

SITE Review Committee Report, 3/27/08
Comparison of Instruments for Student Evaluation of Teaching Effectiveness

At its last meeting, the SITE Review Committee evaluated commercially available student rating forms. Based on this evaluation, the committee recommends that WKU adopt the Student Instructional Report II (SIR II) system that was developed by the Educational Testing Service (ETS). In this report, we present the results of our comparison of the SIR II with the current WKU SITE system.

	<i>Validity</i>	<i>Reliability</i>	<i>National Comparative Data</i>	<i>Local Comparative Data</i>	<i>Comprehensive & Timely Reports</i>	<i>Variety of Delivery Methods</i>	<i>Custom Questions</i>	<i>Written Comments</i>	<i>Cost/Semester</i>
SIR II	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	\$38,760
SITE	No	No	No	Yes	No	No	Limited	Yes	\$14,546

	<i>Preparation & Delivery via Institutional Research</i>	<i>Confidentiality</i>	<i>Scheduling Outside Normal Window</i>	<i>Coefficient Alpha</i>
SIR II	Yes	Yes	Yes	Yes
SITE	Yes	Yes	Yes	Yes

Validity and Reliability:

The SIR II has undergone extensive testing for validity and reliability. This is not the case for the WKU SITE (see previous SITE Review Committee report). The SIR II was originally developed in 1972 and then revised in 1995. The revision includes five dimensions of teaching effectiveness from the original SIR (with some items removed and some added) as well as three new dimensions reflecting more recent emphases in higher education. The specific dimensions measured include course organization and planning, communication, faculty/student interaction, assignments, exams, and grading, supplementary instructional methods, course outcomes, student effort and involvement, course difficulty and pace. There are 3 or more items for each of these dimensions. There is also an item reflecting the overall evaluation of the course. The WKU SITE has fewer dimensions of teaching effectiveness and for those that are included, there is only one item representing the dimension. Pretesting of the SIR II revision involved administration of two forms of the survey with different response formats at 10 two and four year institutions. Item and scale analyses of these data

were conducted to determine the final version of the form, which was then piloted at a variety of colleges over the period of a year. Pilot data were used to determine content validity and reliability (internal consistency, number of students needed for consistency, stability of responses over brief periods of time). Technical reports with further details on the development and validation of this instrument are available from ETS.

National and Local Comparative Data, Comprehensive & Timely Reports: ETS will process SIR II reports within 15 days of receipt of completed surveys. Three copies of the report are provided for each class evaluated, provided there are at least 5 students responding to the survey. (Reliability cannot be guaranteed with fewer than 5 students responding). These reports include average scores for each survey item as well as an overall mean and a comparative mean for each dimension of teaching effectiveness. The comparative mean for each dimension is based on means from similar courses (i.e., same level, type, and subject) in similar institutions that use the SIR II. In addition, items within each dimension are flagged with a + if they are reliably at or above the 90th percentile and with a – if they are reliably at or below the 10th percentile of comparative data. Local comparative data (e.g., institutional, departmental, program) are also available eliminating the need for calculating independent confidence intervals for within department, college, and university comparisons. Combined reports for class, department, or institution as well as data diskette/CDs with item level responses for all classes can be requested from ETS. The WKU SITE report provides statistical information for each item, but because the instrument is used only at WKU, the report cannot provide national comparative data. Moreover, the reliability of the data is not considered -- reports are provided and are used for promotion and tenure decisions even when there are few students responding.

In the SIR II, information on factors that might impact learning (e.g., student effort and involvement; required vs. elective course), principal type of student in the course (e.g., lower or upper division, majors/non-majors, graduate) is also collected and reported. These data are not collected or reported for the WKU SITE. In addition, a narrative explaining how to interpret the SIR II data is provided in the report and there are published guidelines for the use of the data. No guidance is provided for interpretation of the WKU SITE data and there are no formal guidelines for how the data should be used.

Delivery Methods: The SIR II system offers a variety of electronic survey methods (e.g., Blackboard, eCollege, email with unique URL, unique course URL posted on course web page) and a special version of the survey is available for distance learning courses. It is likely that one or more of these methods would interface with Banner.

Custom Questions, SGA Questions, Written Comments: Institution, college, department, or instructor can add up to 10 questions to the SIR II. SGA questions could be administered on the same form and at the same time as the primary SITE. Item level responses for all

classes are available on data diskette/CD for the SIR II, which would allow Institutional Research to prepare separate reports for SGA questions. The SIR II also has space for written comments, and ETS provides transcription and incorporation of the comments into the instructor's report.

