Analysis of IS/IT Student Assessment of Courses and Instruction Instruments

William J. Tastle
tastle@ithaca.edu
Department of Management, Ithaca College
Ithaca, New York 14850-7170, USA

Bruce A. White
bruce.white@quinnipiac.edu
Department of Computer Information Systems, Quinnipiac University
Hamden, Connecticut 06518-1908, USA

Abstract

Student assessment of course instruments from 10 colleges/universities are analyzed to determine the degree of consistency among items deemed important in IS/IT courses. A matrix of general categories and subcategories is identified and the instruments from the 10 institutions are tallied to produce a frequency distribution from which the data are normalized via a probability distribution. The subcategories are ranked and examined. Since these instruments are used to assess teaching effectiveness and play a strong role in tenure and promotion decisions, an argument is made in support of measures better suited for the analysis of ordinal data. Most questions on assessment instruments involve the use of the Likert scale, which is ordinal, but the statistics most commonly used are applicable to ratio and interval scales. The use of the mean in making critical ranking decisions is shown to lack conceptual soundness; measures of consensus and agreement offer much better comparative and analytical results. The calculation of agreement distances is illustrated.

Keywords: course assessment, faculty assessment, consensus, agreement

1. Introduction and Current State

The literature is replete with papers on faculty assessment, course assessment, student reaction, and so-called student evaluation of instruction (though we are most concerned with IS/IT faculty assessment) of some sort or another (Abdullat and Terry, 2005; Aasheim and Teichgelt, 2007; Amoroso, 2005; Baugh, 2004; Ceccucci, 2006; Dettori, et al., 2006; Hernandez, et al, 2004; Joseph, 2007; Landry, et. al., 2006; McDonald and Johnson, 2003; McGinnis and Slauson, 2003; McKell, et. al, 2006; O'Neil, 2006; Paranto and Shillington, 2006; Reynolds, et. al. 2004; Stemler and Chamblin, 2006; Todorova and Mills, 2007; White and McCarthy, 2007), reports by institutions (Univ of New South Wales, 2007; Virginia Polytechnic Institute and State University, 2001), to name but a few.

Missing from these studies is an attempt to examine the evaluation instruments used in each institution to identify commonality in approach and effectiveness in application. Therefore, the purpose of this paper is to make an examination of faculty/course assessment forms gathered from ten US colleges/universities located in the eastern half of the country. The institutions selected were simply all who volunteered their instruments to this study. One of the authors sent emails to a large number of colleges located in institutions across the nation, and all forms that were received were included in this analysis. To ensure confidentiality of the donors of the instruments, we are not

listing the institutions, as promised in the email solicitation.

There are two typical reasons for faculty assessment: The first and one could argue the most important for professional educators, is the opportunity for instructional improvement; the other is the critical evaluation of faculty for the purpose of tenure and promotion (McKeachie, 1997). Thus we seek to address the variance in content of this modest set of evaluation instruments and offer some comments on how they might be better analyzed for the above purposes.

2. Background

As our research began in this very complex field it quickly became evident that many concerns abound when the issue of student evaluation of faculty is mentioned. The following is an illustration of concerns that have been mentioned as faculty have been (unscientifically) interviewed about this matter:
1.   Undergraduate students lack the experience and maturity to judge the instructor and instruction.
2. Student rating forms are little more than a popularity contest with the more entertaining and less demanding instructor getting the higher grade.
3.   Students are unable to make accurate judgments until they have experienced the "real world."
4. Student rating forms are both unreliable and invalid.
5.   There are many extraneous variables that can affect student ratings:
     a. size of class,
     b. gender of the student,
     c. time of day the class is offered,
     d. elective or required course,
     e. student is a major or non-major,
     f. term or semester in which the course is offered,
     g. level of the course (freshman, sophomore, etc.), and
     h. the rank of the instructor.
6.   The grade received by the student is highly correlated with the rating given the course and instructor.
7. How can student evaluations be used to improve instruction?

