived Usefulness: Beliefs about the value, benefit or worth of the concept-based approach to the teaching of introductory statistics (7 items, alpha = .85).</p> <ol style="list-style-type: none"> 1. The concept-based approach to teaching introductory statistics (rather than emphasizing calculations and formulas) makes students better prepared for work. 2. The concept-based approach to teaching introductory statistics (rather than emphasizing calculations and formulas) makes students better prepared for further studies. 3. *Emphasizing concepts and applications in the introductory statistics course (rather than calculations and formulas) is a disservice to our students. 4. *The concept-based approach to teaching introductory statistics is for low achievers only. 5. The concept-based approach to teaching introductory statistics enables students to understand research. 6. I am convinced that the concept-based approach to teaching introductory statistics enhances learning. 7. Teaching introductory statistics using the concept-based approach is likely to be a positive experience for me. |
| <p>Personal Teaching Efficacy: Beliefs about one’s capability to successfully use the concept-based approach to teach introductory statistics (5 items, alpha = .77).</p> <ol style="list-style-type: none"> 1. I will adjust easily to teaching introductory statistics using the concept-based approach. 2. *Concept-based teaching of introductory statistics may be problematic for me. 3. *I do not understand how to organize my introductory statistics course to achieve statistical literacy. 4. *Teaching introductory statistics with emphasis on concepts and their applications (rather than calculations and formulas) may be stressful for me. 5. *I am concerned that using the concept-based approach to teach introductory statistics may result in me being poorly evaluated by my students. |
| <p>Perceived Difficulty: Beliefs about the ease of use, or effort involved in using the concept-based approach to teach introductory statistics (3 items, alpha = .65).</p> <ol style="list-style-type: none"> 1. *Teaching introductory statistics with emphasis on concepts and applications rather than calculations and formulas, can be time consuming. 2. *The preparation required to teach introductory statistics using the concept-based approach is burdensome. 3. *Using active learning strategies (such as projects, group discussions, oral and written presentations) in the introductory statistics course can make classroom management difficult. |

| |
|---|
| <p>Avoidance-Approach: Positive and negative feelings (affect), inclination, proclivity or propensity toward using the concept-based pedagogy to teach introductory statistics (5 items, $\alpha = .69$).</p> <ol style="list-style-type: none"> *I am not comfortable using computer applications to teach introductory statistics. Using computers to teach introductory statistics makes learning fun. *I will avoid using computers in my introductory statistics course. I will incorporate active learning strategies (such as projects, hands-on data analysis, critiquing research articles, and report writing) into my introductory statistics course. *I am hesitant to use computers in my introductory statistics class without the help of a teaching assistant. |
| <p>Behavioral Intention: Likelihood of using the concept-based pedagogy to teach introductory statistics (5 items, $\alpha = .85$).</p> <ol style="list-style-type: none"> I am engaged in the teaching of introductory statistics using the concept-based approach. I am interested in using the concept-based approach to teach introductory statistics. I want to learn more about the concept-based approach to teaching introductory statistics. *Using the concept-based approach to teach introductory statistics is not a priority for me. I plan on teaching introductory statistics according to the concept-based approach. |

References

- Ajzen, I. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2): 179–211. [http://dx.doi.org/10.1016/0749-5978\(91\)90020-T](http://dx.doi.org/10.1016/0749-5978(91)90020-T)
- . 2002. Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology*, 32: 665–683. <http://dx.doi.org/10.1111/j.1559-1816.2002.tb00236.x>
- , and M. Fishbein. 1977. Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84: 888–918. <http://dx.doi.org/10.1037/0033-2909.84.5.888>
- . 1980. *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice Hall.
- . 2004. Questions raised by a reasoned action approach: Comment on Ogden (2003). *Health Psychology*, 23(4): 431–434. <http://dx.doi.org/10.1037/0278-6133.23.4.431>
- Albarracin, D., B.T. Johnson, M. Fishbein, and P.A. Muellerleile. 2001. Theories of reasoned action and planned behavior as models of condom use: A meta-analysis. *Psychology Bulletin*, 127(1): 142–161. <http://dx.doi.org/10.1037/0033-2909.127.1.142>
- Armitage, C. J., and J. Christian. 2003. From attitudes to behaviour: Basic and applied research on the theory of planned behaviour. *Current Psychology: Developmental, Learning, Personality, Social*, 22(3), 187–195. <http://dx.doi.org/10.1007/s12144-003-1015-5>
- Armitage, C. J., and M. Conner. 1999. The theory of planned behaviour: Assessment of predictive validity and “perceived control.” *British Journal of Social Psychology*, 38: 35–54. <http://dx.doi.org/10.1348/014466699164022>
- Asher, J. W. 1997. The role of measurement, some statistics, and some factor analysis in family psychology research. *Journal of Family Psychology*, 11(3): 351–360. <http://dx.doi.org/10.1037/0893-3200.11.3.351>