Cost: With a few exceptions, all instructors are currently evaluated with the SITE in each class they teach in both Fall and Spring semesters. Institutional Research prepares and processes approximately 3,058 SITE course evaluation packets per semester (~85,000 paper and pencil forms) for a total cost of approximately \$14,546. If the same number of SIR II paper and pencil surveys were administered and processed, the total cost would be approximately \$38,760. The additional cost of this survey could be offset by developing an alternative administration procedure whereby a subset of an instructor's courses is selected for evaluation each semester and the courses of pre-tenure faculty are evaluated more often than those of post-tenure faculty. This would also ensure that students do not have to complete the same evaluation form multiple times during the SITE period, which would improve both completion rates and the reliability of the data.

Preparation, Delivery, and Administration: The preparation and delivery of course packets and instructor reports for the WKU SITE is coordinated through Institutional Research. There would be no change in this procedure with the SIR II system because ETS requires an on-campus coordinator for distribution of the survey and the instructor reports. Department heads could continue to edit the list of courses/sections that are scheduled to receive SITE materials. Likewise, administration procedures would change very little though, as noted above, the SIR II offers greater flexibility in the method of delivery. The SIR II, like the WKU SITE takes about 15 minutes to complete.

Confidentiality: With the present SITE, the task of ensuring confidentiality resides with Institutional Research. There would be little change in this task with adoption of the SIR II as Institutional Research would maintain control over the evaluation process as described above.

Scheduling: Scheduling outside the normal administration window is possible with the SIR II as the preparation and delivery of the evaluation packets would be determined jointly by Institutional Research and the various departments.

Coefficient Alpha: If necessary, coefficient alpha reliability estimates using matched pairs of instructor courses could be computed by Institutional Research from the data provided in the SIR II data diskette/CDs. Of course, these estimates would not be available until an instructor had been evaluated with the SIR II at least twice for the same course. However, given that the reliability of the SIR II has been established empirically, it is not critical that these estimates be calculated for the SIR II.

Sources:

Arreola, R. A. (2000). *Developing a comprehensive faculty evaluation system 2/e*. Bolton, MA: Anker Publishing Co. Inc.
B. Cobb (personal communication, 2/26/08).

SITE Review Committee:

Dr. Mel Borland	Gordon Ford College of Business
Dr. Marilyn Gardner	College of Health and Human Services
Dr. Steve Haggbloom	Council of Department Heads
Skyler Jordan	SGA
Dr. Lora Moore	Bowling Green Community College
Dr. Sharon Mutter	University Senate
Dr. Les Pesterfield	Ogden College of Science and Engineering
Dr. John Faine	Potter College of Arts and Letters
Dr. Steve Wininger	College of Education and Behavioral Science

Figure 15.5 Student Instructional Report II Form (front side)



STUDENT INSTRUCTIONAL REPORT II (SIR II)

SIR II Report Number

--	--	--	--	--

This questionnaire gives you the chance to comment anonymously about this course and the way it was taught. Using the rating scale below, mark the one response for each statement that is closest to your view. Fill in the appropriate circle to the right of the statement.

(5) Very Effective
 (4) Effective
 (3) Moderately Effective
 (2) Somewhat Ineffective
 (1) Ineffective
 (0) Not applicable, not used in the course, or you don't know. In short, the statement does not apply to the course or instructor.

As you respond to each statement, think about each practice as it contributed to your learning in this course.

A. Course Organization and Planning

1. The instructor's explanation of course requirements
2. The instructor's preparation for each class period
3. The instructor's command of the subject matter
4. The instructor's use of class time
5. The instructor's way of summarizing or emphasizing important points in class

B. Communication

6. The instructor's ability to make clear and understandable presentations
7. The instructor's command of spoken English (or the language used in the course)
8. The instructor's use of examples or illustrations to clarify course material
9. The instructor's use of challenging questions or problems
10. The instructor's enthusiasm for the course material

C. Faculty/Student Interaction

11. The instructor's helpfulness and responsiveness to students
12. The instructor's respect for students
13. The instructor's concern for student progress
14. The availability of extra help for this class (taking into account the size of the class)
15. The instructor's willingness to listen to student questions and opinions

D. Assignments, Exams, and Grading

16. The information given to students about how they would be graded
17. The clarity of exam questions
18. The exams' coverage of important aspects of the course
19. The instructor's comments on assignments and exams
20. The overall quality of the textbook(s)
21. The helpfulness of assignments in understanding course material

E. Supplementary Instructional Methods

Many different teaching practices can be used during a course. In this section (E), rate only those practices that the instructor included as part of this course.

