eSIR / SIR II Review

Using Fall 2013 Data

Author: Joseph F. van Gaalen, Ph.D., Coordinator, Academic Assessment

1 INTRODUCTION

Florida SouthWestern State College's (FSW) course-level assessment measures include the review of student evaluation of instruction surveys, both in traditional courses through Student Instruction Report II (SIR II) provided by ETS, and in online courses using the eSIR. In both surveys, a 5-point scale from lowest (1) to highest (5) is used. Additional analyses include a college-wide comparison of surveys from both traditional and online courses and college-wide survey response rates.

The 2014/15 Course-level Assessment Plan includes courses that are offered in all three modalities (campus-based, online, and dual enrollment). Based on these criteria, the following courses are included in the study: AMH2010, AMH2020, PSY2012, BSC1010, BSC1010L, FRE1120, FRE1121, SPC2608, SPN1120, SPN1121, HUM2211, HUM2235, HUM2510, GEB1011, ENC1101, ENC1102, MAC1105, MAC1114, MAC1140, EDF2005, and EDF2085. In some cases, 2014/15 Course-level Assessment Plan courses either did not run any sections in Fall 2013, such as FRE1121, or did not yet exist, such as the compressed developmental courses ENC0022, MAT0057, and REA0019.

Beginning Fall 2015, FSW will be using a new instruction evaluation tool, the Student Evaluation of Instruction (SEI), developed within the college itself. This new survey will be ubiquitous across modalities and therefore provide more readily inferable results. As such, this report will serve as a baseline for FSW student evaluation survey results with the new SEI tool.

For additional detail or further analysis not provided in this report, please contact Dr. Joseph van Gaalen, Coordinator of Academic Assessment, Academic Affairs (Joseph.VanGaalen@fsw.edu; x6965).

2 STUDENT SURVEY RESULTS

2.1 RESPONSE RATES

2.1.1 SIR II

During the Fall 2013 semester, 1547 sections across all three campuses and one learning center administered the SIR II instruction evaluation. Of those, 25,334 respondents were collected from an enrollment of 35,566, a response rate of 71.2%.

2.1.2 eSIR

During the Fall 2013 semester, 124 sections conducted through FSW Online administered the eSIR instruction evaluation. Of those, 890 respondents were collected from an enrollment of 2771, a response rate of 32.1%.

2.2 RESULTS AND COMPARISONS FOR COURSE-LEVEL EVALUATION COURSES

2.2.1 SIR II

The overall evaluation mean score (Question 40) across all sections was 4.12. The highest mean score of student evaluation of instruction surveys for overall evaluation from traditional courses (SIR II evaluations) is Introduction to Business (GEB1011), at 4.35 (Table 1). Introductory Spanish II (SPN1121) scored the most 5s in the surveys (70%).

	5 Very effective	4 Effective	3 Moderately effective	2 Somewhat ineffective	1 Ineffective	Mean score
Social Sciences						
AMH2010	44%	35%	18%	3%	1%	4.17
AMH2020	18%	45%	36%	0%	0%	3.75
PSY2012	48%	37%	15%	1%	0%	4.27
Science						
BSC1010	39%	38%	19%	3%	1%	4.12
BSC1010L	38%	29%	18%	8%	7%	3.83
Speech and Foreign Languages						
FRE1120	43%	37%	17%	3%	1%	4.17
FRE1121	n/a	n/a	n/a	n/a	n/a	n/a
SPC2608[#]	39%	33%	27%	1%	0%	4.10
SPN1120	31%	37%	26%	6%	1%	3.91
SPN1121	70%	25%	5%	0%	0%	4.63
Humanities						
HUM2211	49%	22%	24%	4%	1%	4.11
HUM2235	42%	30%	18%	2%	8%	3.96
HUM2510	53%	22%	25%	0%	0%	4.27
Business and Technology						
GEB1011	52%	31%	17%	0%	0%	4.40
English						
ENC1101	42%	31%	19%	6%	2%	4.07
ENC1102	32%	41%	19%	5%	3%	4.00
Mathematics						
MAC1105	41%	33%	19%	5%	2%	4.07
MAC1114	n/a	n/a	n/a	n/a	n/a	n/a
MAC1140	50%	32%	10%	8%	0%	4.24
Education						
EDF2005	36%	45%	17%	2%	0%	4.17
EDF2085	33%	45%	17%	5%	0%	4.06

