TECHNICAL GUIDE[1]

to

SPTE: Student Perceptions of Teaching Effectiveness

General Information

SPTE is a norm-based instrument which includes 39 items from which eight scales are derived by factor analytic statistical techniques. Results for each of the eight scales are reported relative to two separate norm bases: 1) norms in which the instructor's course ratings are reported relative to the ratings for the last ten years of university courses for which we have data, and 2) norms in which the instructor's course ratings are reported relative to the ratings for the past ten years of courses in the instructor's college (e.g., Barton School of Business) or college division (e.g., Liberal Arts and Sciences at WSU is composed of three subdivisions; Humanities, Natural Sciences, and Social Sciences) for which we have data. Each scale is reported using a Scale score, as well as a percentile. In addition, individual item results are given, with raw scores, Scale scores and percentiles. The SPTE is a measure of students' perceptions of teaching and course quality. When used for summative purposes, it should be used in conjunction with other evaluative material.

SPTE Instrument

Creation of Instrument and Selection of Items

In 1975, the Liberal Arts and Sciences Teaching Improvement Committee (LASTIC) convened for the purpose of determining methods of obtaining feedback regarding instruction from students at Wichita State University, Wichita, Kansas.[2] As part of that assignment, an original questionnaire was developed with the purpose of assessing student perceptions of teaching effectiveness. At that time it was called LASTIC after the name of the committee. When the Social Science Research Laboratory took over LASTIC, as requested by the LASTIC committee, the instrument was renamed to Student Perceptions of Teaching Effectiveness (SPTE).

The questionnaire underwent minor revision between 1975 and 1986 (the addition of three items for a total of 39 evaluative items, as well as eleven demographic items). In 1995, the questionnaire was totally revised. The final revision still has 39 evaluative questions and generally the same content. However, the wording is more concise. (See Appendix A for the most current SPTE items and response anchors).

Response Validity Items

In addition to the 39 evaluative items and the eleven demographic items, the instrument includes two items to check for response validity. One item requires the individual to leave the item blank, while the other requires the individual

---

[1]This Technical Guide is intended to provide information regarding the process by which SPTE Results are produced. For information regarding the interpretation of SPTE Results, refer to the SPTE Interpretation Guide.

[2]It is interesting to note that no one on the committee had any background in scale development.

to mark two responses. These items are interspersed among the 39 evaluative items. If either item is marked incorrectly, that student's entire questionnaire is left out of the generation of an instructor's scores. Note that we reject approximately 1.5% of the individual surveys.

To further encourage students to read the items, the items have varying anchors, with the desirable end of the scale alternately corresponding to the low and the high ends of the scale.

Generation of SPTE Results

Unit of Measurement

For each of the 39 items, all of the student responses within a given section are averaged. (The term "section" refers to one "class" of students, for example, all of the students in a Psy 304Q Social Psychology class make up one section.) As a result, the section is treated as the unit of measurement (participant) rather than individual students. This process eliminates variance attributed to individual differences of the students. Thus, only the variance that is attributed to the section, which is more directly affected by the instructor is analyzed.

Factor Analytic Procedures

Conceptually, factor analysis is a statistical procedure for determining the number and nature of underlying constructs from a set of co-varying variables. "Factors" are the dimensions or constructs that are considered to "cause" covariance in a set of variables, for example, "depression" is a construct.

Determining the number of dimensions in the questionnaire initially used the tests that were available at the time; Cattell's Scree (Cattell, 1978) and Kaiser-Guttman (Gorsuch, 1982) both indicating six factors. Later factoring used Velicer's Minimum Average Partials (Velicer, W. F. (1976) & Velicer, W. F., Eaton, C. A., and Fava, J. L. (2000)) and Cattell's Scree test. Both tests showed six factors.

Once we knew how many factors, we used a principle axis factoring. They were then rotated using the PROMAX oblique rotation with a Kaiser normalization. A stable factor structure was achieved which met the criteria for simple structure (Gorsuch, 1982).