Rate the effectiveness of each practice used as it contributed to your learning.

22. Problems or questions presented by the instructor for small group discussions
23. Term paper(s) or project(s)
24. Laboratory exercises for understanding important course concepts
25. Assigned projects in which students worked together
26. Case studies, simulations, or role playing
27. Course journals or logs required of students
28. Instructor's use of computers as aids in instruction

Very Effective	Effective	Moderately Effective	Somewhat Ineffective	Ineffective	Not applicable
5	4	3	2	1	0

Questionnaire continued on the other side. ➡

Copyright © 1985 by Educational Testing Service. All rights reserved. No part of the Student Instructional Report II (SIR II) 57206 • 12402 • CV118M400X may be adapted or reproduced in any form without permission in writing from the publisher. Printed in the U.S.A. CHW95222 Q1057-0 I.N.401651 4

Figure 15.5 (continued) Student Instructional Report II Form (back side)

For the next two sections (F and G), use the rating scale below. Mark the one response for each statement that is closest to your view. Fill in the appropriate circle to the right of each statement.

- (5) **Much More** than most courses
 (4) **More Than** most courses
 (3) **About the Same** as others
 (2) **Less** than most courses
 (1) **Much Less** than most courses
 (0) **Not Applicable**, not used in the course, or you don't know. In short, the statement does not apply to the course or instructor.

Much More than most courses
 More Than most courses
 About the Same as others
 Less than most courses
 Much Less than most courses
 Not Applicable

F. Course Outcomes

29. My learning increased in this course
 30. I made progress toward achieving course objectives
 31. My interest in the subject area has increased
 32. This course helped me to think independently about the subject matter
 33. This course actively involved me in what I was learning

(5) (4) (3) (2) (1) (0)
 (5) (4) (3) (2) (1) (0)
 (5) (4) (3) (2) (1) (0)
 (5) (4) (3) (2) (1) (0)
 (5) (4) (3) (2) (1) (0)

G. Student Effort and Involvement

34. I studied and put effort into the course
 35. I was prepared for each class [writing and reading assignments]
 36. I was challenged by this course

(5) (4) (3) (2) (1) (0)
 (5) (4) (3) (2) (1) (0)
 (5) (4) (3) (2) (1) (0)

H. Course Difficulty, Work Load, and Pace

37. For my preparation and ability, the level of difficulty of this course was:

- (5) Very difficult (4) Somewhat difficult (3) About right (2) Somewhat elementary (1) Very elementary

38. The work load for this course in relation to other courses of equal credit was:

- (5) Much heavier (4) Heavier (3) About the same (2) Lighter (1) Much lighter

39. For me, the pace at which the instructor covered the material during the term was:

- (5) Very fast (4) Somewhat fast (3) Just about right (2) Somewhat slow (1) Very slow

I. Overall Evaluation

40. Rate the quality of instruction in this course as it contributed to your learning (try to set aside your feelings about the course content):

- (5) Very effective (4) Effective (3) Moderately effective (2) Somewhat ineffective (1) Ineffective

J. Student Information

41. Which one of the following best describes this course for you?

- (1) A major/minor requirement (2) A college requirement (3) An elective (4) Other

42. What is your class level?

- (1) Freshman/1st year (2) Sophomore/2nd year (3) Junior/3rd year (4) Senior/4th year (5) Graduate (6) Other

43. Do you communicate better in English or in another language?

- (1) Better in English (2) Better in another language (3) Equally well in English and another language

44. Sex

- (1) Female (2) Male

45. What grade do you expect to receive in this course?

- (1) A (2) A- (3) B+ (4) B (5) B- (6) C (7) Below C

K. Supplementary Questions

If the instructor provided supplementary questions and response options, mark your answers in this section. Mark only one response for each question.

46. (5) (4) (3) (2) (1) (NA) 48. (5) (4) (3) (2) (1) (NA) 50. (5) (4) (3) (2) (1) (NA) 52. (5) (4) (3) (2) (1) (NA) 54. (5) (4) (3) (2) (1) (NA)
 47. (5) (4) (3) (2) (1) (NA) 49. (5) (4) (3) (2) (1) (NA) 51. (5) (4) (3) (2) (1) (NA) 53. (5) (4) (3) (2) (1) (NA) 55. (5) (4) (3) (2) (1) (NA)

L. Student Comments

If you would like to make additional comments about the course or instruction, use a separate sheet of paper. You might elaborate on the particular aspects you liked most as well as those you liked least. Also, how can the course or the way it was taught be improved? An additional form may be provided for your comments. Please give these comments to the instructor.