What is most significant about these concerns is that they were originally described by Lawrence Aleamoni (1987) in 1974, and they are virtually identical to the concerns currently being echoed. The issue of evaluation in IS/IT courses is a matter of continuing interest (Aasheim, et al, 2007; Abdullat and Terry, 2005; Amoroso, 2005; Ceccucci, 2006; Hernández, et al, 2004; Landry, et al, 2006; McDonald and Johnson, 2003; McKell, et al, 2006; Reynolds, 2004; Stemler and Chamblin, 2006; Todorova and Mills, 2007; White and McCarthy, 2007). We begin with an analysis of the evaluation instruments.

3. Analysis of the Evaluation Instruments

It is reasonable to expect that every evaluation instrument is different in content. There might be some content overlap with respect to individual questions, but it is reasonable to expect that institutions have their own particular needs as evidenced by the content of the questions contained in their instrument. With this in mind a preliminary assessment of the instruments was made to identify the major categories to which specific assessment items could be tallied. With a concern only to create a comprehensive listing of topics that could be used to assist in this study it quickly became apparent that the listing would approach the total of all questions across the ten instruments, not a particle solution. Thus, the categories are general in scope and the listing of assessment items is also "generalized" so that survey questions can be "approximately" grouped.

General category and their subcategories:
A. Instructor evaluation
  1. Organization and planning
  2. Scholarship
    a)   Knowledge and competence
    b)   Exams give balanced coverage
    c)   Fair and impartial grading
    d)   Course is challenging
  3. Faculty/student interactions
    a)   Student feedback
    b)   Level of concern
    c)   Understandability
    d)   Helpfulness
  4. Presentation of material
    a) Organized
    b) Enjoys teaching

    c) Uses various teaching methods

    d) Encourages participation

    e) Summarizes major points

  5. Overall effectiveness

    a) Comparison with other instructors/courses

    b) Objectives are achieved

    c) Enhanced student interest in subject

B. Evaluation of student learning

  1. Explanation of grading system and assignments

  2. Tests and assignments reflected course content

  3. Assignments contributed to understanding course

  4. Grading returned in reasonable amount of time

  5. Class attendance necessary to learn material

C. Delivery media and facilities

  1. Software enhanced learning

  2. Hardware enhanced learning

  3. Video enhanced learning

  4. Audio enhanced learning

  5. Room enhanced learning

  6. Readings enhanced learning

D. General relevance

  1. Prepared for future professional success

  2. Relevance to rest of discipline

  3. Relevance to rest of business core

  4. Relevance to general education courses

Each institution's assessment (survey) instrument was reviewed to determine a suitable subcategory. For example, the items "Instructor is prepared for each and every class" and "Instructor's organization of each lecture" would both be tallied as **A4a** – *Instructor evaluation/presentation of material/organized* category. The item "The Instructor brings current ideas to the classroom" would be an **A2a** – *Instructor/scholarship/knowledge and competence* category. If there is a fault in this study it is that not enough researchers were engaged in the assignment of items to specific categories, but we are relatively confident that, in general, we have captured a reasonable assignment for each survey item.

The number of items in each assessment form varied from 7 to 29, mean = 15.6, StDev = 7.01, so direct comparison of data was not possible. The method of analysis chosen was to normalize the values for each institution by taking the probability of occurrence for each category. Thus each institution's evaluation form was equally weighted. [Equality in weights may not be appropriate, for institutions that are predominantly research oriented have an agenda quite different from that of institutions that are predominantly teaching oriented.] The sum of the probabilities associated with each category constituted that items *strength*. Table 1 (see appendix) shows the assignment of the institutional assessment forms to the criteria. Note that the institutions are labeled in Roman numerals.

Institution I has a total of 20 items while institution X has only 7. To provide for equal weighting, table 2 (see appendix) shows the probability matrix. The criteria are denoted only by their alpha-numeric designation.

4. Ranking of Criteria

The right end column (labeled "total") of Table 2 represents the sum of the probabilities associated with each criterion in the first column. Those items with the greater probabilities were identified by institutions as being more important. Table 3 is a ranked ordering of the criteria of greatest usage.

Clearly the top six categories of interest are contained in the major category "Instructor Evaluation." The category of least interest is A4e, "summarizes major points." The evaluation of student learning, major category B, is relegated to a disbursement among the columns. For the institutions represented in this study, evaluation of instructors is a more important criterion than the evaluation of student learning, though one might argue that examinations are sufficient for that item.