- Askew, M., V. Rhodes, M. Brown, D. William, and D. Johnson. 1997. *Effective Teachers of Numeracy. Report of a Study Carried Out for the Teacher Training Agency*. London: King's College London, School of Education.
- Belar, C. 2003. Training for evidence-based practice. *APA Monitor on Psychology*, 34 (5): 56.
- Borg, I., and P. Groenen. 1997. *Modern multidimensional scaling*. New York: Springer-Verlag.
- Briggs, S. R., and J. M. Cheek. 1986. The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54(1):106–148.
<http://dx.doi.org/10.1111/j.1467-6494.1986.tb00391.x>
- Brooks, J. G. and M. G. Brooks. 1993. *In search of understanding: The case for constructivist classrooms*. Association for Supervision and Curriculum Development. Alexandria, VA.
- Byrne, B. M. 2001. *Structural equation modeling with AMOS: Basic concepts, applications and programming*. Mahwah, NJ: Lawrence Erlbaum.
- Cacioppo, J. T., W. L. Gardner, and G. G. Berntson. 1997. Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review*, 1(1): 3–25.
http://dx.doi.org/10.1207/s15327957pspr0101_2
- Caprio, M. 1994. Easing into constructivism. *Journal of College Science Teaching*, 23(6): 210–212.
- Carmines, E. G., and R. A. Zeller. 1979. *Reliability and Validity Assessment*. Newbury Park, CA: Sage Publications.
- Carroll, J. D., and J.J. Chang. 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35: 283–319.
<http://dx.doi.org/10.1007/BF02310791>
- Clark, L. A., and D. Watson. 1995. Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3): 309–319.
<http://dx.doi.org/10.1037/1040-3590.7.3.309>
- Cobb, G. W. 1992. Teaching statistics. In *Heeding the call for change: Suggestions for curricular action*, ed. L. Steen, MAA Notes, No. 22, 3–43.
- Cobb, P. 1994. Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher*, 23(7): 13–20.
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum.
- Comrey, A. L., and H. B. Lee. 1992. *A first course in factor analysis* (2nd ed.). Erlbaum: Hillsdale, NJ.
- Cook, T. D., and D. T. Campbell. 1979. *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.
- Coxon, A. P. M. 1982. *The user's guide to multidimensional scaling*. London: Heinemann Educational Books.
- . 2004. Multidimensional Scaling. In *The SAGE Encyclopedia of Social Science Research Methods*, ed. M.S. Lewis-Beck, et al. London: Sage Publications.
- et al. 2010. *NewMDSX: The User Manual*. <http://www.newmdsx.com>