If you have any comments about this questionnaire, please send them to:
 Student Instructional Report II, Educational Testing Service, Princeton, NJ 08541-0001.

Figure 15.6 SIR II Sample Class Report

STUDENT INSTRUCTIONAL REPORT II

Enrollment	Admin. Date	Report No.	Batch No.
33	00/00	00000	0000



CLASS REPORT SAMPLE

Assessing Courses and Instruction

PERCENTAGES reported below are based on the total number responding, which is: 33

A. Course Organization and Planning <i>Think about each practice as it contributed to your learning in this course.</i>	Omit	Not Applicable	5 Very Effective	4 Effective	3 Moderately Effective	2 Somewhat Ineffective	1 Ineffective	Mean
1. The instructor's explanation of course requirements . . .			39	42	18			4.21
2. The instructor's preparation for each class period . . .			42	42	9	6		4.21
3. The instructor's command of the subject matter . . .			39	42	12	6		4.15
4. The instructor's use of class time . . .			42	30	18	6	3	4.03
5. The instructor's way of summarizing or emphasizing important points in class . . .			27	45	6	18	3	3.76
Overall mean for COURSE ORGANIZATION AND PLANNING is: 4.07 The comparative mean for X-year institutions is: x.xc.								

B. Communication <i>Think about each practice as it contributed to your learning in this course.</i>	Omit	Not Applicable	5 Very Effective	4 Effective	3 Moderately Effective	2 Somewhat Ineffective	1 Ineffective	Mean
6. The instructor's ability to make clear and understandable presentations . . .			30	36	27	6		3.91
7. The instructor's command of spoken English (or the language used in the course) . . .			61	36	3			4.58
8. The instructor's use of examples or illustrations to clarify course material . . .			33	33	30	3		3.97
9. The instructor's use of challenging questions or problems . . .			30	39	27	3		3.97
10. The instructor's enthusiasm for the course material . . .			21	45	30	3		3.85
Overall mean for COMMUNICATION is: 4.06 The comparative mean for X-year institutions is: x.xc.								

C. Faculty/Student Interaction <i>Think about each practice as it contributed to your learning in this course.</i>	Omit	Not Applicable	5 Very Effective	4 Effective	3 Moderately Effective	2 Somewhat Ineffective	1 Ineffective	Mean
11. The instructor's helpfulness and responsiveness to students . . .			39	33	15	12		4.00
12. The instructor's respect for students . . .			45	24	18	9	3	4.00
13. The instructor's concern for student progress . . .			36	33	15	15		3.91
14. The availability of extra help for this class (taking into account the size of the class) . . .			36	33	21	9		3.97
15. The instructor's willingness to listen to student questions and opinions . . .			39	36	12	6	6	3.97
Overall mean for FACULTY/STUDENT INTERACTION is: 3.97 The comparative mean for X-year institutions is: x.xc.								

+ This mean is higher than the comparative mean. See page 4
 - This mean is lower than the comparative mean. See page 4