Table 1. SIR II overall course evaluations for Fall 2013 for courses listed in course-level assessment for 2014-15. [#]SPC2608 was classified as SPC2023 during the Fall 2013 semester.

2.2.2 eSIR

The overall evaluation mean score (Question 41) across all sections was 4.08. The highest mean score of student evaluation of instruction surveys for overall evaluation from traditional courses (eSIR evaluations) is Ancient World – Medieval (HUM2211), at 4.43 (Table 2). The same course also scored the most 5s in the surveys (54%).

	5 Very effective	4 Effective	3 Moderately effective	2 Somewhat ineffective	1 Ineffective	Mean score
Social Sciences						
AMH2010	25%	50%	7%	11%	7%	3.69
AMH2020	33%	46%	18%	3%	0%	4.10
PSY2012	33%	28%	27%	3%	5%	3.84
Science						
BSC1010	38%	43%	15%	3%	0%	4.17
BSC1010L	41%	29%	12%	0%	0%	4.35
Speech and Foreign Languages						
FRE1120	38%	30%	21%	5%	4%	3.91
FRE1121	30%	40%	20%	10%	0%	3.90
SPC2608*[#]	20%	40%	0%	0%	40%	3.00
SPN1120	32%	22%	45%	0%	0%	3.87
SPN1121	27%	31%	31%	10%	0%	3.76
Humanities						
HUM2211	54%	32%	7%	0%	2%	4.43
HUM2235	52%	20%	15%	10%	3%	4.07
HUM2510	44%	37%	16%	2%	0%	4.25
Business and Technology						
GEB1011	41%	31%	21%	0%	5%	4.22
English						
ENC1101	43%	33%	13%	2%	7%	4.11
ENC1102	28%	25%	27%	9%	11%	3.59
Mathematics						
MAC1105	35%	39%	23%	0%	2%	4.06
MAC1114	n/a	n/a	n/a	n/a	n/a	n/a
MAC1140	n/a	n/a	n/a	n/a	n/a	n/a
Education						
EDF2005	51%	27%	7%	5%	9%	4.07
EDF2085*	0%	0%	16%	33%	50%	1.67

*Table 2. eSIR overall course evaluations for Fall 2013 for courses listed in course-level assessment for 2014-15. *EDF2085 & SPC2608 had only 5 survey responses. [#]SPC2608 was classified as SPC2023 during the Fall 2013 semester.*

2.2.3 Course-by-Course eSIR / SIR II Comparison (Overall Course Evaluation Question)

The difference in means of the overall evaluation survey question was analyzed for all courses offered in both traditional and online modalities. The differences were tested for significance using a Welch's t-test according to standard methods (Davis, 1973; McDonald, 2009; Wilkinson, 1999). Since the data are interval-level measurements (i.e. Likert-type ratings), a review of the median or mode is more satisfactory for interpreting the most common feeling in survey response (Jamieson, 2004). However, since a review of the means yields information relating to the standard deviation, and indirectly the skewness and kurtosis of the data (Siegel, 1956), the author feels this is a relevant method. Moreover, the results are not intended to be interpreted using the Likert-type rating definitions (e.g. very effective, effective, etc.), but instead are designed to evaluate shifts in the collective survey responses.