Figure 1 (see appendix) is a line graph of the strength of usage of the individual categories. It is apparent that A2a, "Knowledge and competence of instructor" is of paramount importance followed by A3d and A4a, "Instructor helpfulness" and "Instructor organization", respectively. Before any gene-

ralizations can be made, far more data need to be collected and analyzed.

Usage of Course/Faculty Evaluations

Since these student assessment forms are used to critique faculty performance, and a tenure or promotion decision might be the outcome of a committee or Dean's analysis of these data, the evaluation of these instruments continues as we examine the way in which the data are summarized for comparison purposes.

Students are asked, in each of these instruments, to make their assignment in a Likert category, *strongly agree* (SA), *agree* (A), *neutral* (N), *disagree* (D), or *strongly disagree* (SD). The five item scale is typically used in such a survey. For purposes of illustration, let us assume that the item to be evaluated is "The clarity and audibility of the instructor's speech are excellent." This falls into the A3c category and is number 6 in relative importance (see Table 3 and Figure 1).

SA	A	N	D	SD	Mean
63	29	8	0	0	1.450
102	29	8	0	5	1.451
79	0	0	0	10	1.449
161	0	1	0	20	1.451

Table 4. Example distributions with similar mean values

Suppose four faculty each receive the distributions in Table 4. How could these faculty be compared? The mean value is calculated in the last column and is supposed to indicate a range of performance from 1 (SA) to 5 (SD). Thus, a 1.450 mean indicates a performance equivalent to about half way between SA and A. Unfortunately, no faculty member could attain that measure because the Likert scale, by definition, is an ordinal scale measure. It is well-known that no evidence exists to suggest that Likert scale categories are equally distributed. If a scale is equally distributed it is an interval (also called cardinal) scale. To use a mean value is to presume the scale is at least interval. Making this kind of calculation over an ordinal scale is the equivalent of saying the average of warm and hot is warm-and-a-half!

Not considered in this evaluation of instructor speech is the concern for a possible physical impairment that might prevent one faculty member from being as loud as another, the room acoustics, the outside noise, next room noise, and a host of other conditions that could play a role in causing one faculty member to be ranked below another. Further, using this kind of arbitrary measure, 50% of all faculty members will be below average. The logic for using comparative numbers, or so it goes, is to "encourage" all faculty members towards continuous improvement. Some might argue that such logic does little more than damage faculty morale. We offer a different approach.

Using a newly developed measure of consensus (Tastle and Wierman, 2005a, 2005b, 2007a, 2007b) it is easy to calculate the overall consensus, a measure that provides a collective indication as to the degree of support by the evaluators. Consensus ranges from a low of 0 (all evaluators are equally divided at the extreme categories, i.e., 50% select SA and the other 50% select SD), and a high of 1 (100% of evaluators select the same category). Applying this measure to the distributions in Table 4 the consensus values of Table 5 are calculated.

We note that the highest consensus (77%) is for the first distribution and the least consensus (49%) is for distributions 3 and 4. The

Dist	Mean	Cns
1	1.450	0.773
2	1.451	0.686
3	1.449	0.493
4	1.451	0.497

Table 5. Comparison of mean and consensus

strongest consensus merely means that the evaluators were in greatest agreement for that distribution, but this does not give any information as to which category to select. To successfully select a category (and not a fractional value that has little conceptual value when dealing with ordinal values, like the 1.450 mean), the agreement measure (Tastle and Wierman, 2008) provides the evaluator with an indication as to the ranking of the individual Likert scale categories.

Specifically, Table 6 shows the agreement values for each distribution in Table 4. The first distribution is shown to possess agreement values of 91% for SA, 86% for A, 68% for N, 44% for D, and 14% for SD. Note that it is possible for categories containing zero selections to possess an agreement. This makes sense when one looks at distribution 2 in Table 6. Since there are evaluators who selected N(eutral) and others who selected S(trongly) D(isagree), it is appropriate that the unselected category D(isagree) have some level of agreement. Distribution 1 in Table 6 has a zero frequency in both D(isagree) and S(trongly) D(isagree) but still have an agreement. This is a function of the mathematics and might be considered representative of a population distribution though this aspect of the measure is currently under investigation. In another paper involving dermatological research (Salmoni, et al, under review) it is argued that an 80% consensus is sufficient to establish an acceptable value for analysis purposes. Hence, following this rule-of-thumb, we use 80% as a cut-off indicator and it is thus apparent that distribution 1 in Table 6 has categories SA and A well within the acceptable range. Distribution 2 also shows acceptance at SA and A but at lesser values, but distributions 3 and 4 are acceptable only at the SA level. Given the disparity of agreement values one could conclude that distribution 1 is strongest, with distribution 2 in second position. In fact, a metric distance can be calculated between each distribution to show the degree of proximity between individual distributions. Details are found in Tastle and Wierman (2008) but, at the request of one of the reviewers, a short explanation is provided here.