For explanation of flagging (*), see "Number of Students Responding," page 4

Figure 15.6 (continued) SIR II Sample Class Report

STUDENT INSTRUCTIONAL REPORT II								
D. Assignments, Exams, and Grading Think about each practice as it contributed to your learning in this course.								
	Omit	Not Applicable	5 Very Effective	4 Effective	3 Moderately Effective	2 Somewhat Ineffective	1 Ineffective	Mean
16. The information given to students about how they would be graded			45	39	15			4.30
17. The clarity of exam questions			36	33	21	9		3.97
18. The exams' coverage of important aspects of the course			33	42	18	6		4.03
19. The instructor's comments on assignments and exams			27	39	21	12		3.82
20. The overall quality of the textbook(s)		6	12	45	15	15	6	3.45
21. The helpfulness of assignments in understanding course material			27	48	21	3		4.00
Overall mean for ASSIGNMENTS, EXAMS, AND GRADING is: 3.93 The comparative mean for X-year institutions is: x.xx								
E. Supplementary Instructional Methods Rate the effectiveness of each practice used as it contributed to your learning.								
	Omit	Not Used	5 Very Effective	4 Effective	3 Moderately Effective	2 Somewhat Ineffective	1 Ineffective	Mean
22. Problems or questions presented by the instructor for small group discussions	3	3	21	64	9			***
23. Term paper(s) or project(s)			30	58	12			***
24. Laboratory exercises for understanding important course concepts	12	52	18	9	9			***
25. Assigned projects in which students worked together			33	48	12		6	***
26. Case studies, simulations, or role playing			33	55	12			***
27. Course journals or logs required of students	9	64	9	9	6	3		***
28. Instructor's use of computers as aids in instruction	9	70	9	12				***
Means are not reported (***) for SUPPLEMENTARY INSTRUCTIONAL METHODS.								
F. Course Outcomes Mark the response that is closest to your view.								
	Omit	Not Applicable	5 Much More Than Most Courses	4 More Than Most Courses	3 About the Same as Others	2 Less Than Most Courses	1 Much Less Than Most Courses	Mean
29. My learning increased in this course	3		15	33	33	12	3	3.47
30. I made progress toward achieving course objectives	3		18	30	45	3		3.66
31. My interest in the subject area has increased	3		15	24	33	15	9	3.22
32. This course helped me to think independently about the subject matter	3		24	18	45	9		3.59
33. This course actively involved me in what I was learning	3		27	33	30	6		3.84
Overall mean for COURSE OUTCOMES is: 3.56 The comparative mean for X-year institutions is: x.xx								
G. Student Effort and Involvement Mark the response that is closest to your view.								
	Omit	Not Applicable	5 Much More Than Most Courses	4 More Than Most Courses	3 About the Same as Others	2 Less Than Most Courses	1 Much Less Than Most Courses	Mean
34. I studied and put effort into this course	3		45	12	36	3		4.03
35. I was prepared for each class (writing and reading assignments)	3		33	27	30	3	3	3.88
36. I was challenged by this course	3		33	12	36	9	6	3.59
Overall mean for STUDENT EFFORT AND INVOLVEMENT is: 3.83 The comparative mean for X-year institutions is: x.xx								
+ This mean is higher than the comparative mean. See page 4. - This mean is lower than the comparative mean. See page 4. For explanation of flagging (*), see "Number of Students Responding," page 4.								

Figure 15.6 (continued) SIR II Sample Class Report

ASSESSING COURSES and INSTRUCTION							
H. Course Difficulty, Workload, and Pace Mark the response that is closest to your view.	Omit	Very Difficult	Somewhat Difficult	About Right	Somewhat Elementary	Very Elementary	
37. For my preparation and ability, the level of difficulty of this course was . . .	3	6	42	42	3	3	
	Omit	Much Heavier	Heavier	About the Same	Lighter	Much Lighter	
38. The work load for this course in relation to other courses of equal credit was . . .	3	55	24	15	3		
	Omit	Very Fast	Somewhat Fast	Just About Right	Somewhat Slow	Very Slow	
39. For me, the pace at which the instructor covered the material during the term was . . .	3	6	33	55	3		
Means are not appropriate for COURSE DIFFICULTY, WORKLOAD, and PACE. Review the distribution of students' responses.							
I. Overall Evaluation	Omit	5 Very Effective	4 Effective	3 Moderately Effective	2 Somewhat Ineffective	1 Ineffective	
40. Rate the quality of instruction in this course as it contributed to your learning. (Try to set aside your feelings about the course content.)	3	18	52	21	6	Mean 3.84	
OVERALL EVALUATION mean is: 3.84							
J. Student Information	Omit	Requirement in Major		College Requirement	Effective	Other	
41. Which one of the following best describes this course for you?	3	76		18	3		
42. What is your class level?	Omit	Freshman/ 1st Year	Sophomore/ 2nd Year	Junior/ 3rd Year	Senior/ 4th Year	Graduate	Other
	3			3	85	6	3
43. Do you communicate better in English or in another language?	Omit	Better in English		Better in Another Language	Equally well in English and Another Language		
	3	91		3	3		
44. Sex	Omit	Female			Male		
	3	52			45		
45. What grade do you expect to receive in this course?	Omit	A	A-	B+	B	B-	C
	100						
K. Supplementary Questions	Omit	NA	1	1	1	2	1
46.							
47.							
48.							
49.							
50.							
51.							
52.							
53.							
54.							
55.							
+ This mean is higher than the comparative mean. See page 4. - This mean is lower than the comparative mean. See page 4.							
For explanation of flagging (*), see "Number of Students Responding," page 4.							