The Welch's t-test results of the difference in means of the overall evaluation survey question indicate that for PSY2012, SPN1121, HUM2211, ENC1102, and EDF2085, we must reject the null hypothesis that the difference in the means of the overall evaluation survey question is equal to 0; and we can conclude with a 95% confidence that the differences in scores are not solely due to chance. For the remaining

rubric criteria we cannot reject the null hypothesis, meaning the differences in mean scores for those artifacts can be a result of chance.

Course	mean diff	p-value	Course	mean diff	p-value
AMH2010	0.30	0.056	HUM2235	-0.11	0.679
AMH2020	-0.35	0.215	HUM2510	0.02	0.869
PSY2012	**0.42**	**<0.001**	GEB1011	0.18	0.410
BSC1010	-0.05	0.863	ENC1101	-0.04	0.702
BSC1010L	-0.52	0.069	**ENC1102**	**0.41**	**0.024**
FRE1120	0.26	0.367	MAC1105	-0.01	0.950
FRE1121	n/a	n/a	MAC1114	n/a	n/a
SPC2608	1.10	0.224	MAC1140	n/a	n/a
SPN1120	0.04	0.865	EDF2005	0.09	0.749
SPN1121	**0.87**	**0.010**	**EDF2085**	**2.40**	**0.002****
HUM2211	**-0.32**	**0.048***			

*Table 3. Difference in means of the overall evaluation survey question for all courses offered in both traditional and online modalities between SIR II and eSIR. Positive mean scores indicate SIR II > eSIR. Bold-faced denotes significant results. *Denote marginal significance as defined by Johnson (2013). **Small sample size may result in Type I errors (false positives) approximately 30% of the time for all statistically significant results (de Winter, 2013).*

Based on the work of Johnson (2013), there is a 17-25% chance that the marginally significant results depicted in Table 3 may be false positives (i.e. Type I errors). These marginal results, defined as those within the 95-99% confidence level, include HUM2211 and EDF2085. Further, the sample size for EDF2085 online survey was small (5 respondents). Reduced sample sizes of this magnitude have been shown to result in Type I errors (false positives) approximately 30% of the time for all statistically significant results (de Winter, 2013). Type II errors (false negatives) can also be of concern here. Therefore, we must bear this in mind when rejecting the null hypothesis that the difference in the means of EDF2085, and, to a lesser extent, HUM2211.

3 College-Wide SIR II / eSIR Comparison (All Common Questions)

To date there has been no direct comparison of survey results at FSW between the traditional course SIR II evaluation and the online course eSIR evaluation. Questions that can be compared reasonably and effectively were identified and overall evaluation scores college-wide were compared between online and traditional. Common questions and their identifying question number are labeled with an ID letter and listed in Table 4. Twelve questions used identical language between the SIR II and eSIR. A further 12 questions were identified as being similar enough as to have results evaluated without any hindrance towards interpretation of the results. Question IDs A-X comparisons are provided in Figures 1-3.

All survey results of Question IDs Q through W show statistically significant differences in the means (Table 4). We must reject the null hypothesis that the difference in the means of the two surveys (eSIR and SIR II) are equal to 0, and we can conclude with a 95% confidence that the differences in survey results are not solely due to chance. In the most extreme case, Question ID W, results from a Welch's T-test exhibit significant differences in mean survey rating {$t(399)=-0.44$, $p=4.55\times10^{-53}$}. No effect size was calculated for this analysis as survey results are typical of the field and therefore already meaningful (Wilkinson, 1999). Based on the work of Johnson (2013), there is a 17-25% chance the marginally significant results depicted in Table 4 may be false positives (i.e. Type I errors). These marginal results, defined as those within the 95-99% confidence level, include Question IDs H and J.