5. Measuring Distance between Agreement Distributions

Given two frequency distributions, F_1 and F_2, for which the agreement distributions, Agt_1 and Agt_2, are calculated using (1), a distance between the distributions can be determined. For each category in each frequency distribution there is a corresponding agreement value (see table 6 for an illustration). The distance is calculated using

$$c_n \sqrt{\sum_i^n (Agt_1^i - Agt_2^i)^2} \qquad (1)$$

where n = the number of categories, c_n is a constant for each n, Agt_1 is one agreement distribution, Agt_2 is the second agreement distribution, Agt^i is the ith category in n. The constant is responsible for giving the maximum separation a value of 1 and a range of distance is limited to the unit interval, [0..1]. In the case of a five category scale, c_n is 0.63612 (see Tastle and Wierman, 2008).

The illustration in Row 1 of Table 7 (the first grayed section) shows the maximum possible distance (this maximum possible distance has not yet been mathematically proven) between two distributions in which the survey participants have chosen extreme positions. Observe that in the first grayed "top" distribution of row 1 has all values at SD while the second "bottom" distribution has all values at SA. By definition, the consensus of each separate distribution is 1.0 because all values for each of these two distributions in Row 1 are confined to only one single category, and when compared it is easy to see that the distance between the two distributions is maximized. We define this maximum distance at 1.0; it is not possible to separate the distributions any further.

	SA	A	N	D	SD	Dist
1	0	0	0	0	5	1.000
	5	0	0	0	0	
2	0	0	0	1	4	0.957
	5	0	0	0	0	
3	0	0	1	1	3	0.864
	5	0	0	0	0	
4	0	0	1	2	2	0.830
	5	0	0	0	0	
5	0	0	1	2	2	0.777
	4	1	0	0	0	
6	0	0	1	2	2	0.670
	3	1	1	0	0	
7	0	0	1	2	2	0.512
	2	1	1	1	0	
8	0	0	1	2	2	0.283
	0	1	2	2	0	
9	0	0	1	2	2	0.063
	0	0	2	1	1	
10	0	0	1	2	2	0.000
	0	0	1	2	2	

Table 7. Illustration of the agreement distance from maximum disagreement (frequency pairs shown in row 1) to minimal disagreement (row 10).

It is the agreement distribution for each of these top and bottom distributions that are compared and a distance calculated. Thus, the agreement distances become smaller as the agreement values become equal (see rows 8-10). It is important to understand that actual frequency values are not compared, rather, the agreement measure calculated on those frequencies. This permits us to calculate a distance without regard to the number of items constituting the frequency distribution.

Dist 1 and Dist 2	0.0318
Dist 1 and Dist 3	0.1040
Dist 1 and Dist 4	0.1022
Dist 2 and Dist 3	0.0722
Dist 2 and Dist 4	0.0704
Dist 3 and Dist 4	0.0019

Table 8. Distances between distributions

Returning to our data in Table 6, the information distances between distributions are shown in Table 8. More information on calculating a distance between distributions is available in Tastle and Wierman (2008).

It is apparent that the closest distance is between distributions 3 and 4 of Table 6 which is clearly seen by inspection. The furthest distance is between distributions 1 and 3, with a close second between distributions 1 and 4. It should be recalled that the mean for all of these distributions is essentially identical, so any attempt to identify a similarity ranking using only the mean is circumspect. Using the agreement distance, a similarity ordering can be justified.