For an overview of the processes the factor analysis of the SPTE data includes each of these steps. Cattell's scree yields six first order factors as did Velicier's MAP . Extraction of the SPTE factors typically (we've done it several times as data increased) resulted in fixed communality estimates that are above 0.6 for 36 of the 39 variables (see Appendix B for the most current communality estimates). The factors were then rotated using a PROMAX rotation as described above. The first order factor pattern may be found in appendix C and the associated factor correlations in Appendix D.

As the initial factors were correlated, a second-order factoring revealed two second-order factors. This resulted in similar fixed communality estimates, and was conducted in the same manner as is described for the first-order factor analysis. The results of the 2nd order factoring may be found in Appendix E.

The SPTE data[3] are factor analyzed at regular intervals to determine that the factor structure remains stable. Over the last 25 years, the factor structure has remained stable across versions with the possible exception is that on some occasions, Workload and Difficulty were not clearly separated.

Generation of Factor Scores

For each of the eight scales (six first order and two second-order factors), factor scores are estimated using the regression method. For each of the eight scales, all variables are used as predictors using the variable's loading for that particular factor (whether salient or not). When the factor scores are estimated, a correlation matrix of the factor scores with the factors is calculated. The diagonal of this matrix is essentially equivalent to an alpha coefficient for each of the eight scales (Gorsuch, 1982). For each scale, these equivalent alpha values are consistently over 0.9 except for the 'Difficulty' score which was a respectable .85.

The Necessity of Correcting for Bias

Our research has demonstrated that some variables outside the instructor's control tend to be systematically related to instructors' ratings. These variables produce variance in ratings that is not related to what the instructor actually does in class. The most common of these include *a priori* student motivation (how much the student wanted to take the class) and class size. *A priori* student motivation is positively related to student evaluations, while class size is negatively related. Students who report that they really want to take a class prior to its start typically evaluate the class as more valuable to them (scores for the *Course Value* scale are affected). While less of a factor, students in smaller classes (in which an instructor may be more accessible, individual attention may be more readily available, and the nature of the course itself may be more enjoyable) tend to rate the instructors' of those classes higher with regard to *Rapport with Students* and *Course Value.* Conversely, ratings of instructors are lower from students who report very little *a priori* motivation as well as those from larger classes. Of these two, *a priori* student motivation appears to be more strongly related to instructors' ratings than does class size. However, both the correlations between student ratings (of instructors) and *a priori* student motivation and student ratings and class size are strong enough that instructor's ratings should be corrected. The correlations are recalculated each semester, to be sure that the correction of the factors scores is precise. The most current correlations are shown in Appendix F.

Measurement of Bias

Class size

The measurement of class size is determined by the number of raters that complete an SPTE questionnaire during the evaluation period. Current research is directed toward determining if the number of SPTE raters is a sufficient proxy for class size (the number of raters for each class is being compared to enrollment at the end of the semester).

A Priori Student Motivation

A priori *student motivation is determined through four questions which are included on the SPTE questionnaire, but* are not included among the 39 evaluative items. The individual responses to each of these four items are averaged and summed for the class, producing a mean *a priori* student motivation score for the class. See Appendix F for these items. See Appendix G for the correlations of these two variables with the SPTE scales.

---

[3] SPTE norm bases ("SPTE data") always include the most recent ten years of data. When there are revisions in SPTE such as there were in 1995, the norm bases must be recreated and as such, may contain less than ten years of data.

Correction Procedure

Each of the individual items is corrected for both class size and *a priori* student motivation. Correction of items involves the residualization of each item by class size and student motivation. Residualization involves using multiple regression techniques to predict each item from *a priori* student motivation and class size. The predicted value of the item is then subtracted from the actual value of the item, and this results in a measure of the item that is unrelated to either *a priori* student motivation or class size.

Factor scores are also corrected for class size and *a priori* student motivation. The uncorrected items are used to produce the factor scores. Then the factors themselves are corrected using multiple regression techniques to residualize the factor score by *a priori* student motivation and class size.

Norm Bases

Once the corrected items and corrected factor scores have been calculated for each section, these scores are added to the norm bases described earlier. Norms are updated and recalculated each semester (Spring, Summer, and Fall) of each year.