SIR II			eSIR	
Q #	*Question*	*ID*	*Q #*	*Question*
1	*The instructor's explanation of course requirements.*	*A*	*1*	*The instructor's explanation of course requirements.*
2	*The instructor's preparation for each class period.*	*B*	*2*	*The instructor's preparedness for this course.*
5	*The instructor's way of summarizing or emphasizing important points in class.*	*C*	*5*	*The instructor's summaries of important material.*
8	*The instructor's use of examples or illustrations to clarify course material.*	*D*	*8*	*The instructor's use of examples or illustrations to clarify course material.*
9	*The instructor's use of challenging questions or problems.*	*E*	*9*	*The instructor's use of challenging questions or problems.*
12	*The instructor's respect for students.*	*F*	*14*	*The respect for students shown by the instructor.*
13	*The instructor's concern for student progress.*	*G*	*15*	*The concern for student progress shown by the instructor.*
16	*The information given to students about how they would be graded.*	*H*	*17*	*The information given to students about grading.*
18	*The exams' coverage of important aspects of the course.*	*I*	*20*	*The exam coverage of important aspects of the course.*
19	*The instructor's comments on assignments and exams.*	*J*	*21*	*The instructor's comments on assignments and exams.*
21	*The helpfulness of assignments in understanding course material.*	*K*	*19*	*The effectiveness of assignments in contributing to learning.*
22	*Problems or questions presented by the instructor for small group discussions.*	*L*	*23*	*Problems or questions assigned by the instructor for group discussions.*
24	*Laboratory exercises for understanding important course concepts.*	*M*	*24*	*Laboratory exercises.*
25	*Assigned projects in which students worked together.*	*N*	*25*	*Projects in which students work together in teams.*
26	*Case studies, simulations or role playing.*	*O*	*26*	*Case studies, simulations or role playing.*
27	*Course journals or logs required of students.*	*P*	*27*	*Journals or logs required of students.*
29	*My learning increased in this course.*	*Q*	*30*	*My learning in this course was…*
30	*I made progress toward achieving course objectives.*	*R*	*31*	*My progress toward achieving the course objectives was…*
31	*My interest in the subject area has increased.*	*S*	*32*	*My interest in the course's subject area increased…*
32	*This course helped me to think independently about the subject matter.*	*T*	*33*	*This course helped me think independently about the subject matter…*
33	*This course actively involved me in what I was learning.*	*U*	*34*	*This course actively involved me in what I was learning…*
34	*I studied and put effort into this course.*	*V*	*35*	*The effort I put into this course was…*
36	*I was challenged by this course.*	*W*	*37*	*I was academically challenged by this course…*
40	*Which one of the following best describes this course for you?*	*X*	*41*	*Rate the overall effectiveness of instruction in this course as it contributed to your learning (set aside your feelings about the course content):*

Table 4. Common survey questions among SIR II and eSIR course evaluation forms.

Figure 1. Survey response distributions for SIR II / eSIR comparable questions. Question IDs: A through H. Red line represents SIR II results. Black line represents eSIR results. Responses 5-4-3-2-1 represent 'very effective', 'effective', 'moderately effective', 'somewhat ineffective', and 'ineffective'.