6. Conclusion

In examining the student assessment instruments from ten colleges/universities in the eastern portion of the USA it is apparent that each institution's data needs are vastly different. Analysis of the individual instruments shows an emphasis on certain items of interest, the most significant ones centering on the student perceived instructor effectiveness. The 10 instruments were classified into a set of major and minor categories

to create a frequency distribution. From this a probability distribution was calculated and used to identify those categories that were strongest. It was noted that the general categories of instructor scholarship, faculty-student feedback, presentation of material, and overall effectiveness dominated the assessment instruments. In the subcategories the items of dominance are knowledge and competence, helpfulness, organization, and enjoying teaching.

Since these assessment forms are used to evaluate teaching skill and to render assistance in tenure and promotion decisions, an argument is made, via illustration, that the use of the "mean" for measuring averageness is conceptually inappropriate and invalid. Since students assessment according to a Likert scale of defined categories (strongly agree to strongly disagree), and there is no evidence to suggest that the scale of categories is uniform, applying an interval measure lacks suitability, though it is commonly used because of its simplicity of calculation. Another set of measures is suggested as being far more meaningful and conceptually sound: consensus and agreement.

7. Future Research

Given the interesting results of this paper, additional research involving a representative sample of assessment instruments from IS/IT departments across the country should be undertaken with the goal of identifying those qualities that yield the strongest teachers. Such information would be of significant benefit to all IS/IT instructors. Further, guidance in how to compare and properly assess the instruments would benefit the discipline.

8. References Cited

Aasheim, C, J A Gowan, and H Reichgelt. Establishing an Assessment Process for a Computing Program. In *The Proceedings of ISECON 2006*, v 23 (Dallas): §3143. ISSN: 1542-7382. (Revised in *Information Systems Education Journal* 5(1). ISSN: 1545-679X.)

Abdullat, A A and N Terry. Assessing the Effectiveness of Virtual Learning in a Graduate Course in Computer Informa-

tion Systems. In *The Proceedings of ISECON 2004*, v 21 (Newport): §2112. ISSN: 1542-7382. (Revised in *Information Systems Education Journal 3(34).* ISSN: 1545-679X.)

Aleamoni, LM(1987), Typical faculty concerns about student evaluation of teaching. In: Aleamoni LM. , editor. Techniques for evaluations and improving instruction. New directions for teaching and learning, no. 31. San Francisco: Jossey-Bass.

Amoroso, D L. Use of Online Assessment Tools to Enhance Student Performance in Large Classes. In *The Proceedings of ISECON 2004*, v 21 (Newport): §3142. ISSN: 1542-7382. (Revised in *Information Systems Education Journal 3(4).* ISSN: 1545-679X.)

Baugh, J M. Assessment of Spreadsheet and Database Skills in the Undergraduate Student. In *The Proceedings of ISECON 2003*, v 20 (San Diego): §2111. ISSN: 1542-7382. (Revised in *Information Systems Education Journal 2(30).* ISSN: 1545-679X.)

Ceccucci, W. An Alternative Testing Strategy for Advanced Programming Courses. In *The Proceedings of ISECON 2005*, v 22 (Columbus OH): §2343. ISSN: 1542-7382. (Revised in *Information Systems Education Journal 4(89).* ISSN: 1545-679X.)

Dettori, L, T A Steinbach, and M Kalin. Is this Course Right for You? Using Self-Tests for Student Placement. In *The Proceedings of ISECON 2005*, v 22 (Columbus OH): §2312. ISSN: 1542-7382. (Revised in *Information Systems Education Journal 4(77).* ISSN: 1545-679X.)

Hernández, L. O., Wetherby, K., and Pegah, M. 2004. Dancing with the devil: faculty assessment process transformed with web technology. In *Proceedings of the 32nd Annual ACM SIGUCCS Conference on User Services* (Baltimore, MD, USA, October 10 - 13, 2004). SIGUCCS '04. ACM, New York, NY, 60-65. DOI= http://doi.acm.org/10.1145/1027802.10 27818

Joseph, P A. Ethics in the Pedagogy of Information Systems. In *The Proceedings of ISECON 2005*, v 22 (Columbus OH): §3545. ISSN: 1542-7382. (Revised in *Information Systems Education Journal 5(23).* ISSN: 1545-679X.)

Landry, J P, J H Pardue, H E Longenecker, J H Reynolds, L J McKell, and B A White. Using the IS Model Curriculum and CCER Exit Assessment Tools for Course-level Assessment. In *The Proceedings of ISECON 2005*, v 22 (Columbus OH): §2123. ISSN: 1542-7382. (Revised in *Information Systems Education Journal 4(73).* ISSN: 1545-679X.)