A large number of sections are evaluated during each of the fall and spring semesters. For example, in the Fall of 2018, 1,418 sections were evaluated. Due to the large number of sections being evaluated, the university norm base increases dramatically each semester. College or division norms depend on degree of usage in the college or division. Norms for an individual college/division are created when at least 70 sections in that college/division have been evaluated by SPTE. Typically, both the university norms and the college or division norms include the most recent ten years of data.

Converting to Scale Scores

After the norm bases have been updated, the scores on each of the eight scales and each of the 39 individual items are converted to Scale scores. Essentially, each scale (and each item) is standardized according to the norm base being used (either university or college/division), and is reported as a "standard ten" or "Scale" score. Scale scores have an approximate range of one to ten (hence, the "standard ten"), with an average of 5.5 and a standard deviation of 2. Each scale (factor) score and each item score are converted to a Scale score, rounded to two decimal places[4]. The Scale score indicates an instructor's position relative to the average (of either all university classes or of all the classes in your college or division) for that scale (or item). Depending upon the scale, a high Scale score indicates either a desirable score or that the instructor placed high demands upon students.

Calculation of Percentiles

Percentiles are also calculated for each of the eight scales and each of the 39 items, using the norms previously described. Percentiles are calculated directly from Scale scores within each norm base for each scale (or item). Each of the Scales are recorded from lowest to highest. Frequency counts of each Scale are made. Any Scale value above

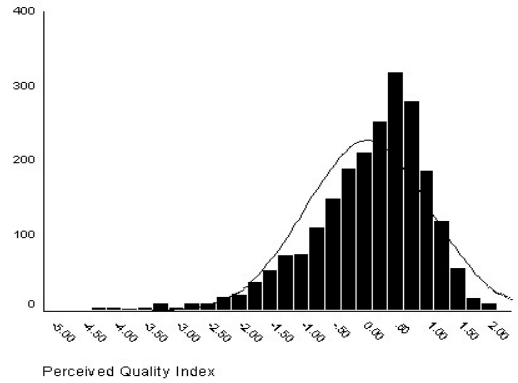
---

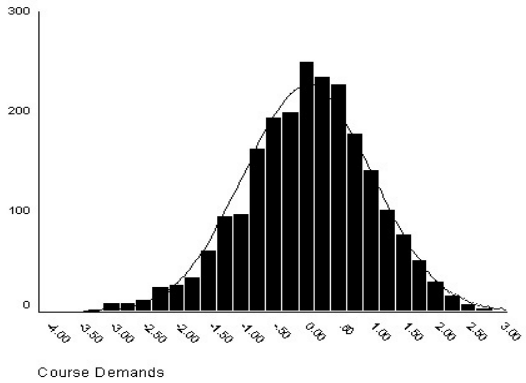
[4] For each of the scale (and item) scores a Scale is calculated by the following equation:

$$X_{STEN} = \left(\frac{X_{scale} - \mu_{scale}}{\sigma_{scale}}\right)2 + 5.5$$

11.5 (3 standard deviations above the mean) are set at a “ceiling” of 11.5, while those scores lower than -.5 (three standard deviations below the mean) are set at a ‘floor” value of -.5. The frequency of each of the Scale values is divided by the total number of Scales in the norm base. Once multiplied by 100, the resulting values become the associated percentiles for the Scale values. An individual instructor’s Scale, then, is associated with a calculated percentile which is reported in addition to the Scale value.

The distributions of z-scores for the Perceived Quality Index and Course Demands from three semesters of data may be seen in the following figures with a normal curve imposed:





You will notice that while Course Demands appears normally distributed (-.287), the Perceived Quality Index is somewhat negatively skewed (-1.204). While such skewness is desirable for the overall quality of teaching at WSU, estimation of percentiles from standard scores can misrepresent an individual's score relative to all others in the norm base. In other words, with many high scores bunched together, even a small change in a raw score can make a dramatic change in the percentile. Scale scores, however, tell you where you stand relative to the mean and, therefore, are not as sensitive to small changes in the raw score and are the measure of choice for evaluative use.