Arreola, R. A. (2000). *Developing a Comprehensive Faculty Evaluation System 2/e*. Bolton, MA: Anker Publishing Co., Inc.

Figure 15.6 (continued) SIR II Sample Class Report

INTERPRETING SIR II

The SIR II is designed to:

- Identify areas of strength and/or areas for improvement.
- Provide information on new teaching methods or techniques used in class.
- Provide feedback from students about their courses.

NUMBER OF STUDENTS RESPONDING

The number of students responding can affect the results when the class is very small (fewer than 10 students are enrolled), or when fewer than two-thirds of the students enrolled in the class actually respond. For this reason, a Class Report will not be produced when fewer than five students responded, that is, fewer than five completed answer sheets were received for a class.

The degree of accuracy for each item mean increases as the number of students responding increases. For example, the estimated reliability for the Overall Evaluation Item is .78 if 10 students respond; .88 if 20 students respond; and .90 if 25 students respond. (A full discussion of the reliability of student evaluation items can be found in *SIR Report No. 3*.) To call attention to possible reliability concerns, a report will be flagged (*) for one or more of the following.

- * The number responding will be flagged when: 10 or fewer students responded or less than 60 percent of the class responded (this calculation is based on information from the *Instructor's Cover Sheet*).
- * An item mean will not be reported when: 50 percent or more of the students did not respond, or marked an item "Not Applicable," or fewer than five students responded to an item.
- * An overall mean is not reported when one or more item means are not reported.

COMPARATIVE DATA (NOT AVAILABLE FOR SIR II PILOT)

The comparative means used throughout this report are based on user data from a sample of two and four year colleges and universities. An institution is identified by type — two-year or four-year — on the Processing Request Form that is returned with the questionnaires for scoring. Either two-year or four-year comparative data are used, based on that identification.

These data are comparative rather than normative. That is, they are prepared by combining class reports from institutions at which the questionnaire was administered. The data are updated periodically and are developed and published separately for two-year and for four-year institutions in the *Comparative Data Guides*.

The *Comparative Data Guides* for both two- and four-year colleges contain data analyzed for: size of class, level of class (freshman/sophomore and junior/senior), type of class (lecture, discussion, lab), and several different subject areas. A copy of the appropriate *Guide* is sent to Institutional Coordinators with the SIR II reports.

Local Comparative Data: Equally important and useful are an institution's own comparative data. Such local comparative data — e.g., an Institutional Summary, departmental summaries, program summaries — are available to any user institution. Forms for ordering these reports are included in the *Institutional Coordinator's Manual*.

Understanding Mean Ratings

Ratings can vary by class size and discipline. The *Comparative Data Guides* provide data by various categories to assist users in interpreting the SIR II reports. Please refer to the *Guide* and to the SIR II Guidelines for further information. Since student ratings typically tend to be favorable, it is important to have comparative data to interpret a report fully. For example, while a 3.6 is numerically above average on a 5-point scale, it may be average or even slightly below average in comparison to other means for items in SIR II.

What Makes a Score Difference Significant?

The mean scores on all of the items and scales in this report have been compared against the scores obtained by all of the classes in one of the appropriate comparative data groups (two-year or four-year institutions). Specifically, the scores have been compared against the score values corresponding to the 10th percentile and 90th percentile in the comparative group. If the results indicate a score is sufficiently reliable and is below the 10th percentile or above the 90th percentile, it will be flagged in the report as follows:

- + This class mean is reliably at or above the 90th percentile.
- This class mean is reliably at or below the 10th percentile.

Scores above the 90th percentile or below the 10th percentile are flagged when there is appropriate statistical confidence that the "true scores" (i.e., the scores that would be obtained if there were no measurement error) fall within these ranges. If a score is flagged with a +, there is less than one chance in 20 that the "true score" is below the 90th percentile; if a score is flagged with a —, there is less than one chance in 20 that the "true score" is above the 10th percentile. (One chance in 20 is the commonly accepted measurement standard for a 95% confidence level.)

Because measurement error varies from class to class, instructors with identical means on the SIR II items may not have the same items flagged. In particular, measurement error tends to be larger when the number of respondents is low and when disagreement among the respondents is high. For example, instructors in small classes are likely to have fewer items flagged than those in large classes because there is less confidence of the reliability of means in small classes.

Copyright © 1995 by Educational Testing Service. All rights reserved.

57208-12723-SIRP1-4ELX 7/24/95