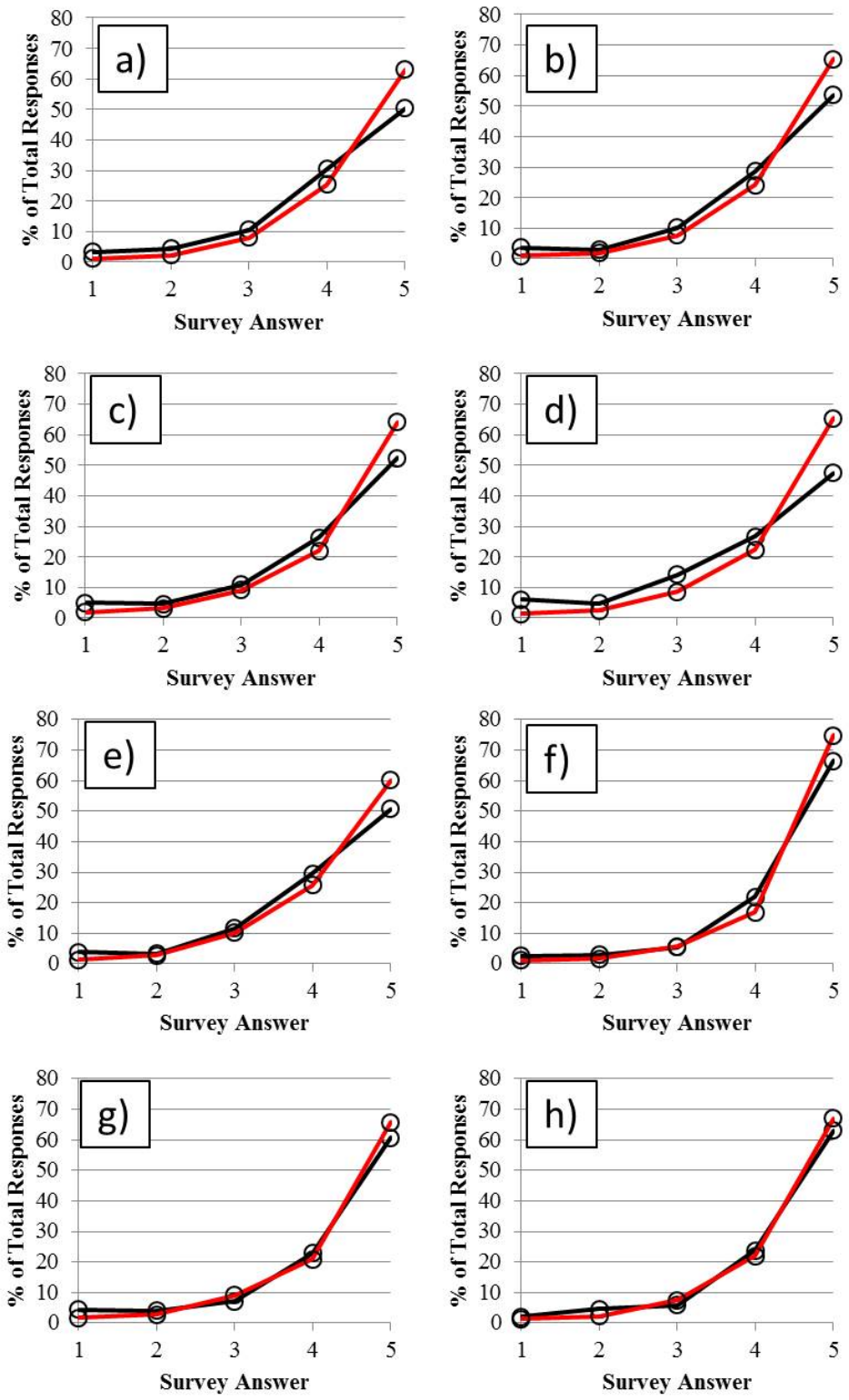


Figure 2. Survey response distributions for SIR II / eSIR comparable questions. Question IDs: I through P. Red line represents SIR II represents SIR II results. Black line represents eSIR results. Responses 5-4-3-2-1 represent 'very effective', 'effective', 'moderately effective', 'somewhat ineffective', and 'ineffective'.