McDonald, D S and R D Johnson. Grade Distribution and Its Impact on CIS Faculty Evaluations: 1992-2002. In *The Proceedings of ISECON 2003*, v 20 (San Diego): §2244. ISSN: 1542-7382. (Revised in *Information Systems Education Journal 1(42).* ISSN: 1545-679X.)

McGinnis, D R and G J Slauson. Advancing Local Degree Programs Using the IS Model Curriculum. In *The Proceedings of ISECON 2003*, v 20 (San Diego): §2133. ISSN: 1542-7382. (Revised in *Information Systems Education Journal 1(37).* ISSN: 1545-679X.)

McKeachie, W. J., (1997), "Student Ratings: The Validity of Use." American Psychologist, 52(11), pp. 1218-1225.

McKell, L J, J H Reynolds, H E Longenecker, J P Landry, and J H Pardue. The Center for Computing Education Research (CCER): A Nexus for IS Institutional and Individual Assessment. In *The Proceedings of ISECON 2005*, v 22 (Columbus OH): §2122. ISSN: 1542-7382. (Revised in *Information Systems Education Journal 4(69).* ISSN: 1545-679X.)

O'Neil, T D. The Effective Use of Web-based Training and Assessment in a Computer Literacy Course. In *The Proceedings of ISECON 2005*, v 22 (Columbus OH): §2532. ISSN: 1542-7382. (Revised in *Information Systems Education Journal 4(106).* ISSN: 1545-679X.)

Paranto, S and L Shillington. Is it Possible to Assess Information Systems Skills using a Multiple-Choice Exam? In *The Pro-*

ceedings of ISECON 2005, v 22 (Columbus OH): §3532. ISSN: 1542-7382. (Revised in *Information Systems Education Journal* 4(24). ISSN: 1545-679X.)

Reynolds, J H, H E Longenecker, J P Landry, J H Pardue, and B Applegate. Information Systems National Assessment Update: The Results or a Beta Test of a New Information Systems Exit Exam Based on the IS 2002 Model Curriculum. In *The Proceedings of ISECON 2003*, v 20 (San Diego): §3415. ISSN: 1542-7382. (Revised in *Information Systems Education Journal* 2(24). ISSN: 1545-679X.)

Salmoni, A. J, S. Coxall, M. Gonzalez, W. Tastle, and A. Y. Finley, (under review), " Defining a postgraduate curriculum in dermatology for general practitioners: a needs analysis using a modified Delphi method."

Stemler, L and C Chamblin. The Role of Assessment in Accreditation: A Case Study for an MIS Department. In *The Proceedings of ISECON 2005*, v 22 (Columbus OH): §3565. ISSN: 1542-7382. (Revised in *Information Systems Education Journal* 4(39). ISSN: 1545-679X.)

Tastle, W. and M. Wierman, 2005a, "Consensus and Dissention: Theory and Properties." North American Fuzzy Information Processing Society (NAFIPS) Conference, Ann Arbor, MI.

Tastle, W. and M. Wierman, 2005b, "A Tool for the Analysis of Ordinal Scale Sate: Measuring Consensus, Agreement, and Dissent." 5th International Conference on Methods and Techniques in Behavioral Research, Wageningen, The Netherlands.

Tastle, W. J. and M. J. Wierman, 2007a, "Using Consensus to Measure Weighted Targeted Agreement." NAFIPS 2007 Conference Proceedings.

Tastle, W. and M. Wierman, 2007b, "Determining Risk Assessment Using the Weighted Ordinal Agreement Measure." Journal of Homeland Security, http://www. homelandsecurity.org/new-journalArticles/displayArticle2.asp?article =157, June 2007.

Tastle, W. J. and M. J. Wierman, 2008, "Agreement, Agreement Distributions, and Distance." North American Fuzzy Information Processing Society (NAFIPS) Conference Proceedings, New York City, NY.

Todorova, N and A Mills. Development of Assessment Portfolios for IS Majors. In *The Proceedings of ISECON 2004*, v 21 (Newport): §4112. ISSN: 1542-7382. (Revised in *Information Systems Education Journal* 5(25). ISSN: 1545-679X.)