Generating SPTE Results for Each Section (Instructor)

For each instructor (course), SPTE results are generated as Scale scores and percentiles for that section. These results are returned to the instructor of the section. Two Scale scores are reported for each of the eight scales, one generated using the university norm base and the other using the college or division norm base. For the items, only one set of Scale scores is generated using the college or division norms. However, if an insufficient number of instructors in the college or division have used SPTE (less than 70), the university normed items are presented. Similarly, two percentile scores are reported for each of the eight scales, and one percentile score for the items.

References

Cattell, R. B. (1978). The position of factor analysis in psychological research. In R. B. Cattell (Ed.), The Scientific Use of Factor Analysis in Behavioral and Life Sciences (pp. 3-14). New York: Plenum Press.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. Psychometrika, 41, 321-327.

Gorsuch, R. L. (1983). Factor Analysis. Hillsdale, NJ: Lawrence Erlbaum Associates.

Burdsal, C. & Bardo, J. (1986). Measuring student's perceptions of teaching: dimensions of evaluation. Educational Psychological Measurement, 46, 63-79

Jackson, D., Teal, C. R., Raines, S. J., Nansel, T. R., Force, R. C. & Burdsal, C. A. (2000). The Dimensions of students' perceptions of teaching effectiveness. Educational Psychological Measurement, 59(4), 580-596.

Velicer, W. F., Eaton, C. A., and Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. Pp. 41-71 in R. D. Goffin and E. Helmes, eds., Problems and solutions in human assessmen Boston: Kluwer.

Harrison, P. D., Douglas, D. K and Burdsal, C. A. (2004) The relative merits of different types of overall evaluations of teaching effectiveness. Research in Higher Education 45(3), 311-323.

Burdsal, C. A. & Harrrison, P. D. (2008) Further evidence supporting the validity of both a multidimensional profile an an overall evaluation of teaching effectiveness. Assessment and Evaluation in Higher Education, 33, 5.

Appendix A: SPTE II Instrument[5]

9.	The instructor came across as a person as very well	not at all well as a teacher.	[1]	[2]	[3]	[4]	[5]
10.	I usually went to this class with great reluctance	eager anticipation	[1]	[2]	[3]	[4]	[5]
11.	For this class, the method of presentation of appropriate material (lecture, discussion, etc.) was	inappropriate	[1]	[2]	[3]	[4]	[5]
12.	The types of evaluations (exams, assignments, papers) used in determining the grades were	inappropriate appropriate	[1]	[2]	[3]	[4]	[5]
13.	The time required to prepare for this class was extensive	minimal	[1]	[2]	[3]	[4]	[5]
14.	Regarding my progress in this course, the instructor was concerned and actively helpful.	strongly disagree strongly agree	[1]	[2]	[3]	[4]	[5]
15.	I expect my retention of the course material will be low	high	[1]	[2]	[3]	[4]	[5]
16.	The objectives of the course were unclear	clear	[1]	[2]	[3]	[4]	[5]
17.	The number of evaluations (exams, assignments, papers, etc.) used in determining grades was	sufficient	[1]	[2]	[3]	[4]	[5]insufficient
18.	The instructor's syllabus was	poor/non-existent	[1]	[2]	[3]	[4]	[5]excellent
19.	The difficulty level of the material presented was	high	[1]	[2]	[3]	[4]	[5]low
21.	The required readings, text, and other materials were relevant to course content.	mostly	[1]	[2]	[3]	[4]	[5]not at all
22.	The instructor challenged me intellectually.	never	[1]	[2]	[3]	[4]	[5]frequently
23.	The instructor graded	very easy	[1]	[2]	[3]	[4]	[5]very hard
24.	The number of assignments was	minimal	[1]	[2]	[3]	[4]	[5]extensive

[5] Items 1-8 are omitted as they are the motivation scale and some demographics. Also, the two validity items are also omitted.