- (accessed May 12, 2011).
- Cronbach, L. J. 1947. Test “reliability”: Its meaning and determination. *Psychometrika*, 12: 1–16. <http://dx.doi.org/10.1007/BF02289289>
- . 1951. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3): 297–334. <http://dx.doi.org/10.1007/BF02310555>
- , and P.C. Meehl. 1955. Construct validity in psychological tests. *Psychological Bulletin*, 52: 281–302. <http://dx.doi.org/10.1037/h0040957>
- Cruise, S. M., C. A. Lewis, and C. McGuckin. 2006. Internal reliability and temporal stability of the Oxford Happiness Questionnaire short-form: Test-retest data over two weeks. *Social Behavior and Personality*, 34(2): 123–126. <http://dx.doi.org/10.2224/sbp.2006.34.2.123>
- Davis, F., R. Bagozzi, and R. Warshaw. 1989. User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 35: 982–1003. <http://dx.doi.org/10.1287/mnsc.35.8.982>
- Dennen, V. P. 2004. Cognitive apprenticeship in educational practice: Research on scaffolding, modeling, mentoring, and coaching as instructional strategies. In *Handbook of research on educational communications and technology*, ed. D. H. Jonassen, 813–828. Lawrence Erlbaum Associates. NJ.
- delMas, R. C., J. Garfield, and B. Chance. 1999. A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, 7(3). <http://www.amstat.org/publications/jse/secure/v7n3/delmas.cfm> (accessed January 12, 2011).
- delMas, R., J. Garfield, A. Ooms, and B. Chance. 2006. Assessing students' conceptual understanding after a first course in statistics. In *Proceedings of the Annual Meeting of The American Educational Research Association (AERA)*. San Francisco, CA.
- DeVellis, R. F. 2003. *Scale development: Theory and applications*. Thousand Oaks: Sage Publications.
- Erlandson, D. A., E. L. Harris, B. Skipper, and S. D. Allen. 1993. *Doing naturalistic inquiry: A guide to methods*. Newbury Park, CA: Sage Publications.
- Fabrigar, L. R., D. T. Wegener, and R. C. MacCallum. 1999. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3): 272–299. <http://dx.doi.org/10.1037/1082-989X.4.3.272>
- Fishbein, M., and I. Ajzen. 1975. *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Fitzgerald, L. F., and L. J. Hubert. 1987. Multidimensional scaling: Some possibilities for counseling psychology. *Journal of Counseling Psychology*, 34(4): 469–480. <http://dx.doi.org/10.1037/0022-0167.34.4.469>
- Floyd, F. J., and K. F. Widaman. 1995. Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3): 286–299. <http://dx.doi.org/10.1037/1040-3590.7.3.286>
- Fosnot, C. T. 1996. *Constructivism: Theory, Perspectives, and Practice*, Vol. I. New York: Teachers College Press.
- . 2005. *Constructivism: Theory, perspectives, and practice*, Vol. II. New York:

- Teachers College Press.
- et al. 2007. *Contexts for learning mathematics, K-6*, Portsmouth, NH: Heinemann Press and Harcourt School Publishers.
- Franklin, C., and J. Garfield. 2006. The GAISE project: developing statistics education guidelines for pre K-12 and college courses. In *Thinking and reasoning with data and chance*, ed. G. F. Burrill and P.C. Elliot, 345–375. NCTM Yearbook. <http://www.amstat.org/Education/gaise/GAISECollege.htm> (accessed January 12, 2011).
- Froelich, A., W. R. Stephenson, and W. Duckworth. 2008. Assessment of materials for engaging students in statistical discovery. *Journal of Statistics Education*, 16(2). www.amstat.org/publications/jse/v16n2/froelich.html (accessed January 12, 2011).
- Gal, I. 2000. Statistical literacy: Conceptual and instructional issues. In *Perspectives on adults learning mathematics*, ed. D. Coben, J. O’Donoghue, and G. E. Fitzsimons, 135-150. Boston: Kluwer Academic Publishers.
- Garfield, J. 1993. Teaching statistics using small-group cooperative learning. *Journal of Statistics Education*, 1(1). <http://www.amstat.org/publications/jse/v1n1/garfield.html> (accessed January 12, 2011).
- , and D. Ben-Zvi. 2007. How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3): 372–396. <http://dx.doi.org/10.1111/j.1751-5823.2007.00029.x>
- . 2009. Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment. *Teaching Statistics*, 31(3): 72–77. <http://dx.doi.org/10.1111/j.1467-9639.2009.00363.x>
- Garfield, J., B. Hogg, C. Schau, and D. Whittinghill. 2002. First courses in statistical science: The status of educational reform efforts. *Journal of Statistics Education*, 10(2). <http://www.amstat.org/publications/jse/v10n2/garfield.html> (accessed January 12, 2011).
- Gorsuch, R. L. 1983. *Factor analysis*. Lawrence Erlbaum Associates: Hillsdale, NJ.
- . 1997. Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, 68(3): 532–560. http://dx.doi.org/10.1207/s15327752jpa6803_5
- Green, B.F. 1978. In defense of measurement. *American Psychologist*, 33: 664–670. <http://dx.doi.org/10.1037/0003-066X.33.7.664>
- Green, J. J., C. C. Stone, A. Zegeye, and T. A. Charles. 2009. How much math do students need to succeed in business and economics statistics? An ordered probit analysis. *Journal of Statistics Education*, 17(3). www.amstat.org/publications/jse/v17n3/green.html (accessed January 12, 2011).
- Hall, G., and S. Hord. 1987. *Change in schools: Facilitating the process*. Albany, NY: State University of New York Press.
- Handal, B. 2003. Profiling teachers: Constructivist- and behaviorist-oriented mathematics. *The International Online Journal of Science and Mathematics Education*, 3.
- Hassad, R. A. 2003. Teaching introductory statistics in the social and behavioral sciences – Approach and rationale. In *Proceedings of the American Statistical Association*