University of New South Wales, Faculty of Business, School of Information Systems, Technology and Management, 2007, "INFS3605, Project Workshop." Last viewed: 29 June 2008, URL: http://wwwdocs.fce.unsw.edu.au/sistm/current/CourseOutlines/2007S1INFS3605CourseOutline.pdf

Virginia Polytechnic Institute and State University, Pamplin College of Business, 2001, "Schev Outcomes Assessment Report." Last viewed 29 June 2008, URL: http://www.aap.vt.edu/department%20reports/Business%20Info%20Technology%20-%20grad.htm

White, B A and R V McCarthy. The Development of a Comprehensive Assessment Plan: One Campus' Experience. In *The Proceedings of ISECON 2007*, v 24 (Pittsburgh): §3524. ISSN: 1542-7382. (Revised in *Information Systems Education Journal* 5(35). ISSN: 1545-679X.)

Appendices

Criteria	I	II	III	IV	V	VI	VII	VIII	IX	X
A. Instructor Evaluation										
1. organization and planning	2	2	1	1		1				
2. scholarship										
a) knowledge and competence		2	1	1	4		1		1	
b) exams give balanced coverage					1		1			
c) fair and impartial grading	1					1	1	1		1
d) course is challenging			2	1				3		1
3. faculty/student interactions										
a) student feedback	1					1	1	1		1
b) level of concern	1		1				1	3		
c) understandable	1		2	1	1	1		3	1	
d) helpfulness	2		1	2	1			1	3	1
4. Presentation of material										
a) organized	1	1	1	1			2	3	1	1
b) enjoys teaching	2	1	1	2				3	1	1
c) uses various teaching methods	1		1	2				2	1	
d) encourages participation			1	1	1			2		
e) summarizes major points								1		
5. Overall effectiveness										
a) comparison with other instructors/courses	1	1		1		2	1	2	1	
b) objectives are achieved	3		2	1			1			
c) enhanced interest in subject	1	1		1		2				
B. Evaluation of student learning										
1. explanation of grading system & assignments	2		2			1		1	1	1
2. tests & assignments reflected course content			2				1	1	1	
3. assignments contributed to understanding course		1	2	1		1		1		
4. grading returned in reasonable amt of time	1		1					1		
5. class attendance necessary to learn material					1					
C. Delivery media and facilities										
1. software enhanced learning									1	
2. hardware enhanced learning									1	
3. video enhanced learning									1	
4. audio enhanced learning									1	
5. room enhanced learning			3						1	
6. readings enhanced learning			2							
D. General relevance										
1. prepared for future professional success									1	
2. relevance to rest of discipline									1	
3. relevance to rest of business core									1	
4. relevance to general ed courses									1	

Institutions

Table 1. Assignment of institutional assessment forms into the criteria matrix. Each column represents a frequency distribution.

Normalized Data											
Criteria	I	II	III	IV	V	VI	VII	VIII	IX	X	Total
A1	0.100	0.133	0.048	0.063	0	0.100	0	0	0	0	0.443
A2a	0	0.133	0.048	0.063	0.500	0	0.100	0	0.050	0	0.893
A2b	0	0	0	0	0.125	0	0.100	0	0	0	0.225
A2c	0.050	0	0	0	0	0.100	0.100	0.034	0	0.143	0.427
A2d	0	0	0.095	0.063	0	0	0	0.103	0	0.143	0.404
A3a	0.050	0	0	0	0	0.100	0.100	0.034	0	0.143	0.427
A3b	0.050	0	0.048	0	0	0	0.100	0.103	0	0	0.301
A3c	0.050	0	0.095	0.063	0.125	0.100	0	0.103	0.050	0	0.586
A3d	0.100	0	0.048	0.125	0.125	0	0	0.034	0.150	0.143	0.725
A4a	0.050	0.067	0.048	0.063	0	0	0.200	0.103	0.050	0.143	0.723
A4b	0.100	0.067	0.048	0.125	0	0	0	0.103	0.050	0.143	0.636
A4c	0.050	0.000	0.048	0.125	0	0	0	0.069	0.050	0	0.342
A4d	0	0.067	0.048	0.063	0	0	0	0.069	0	0	0.246
A4e	0	0	0	0	0	0	0	0.034	0	0	0.034
A5a	0.050	0.067	0	0.063	0	0.200	0.100	0.069	0.050	0	0.598
A5b	0.150	0.000	0.095	0.063	0	0.000	0.100	0	0	0	0.408
A5c	0.050	0.067	0	0.063	0	0.200	0	0	0	0	0.379
B1	0.100	0	0.095	0	0	0.100	0	0.034	0.050	0.143	0.523
B2	0	0	0.095	0	0	0	0.100	0.034	0.050	0	0.280
B3	0	0.067	0.095	0.063	0	0.100	0	0.034	0	0	0.359
B4	0.050	0	0.048	0	0	0	0	0.034	0	0	0.132
B5	0	0	0	0	0.125	0	0	0	0	0	0.125
C1	0	0	0	0	0	0	0	0	0.050	0	0.050
C2	0	0	0	0	0	0	0	0	0.050	0	0.050
C3	0	0	0	0	0	0	0	0	0.050	0	0.050
C4	0	0	0	0	0	0	0	0	0.050	0	0.050
C5	0	0.200	0	0	0	0	0	0	0.050	0	0.250
C6	0	0.133	0	0	0	0	0	0	0	0	0.133
D1	0	0	0	0	0	0	0	0	0.050	0	0.050
D2	0	0	0	0	0	0	0	0	0.050	0	0.050
D3	0	0	0	0	0	0	0	0	0.050	0	0.050
D4	0	0	0	0	0	0	0	0	0.050	0	0.050