25.	The instructor's manner of presentation (voice level, rate of speaking, etc.) was	conducive to learning	[1]	[2]	[3]	[4]	[5] poor/distracting
26.	The instructor was in control of the class.	usually	[1]	[2]	[3]	[4]	[5]rarely
27.	As a result of this course, my knowledge of the subject has increased	very little	[1]	[2]	[3]	[4]	[5]a great deal
28.	The instructor's classroom presentation was prepared	well prepared	[1]	[2]	[3]	[4]	[5]poorly
29.	The content of the exams and other evaluations were consistent with the material presented in the course.	very much so	[1]	[2]	[3]	[4]	[5] not at all
30.	The speed at which the instructor covered the material was	slow	[1]	[2]	[3]	[4]	[5] fast
31.	The instructor responded fully to questions from students.	rarely	[1]	[2]	[3]	[4]	[5]always
32.	As a result of this course, my interest in the subject has been	stimulated	[1]	[2]	[3]	[4]	[5]stifled
33.	Overall, the instructor was	well organized	[1]	[2]	[3]	[4]	[5]poorly organized
34.	The method of assigning grades was	unclear	[1]	[2]	[3]	[4]	[5]clear
35.	The workload for this class was	light	[1]	[2]	[3]	[4]	[5]heavy
36.	The instructor treated students	respectfully	[1]	[2]	[3]	[4]	[5]disrespectfully
37.	I find that this course has been	of very little value	[1]	[2]	[3]	[4]	[5]very useful
38.	The instructor's knowledge of the subject appeared	poor	[1]	[2]	[3]	[4]	[5]exceptional
39.	The instructor was aware when students were having difficulty understanding a topic.	usually	[1]	[2]	[3]	[4]	[5]rarely
40.	The grade I expect to earn reflects my performance in the course.	very well	[1]	[2]	[3]	[4]	[5]not at all
41.	The amount of material presented in this course was	minimal	[1]	[2]	[3]	[4]	[5]extensive
42.	The instructor was enthusiastic about the class.	strongly agree	[1]	[2]	[3]	[4]	[5]strongly disagree
43.	Would you recommend this course to another student?	definitely yes	[1]	[2]	[3]	[4]	[5]definitely no
45.	The feedback given by the instructor on my course work was	adequate	[1]	[2]	[3]	[4]	[5]inadequate

46. The instructor's ability to convey the key concepts of this course in a clear and meaningful manner was	**very poor**	**[1]**	**[2]**	**[3]**	**[4]**	**[5]very good**
47. This course was	**very hard**	**[1]**	**[2]**	**[3]**	**[4]**	**[5]very easy**
48. I felt free to ask questions and make comments.	**never**	**[1]**	**[2]**	**[3]**	**[4]**	**[5]usually**
49. The ability of the instructor to answer questions was	**poor**	**[1]**	**[2]**	**[3]**	**[4]**	**[5]excellent**

Appendix B: First Order Factoring

Communality Estimates

Item #	Initial	Extraction
i9	.839	.842
i10	.807	.807
i11	.847	.841
i12	.820	.833
i13	.757	.743
i14	.769	.764
i15	.827	.846
i16	.825	.820
i17	.698	.720
i18	.589	.555
i19	.807	.805
i21	.523	.490
i22	.709	.703
i23	.600	.563
i24	.733	.748
i25	.817	.798
i26	.635	.628
i27	.852	.860
i28	.900	.882
i29	.742	.734
i30	.427	.345
i31	.857	.835
i32	.871	.887
i33	.890	.867
i34	.728	.738
i35	.879	.950
i36	.717	.690
i37	.889	.894
i38	.733	.661
i39	.797	.786
i40	.725	.728
i41	.695	.673
i42	.724	.695
i43	.891	.889

i45	.799	.785
i46	.915	.914
i47	.881	.902
i48	.779	.799
i49	.911	.894

Extraction Method: Principal Axis Factoring. Seven Iterations

Appendix C: Factor Pattern[6]