- Joint Statistical Meetings*. San Francisco, CA.
<http://www.eric.ed.gov/PDFS/ED506521.pdf> (accessed January 12, 2011).
- . 2006. Reflection on training, experience, and introductory statistics: A mini-survey of tertiary level statistics instructors. In *Proceedings of the ISI-IASE International Conference on Teaching Statistics (ICOTS-7)*, ed. A. Rossman and B. Chance. Salvador, Bahia, Brazil.
<http://www.stat.auckland.ac.nz/~iase/publications/17/C141.pdf> (accessed January 12, 2011).
- . 2007. Development and validation of a scale for measuring instructors' attitudes toward concept-based or reform-oriented teaching of introductory statistics in the health and behavioral sciences. PhD dissertation, Touro University International.
- . 2009. Reform-oriented teaching of introductory statistics in the health, social and behavioral sciences – Historical context and rationale. *International Journal of Human and Social Sciences*, 4(2): 132–137.
<http://www.waset.org/journals/ijhss/v4/v4-2-17.pdf> (accessed January 12, 2011).
- . 2010. Toward improving The quality of doctoral education: A focus on statistics, research methods, and dissertation supervision. In *Proceedings of the ISI-IASE International Conference on Teaching Statistics (ICOTS-8)*, ed. C. Reading. Ljubljana, Slovenia.
http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_C121_HASSAD.pdf (accessed January 12, 2011).
- , and A. P. M. Coxon 2007. Development and validation of a scale to measure instructors' attitudes toward concept-based teaching of introductory statistics in the health and behavioral sciences at the tertiary level. In *Proceedings of the International Statistical Institute (ISI)*, 56th Session. Lisbon, Portugal.
http://www.stat.auckland.ac.nz/~iase/publications/isi56/CPM81_Hassad.pdf (accessed January 12, 2011).
- Haynes, S. N., D. R. Richard, and E. S. Kubany. 1995. Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3): 238–247. <http://dx.doi.org/10.1037/1040-3590.7.3.238>
- Hogue, A., H. A. Liddle, A. Singer, and J. Leckrone. 2005. Intervention fidelity in family-based prevention counseling for adolescent problem behaviors. *Journal of Community Psychology*, 33(2): 191–211. <http://dx.doi.org/10.1002/jcop.20031>
- Jaccard, J., and H. Blanton. 2005. The origins and structure of behavior: Conceptualizing behavior in attitude research: In *Handbook of Attitudes and Attitude Change*, ed. D. Albarracín, B. T. Johnson, and M. P. Zanna. Mahwah, NJ: Erlbaum.
- Jaworska, N., and A. Chupetlovska-Anastasova. 2009. A review of multidimensional scaling (MDS) and its utility in various psychological domains. *Tutorials in Quantitative Methods for Psychology*, 5(1): 1–10.
- Jenkins, O. F. 2010. Developing teachers' knowledge of students as learners of mathematics through structured interviews. *J Math Teacher Educ*, 13(2):141–154.
<http://dx.doi.org/10.1007/s10857-009-9129-9>
- Kahveci, A. 2009. Exploring chemistry teacher candidates' profile characteristics, teaching attitudes and beliefs, and chemistry conceptions. *Chem. Educ. Res. Pract.*, 10(2): 109–120. <http://dx.doi.org/10.1039/b908248b>