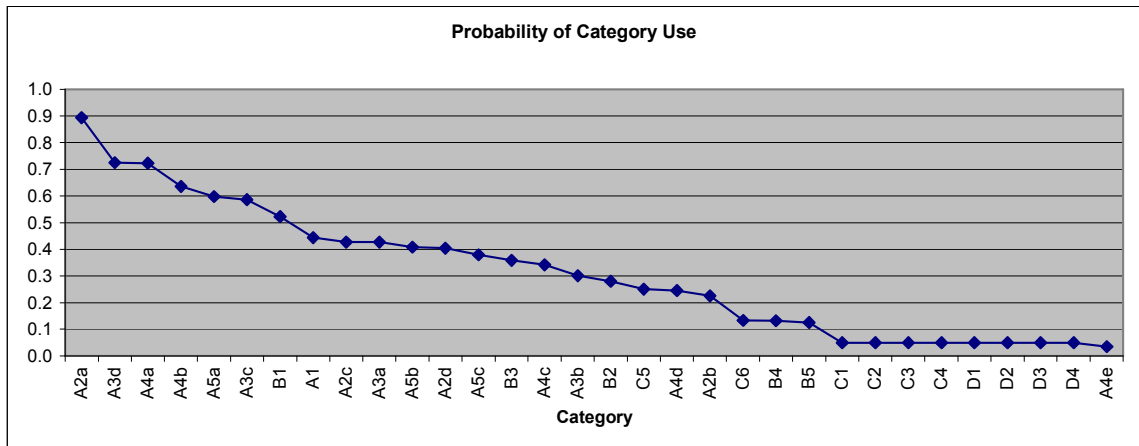**Table 2.** Normalization of table 1. Each column represents a probability distribution.

Figure 1. Degree of usage of evaluation categories by order of probability.

Hierarchy

Rank	Criteria	Rank	Criteria	Rank	Criteria	Rank	Criteria
1	A2a	9	A2c	17	B2	24	C2
2	A3d	9	A3a	18	C5	24	C3
3	A4a	11	A5b	19	A4d	24	C4
4	A4b	12	A2d	20	A2b	24	D1
5	A5a	13	A5c	21	C6	24	D2
6	A3c	14	B3	22	B4	24	D3
7	B1	15	A4c	23	B5	24	D4
8	A1	16	A3b	24	C1	32	A4e

Table 3. Ranked criteria according to probability of use.

		SA	A	N	D	SD
1		63	29	8	0	0
		0.911	0.863	0.683	0.437	0.140
2		102	29	8	0	5
		0.903	0.829	0.653	0.419	0.132
3		79	0	0	0	10
		0.888	0.753	0.585	0.376	0.112
4		161	0	1	0	20
		0.888	0.754	0.587	0.378	0.113

Table 6. Agreement values for each distribution of Table 4.