SPTE Item #	Factor					
	Course Design	Course Value	Difficulty	Rapport w/ Students	Grading Quality	Workload
i9	.311	.093	-.051	.622	-.037	.008
i10	-.050	-.814	.056	-.095	.056	.067
i11	.714	.142	-.032	.003	.122	-.014
i12	.275	.110	-.039	.054	.589	.027
i13	-.031	.016	.766	.036	-.065	.266
i14	.149	.162	-.004	.555	.118	.105
i15	.065	.954	-.141	-.102	-.044	.111
i16	.601	.150	-.053	-.074	.320	.010
i17	.188	.109	.001	.014	.615	.237
i18	.698	-.107	-.015	-.039	.226	.095
i19	-.192	.005	.933	.097	.041	-.125
i21	.324	.128	.232	.002	.310	-.038
i22	.150	.311	.647	.173	.010	-.057
i23	.163	.034	.490	-.217	-.379	.149
i24	-.089	-.046	.424	.057	.160	.691
i25	-.711	-.179	.086	-.200	.179	-.060
i26	.783	.025	.091	.013	-.073	-.103
i27	.219	.761	.178	-.112	.083	-.061
i28	-1.038	.006	-.019	.162	-.005	.012
i29	.457	.065	.052	-.022	.460	.002
i30	.233	-.071	.515	-.241	.084	.006
i31	.584	-.078	-.015	.504	-.039	-.038
i32	.027	.940	-.020	-.023	-.018	-.082
i33	1.047	-.087	-.016	-.141	.058	.004
i34	.496	-.183	.029	.042	.565	-.071
i35	-.042	-.015	.708	.032	-.039	.546
i36	.320	-.231	.020	.675	.103	-.092
i37	.073	.931	.078	-.098	.042	-.017
i38	.720	.038	.177	.066	-.056	-.161
i39	.312	.150	-.049	.448	.094	.082
i40	.064	.303	-.194	.167	.399	.060
i41	.177	-.043	.780	-.090	.103	.076
i42	.476	.164	.036	.368	-.123	-.031
i43	.197	.660	-.097	.061	.094	-.050
i45	.366	.085	-.049	.322	.249	.123
i46	.711	.198	-.090	.105	.014	.031
i47	-.090	-.086	.921	.055	-.097	-.007

[6] The factor pattern is recalculated each semester as additional data are added to the norm base. The factor pattern shown here is calculated with data from Fall 1995 through Summer 2018. Items 20 and 44 were omitted as they are the validity items.

i48	.145	.037	-.046	.744	.023	.067
i49	.724	.016	-.017	.357	-.117	-.043

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.
a. Rotation converged in 8 iterations.

Appendix D: Factor Correlations for the 1st Order Factors

Factor Correlation Matrix

Factor	1	2	3	4	5	6
1	1.000	.747	-.034	.634	.571	.030
2	.747	1.000	.044	.698	.593	-.025
3	-.034	.044	1.000	.131	.207	.227
4	.634	.698	.131	1.000	.513	.008
5	.571	.593	.207	.513	1.000	.012
6	.030	-.025	.227	.008	.012	1.000

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.

Appendix E: Second Order Factoring

Communalities

	Initial	Extraction
Course Design	.680	.770
Course Value	.685	.783
Difficulty	.172	.713
Rapport	.628	.694
Grading Quality	.543	.595
Workload	.086	.073

Extraction Method: Principal Axis Factoring.

2nd Order Factors

	PQI	Course Demands
Course Design	0.884	0.114
Course Value	0.888	0.032
Difficulty	0.012	0.846
Rapport	0.819	-0.084
Grading Quality	0.748	-0.119
Workload	-0.056	0.258

Rotation converged in 3 iterations.

Correlation between PQI and Course Demands: .094

Appendix F: *a priori* Motivation Scale

The *a priori* motivation scale is comprised of the following four items:

1.	Prior to enrolling in this class, I expected it to be of	**little or no value**	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**great value**
2.	When I enrolled for this course, I wanted to take this course	**very much**	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**not at all**
3.	I took this course because I had to.	**strongly agree**	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**strongly disagree**
4.	I took this course because I was interested in the subject	**very much**	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**not at all**

The Cronbach's Alpha for the scale is .902 and is base upon 1995 to 2018 data.

Appendix G: Correlations between Each Factor and Class Size and *a priori* Motivation[7]

Correlations

Scale	Motivation Score	Number of SPTEs
Course Design	.301	-.041
Course Value	.671	-.179
Difficulty	-.090	-.026
Rapport	.353	-.205
Grading Quality	.290	-.116
Workload	-.145	-.173
PQI	.487	-.147
Course Demands	-.106	-.021

[7] The correlations between the factors and *a priori* motivation and class size are recalculated each semester as additional data are added to the norm bases. The correlations shown here were calculated with data from Fall 1995 through Summer 2018, for the all university norm base.