- Kaiser, F. G., and P. W. Schultz. 2009. The attitude-behavior relationship: A test of three models of the moderating role of behavioral difficulty. *Journal of Applied Social Psychology*, 39(1): 186–207. <http://dx.doi.org/10.1111/j.1559-1816.2008.00435.x>
- Kline, T. 2005. *Psychological testing: A practical approach to design and evaluation*. Sage: Thousand Oaks, CA.
- Kloeblen, A. S., and S. S. Batish. 1999. Understanding the intention to permanently follow a high folate diet among a sample of low-income pregnant women according to the Health Belief Model. *Health Educ. Res*, 14(3): 327–338. <http://dx.doi.org/10.1093/her/14.3.327>
- Kruskal, J. B., and R. N. Shepard. 1974. A nonmetric variety of linear factor analysis, *Psychometrika*, 39 (2): 123–157. <http://dx.doi.org/10.1007/BF02291465>
- Kruskal, J. B., and M. Wish. 1978. *Multidimensional Scaling*. Newbury Park, CA: Sage.
- Leonhard, C., K. Mulvey, D. R. Gastfriend, and M. Shwartz. 2000. The addiction severity index: A field study of internal consistency and validity. *Journal of Substance Abuse Treatment*, 18(2): 129–35. [http://dx.doi.org/10.1016/S0740-5472\(99\)00025-2](http://dx.doi.org/10.1016/S0740-5472(99)00025-2)
- Leont'ev, A. N. 1972. The problem of activity in psychology. *Voprosy filosofii*, 9: 95–108.
- Liu, C. H., and R. Matthews. 2005. Vygotsky's philosophy: Constructivism and its criticisms examined. *International Education Journal*, 6(3): 386–399.
- Loewenthal, K. M. 1996. *An introduction to psychological tests and scales*. London: UCL Press Limited.
- Lord, F. M., and M.R. Novick. 1968. *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Luft, J. A., and G. H. Roehrig. 2007. Capturing science teachers' epistemological beliefs: The development of the Teacher Beliefs Interview. *Electronic Journal of Science Education*, 11(2): 38–63. <http://ejse.southwestern.edu/article/view/7794> (accessed January 12, 2011).
- MacInnes, D. 2003. Evaluating an assessment scale of irrational beliefs for people with mental health problems. *Nurse Researcher*, 10(4): 53–67.
- Meehl, P. E. 1954. *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press. <http://dx.doi.org/10.1037/11281-000>
- McGowan, H. M. 2009. Experimentation Methodologies for Educational Research with an Emphasis on the Teaching of Statistics. PhD Dissertation, University of Michigan.
- Mills, J. D. 2002. Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education*, 10(1). <http://www.amstat.org/publications/jse/v10n1/mills.html> (accessed January 12, 2011).
- Moore, D.S. 1998. Statistics among the liberal arts. *Journal of the American Statistical Association*, 93(444): 1253-1259. <http://dx.doi.org/10.2307/2670040>
- Muldoon, M. F., S. Barger, J. Flory, and S. Manuck. 1998. What are quality of life measurements measuring? *British Medical Journal*, 316 (7130): 542–545.

- Mvududu, N. 2003. A cross-cultural study of the connection between students' attitudes toward statistics and the use of constructivist strategies in the course. *Journal of Statistics Education*, 11(3).
<http://www.amstat.org/publications/jse/v11n3/mvududu.html> (accessed January 12, 2011).
- Nunnally, J. C. 1967. *Psychometric theory*. New York: McGraw Hill.
- . 1978. *Psychometric theory* (2nd ed.). New York: McGraw Hill.
- , and I. H. Bernstein. 1994. *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Oppenheim, B. 1982. An exercise in attitude measurement. In *Social psychology: A practical manual*, ed. G.M. Breakwell, H. Foot, and R. Gilmour. London: Macmillan Press.
- Patton, M. Q. 1990. *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage Publications.
- Pecore, J. 2009. A case study of secondary teachers facilitating a historical problem-based learning instructional unit. PhD dissertation, Georgia State University.
- Piaget, J. 1950. *The psychology of intelligence*. London. Routledge & Kegan Paul.
- . 1967. *Biologie et connaissance* (Biology and knowledge). Paris: Gallimard.
- . 1977. *Psychology and epistemology: Towards a theory of knowledge*. New York: Penguin.
- Ravitz, J.L., H.J. Becker, and Y. Wong. 2000. Constructivist compatible beliefs and practices among U.S. teachers. *Teaching, Learning and Computing Report 4*. University of California.
- Riel, M., and H. Becker. 2000. The beliefs, practices, and computer use of teacher leaders. In *Proceedings of the Annual Meeting of the American Educational Research Association (AERA)*. New Orleans.
- Robinson, J. P., P. R. Shaver, and L. S. Wrightsman. 1991. *Measures of personality and social psychological attitudes*. San Diego: Academic Press.
- Rosenberg, M., and C. Hovland. 1960. Cognitive, affective and behavioral components of attitudes. In *Attitude organization and change*, ed. C. I. Hovland and M. J. Rosenberg, 1–14. New Haven, CT: Yale University Press.
- Roskam, E. E. 1972. MDS by metric transformation of the data. *Psychologie*, 27.
- , and J. C. Lingoes. 1970. Minissa-I: A FORTRAN IV (G) program for the smallest space analysis of square symmetric matrices. *Behavioral Science*, 15: 204–205.
- Russell, D.W. 2002. In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin*, 28(12): 1629–1646.
<http://dx.doi.org/10.1177/014616702237645>
- Schmitt, N. 1996. Uses and abuses of coefficient alpha, *Psychological Assessment*, 8(4): 350–353. <http://dx.doi.org/10.1037/1040-3590.8.4.350>
- Schwarz, N. 2007. Attitude construction: Evaluation in context. *Social Cognition*, 25 (5): 638–656. <http://dx.doi.org/10.1521/soco.2007.25.5.638>
- Shulman, L.S. 1986. Those who understand: Knowledge growth in teaching, *Educational Researcher*, 15(2): 4–14.

- Singer, J., and J. Willett. 1993. Lessons we can learn from recent research on teaching: It's not just the form, it's the authenticity. In *Proceedings of the American Statistical Association Joint Statistical Meetings*. San Francisco, CA.
- Skinner, B. F. 1974. *About behaviorism*. New York: Knopf.
- Skemp, R. R. 1987. *The psychology of learning mathematics*. Lawrence Erlbaum Associates: NJ Hillsdale.
- Smith, M. B. 1947. The personal setting of public opinion. *Public Opinion Quarterly*, 11: 507–23. <http://www.jstor.org/stable/2745499>
- Spence, I. 1979. A simple approximation for random rankings stress values. *Multivariate Behavioral Research*, 14: 355–365. http://dx.doi.org/10.1207/s15327906mbr1403_5
- Steen, L. A., ed. 2004. *Achieving quantitative literacy: An urgent challenge for higher education*. Washington, DC: Mathematical Association of America.
- Steinhorst, R. K., and C. M. Keeler. 1995. Developing material for introductory statistics courses from a conceptual, active learning viewpoint. *Journal of Statistics Education*, 3(2). <http://www.amstat.org/publications/jse/v3n3/steinhorst.html> (accessed January 12, 2011).
- Streiner, D. L., and G. R. Norman. 1989. *Health measurement scales: A practical guide to their development and use*. Oxford: Oxford University Press.
- Taylor, S., and P. A. Todd. 1995. Understanding information technology usage: A test of competing models. *Information Systems Research*, 6(2):144–176. <http://dx.doi.org/10.1287/isre.6.2.144>
- Thompson, P. W. (2000). Radical constructivism: Reflections and directions. In *Radical constructivism in action: Building on the pioneering work of Ernst von Glasersfeld*, ed. P. Steffe, and P. W. Thompson, 412–448. London: Falmer Press.
- Trigwell, K., and M. Prosser. 2004. Development and use of the approaches to teaching inventory. *Educational Psychology Review*, 16 (4): 409–424.
- Tschannen-Moran, M., C. Uline, A. W. Hoy, and T. Mackley. 2000. Creating smarter schools through collaboration. *Journal of Educational Administration*, 38(3): 247–271. <http://dx.doi.org/10.1108/09578230010342312>
- Venkatesh, V., and F. D. Davis. 2000. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46:186–205. <http://dx.doi.org/10.1287/mnsc.46.2.186.11926>
- Viswanathan, M. 2005. *Measurement Error and Research Design*. Thousand Oaks: Sage Publications.
- von Glasersfeld, E. 1987. Learning as a constructive activity. In *Problems of representation in the teaching and learning of mathematics*, ed. C. Janvier, 3–17. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Vygotsky, L.S. 1962. *Thought and language*. MIT Press. Cambridge, MA. <http://dx.doi.org/10.1037/11193-000>
- . 1978. *Mind in Society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wallace, D. S., R. M. Paulson, C. G. Lord, and C. F. Bond. 2005. Which behaviors do attitudes predict? Meta-analyzing the effects of social pressure and perceived

- difficulty. *Review of General Psychology*, 9(3): 214–227.
<http://dx.doi.org/10.1037/1089-2680.9.3.214>
- Woolley, S. L., W. J. Benjamin, and A. W. Woolley. 2004. Construct validity of a self-report measure of teacher beliefs related to constructivist and traditional approaches to teaching and learning. *Educational and Psychological Measurement*, 64(2): 319–331. <http://dx.doi.org/10.1177/0013164403261189>
- Yu, C. H. 2001. An introduction to computing and interpreting Cronbach coefficient alpha in SAS. In *Proceedings of 26th SAS User Group International Conference*. Cary, NC: SAS Institute Inc.
- Zimbardo, P.G., and M.R. Leippe. 1991. *The psychology of attitude change and social influence*. New York: McGraw-Hill.