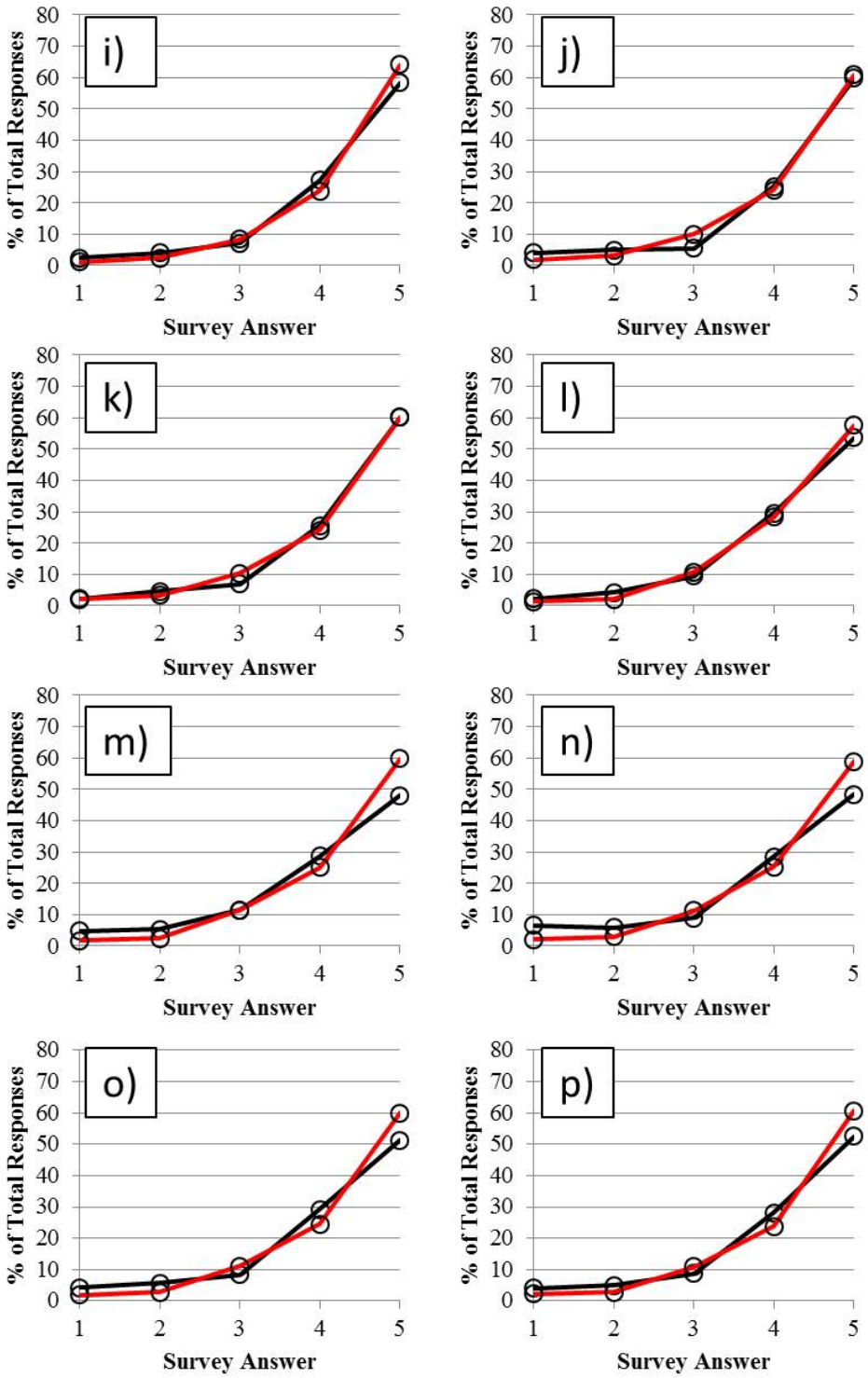
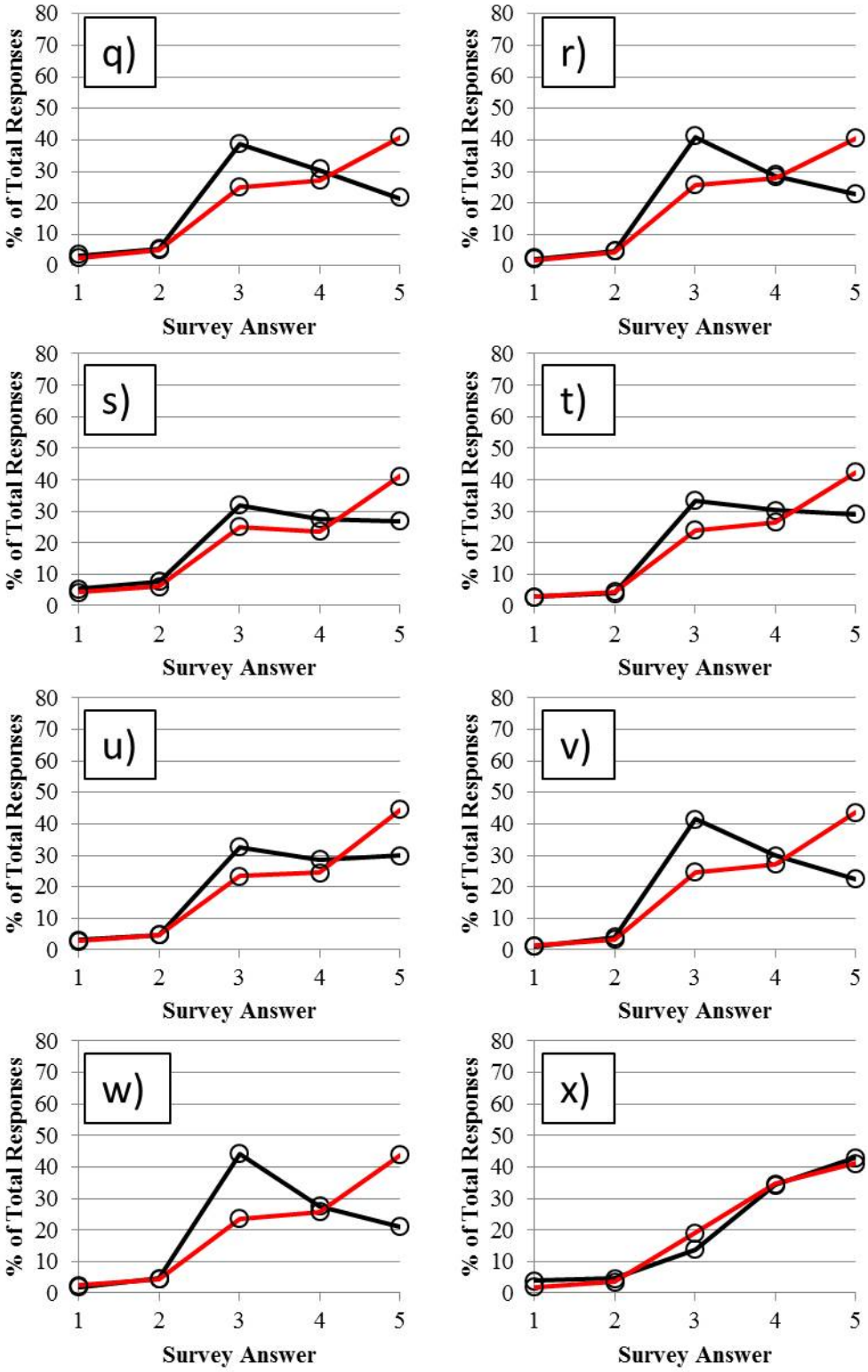


Figure 3. Survey response distributions for SIR II / eSIR comparable questions. Question IDs Q through X. Red line represents SIR II represents SIR II results. Black line represents eSIR results. Responses 5-4-3-2-1 represent 'much more than most courses', 'more than most courses', 'about the same as others', 'less than most courses', and 'much less than most courses'.



Comparison results depict a somewhat more successful demonstration of examples and illustrations used in traditional courses when compared with online courses (Question ID: D). In traditional courses, 65.4% of students surveyed responded 'very effective' compared with only 47.4% of online students surveyed.

ID	mean diff	p-value
A	**-0.26**	**<0.001**
B	**-0.24**	**<0.001**
C	**-0.26**	**<0.001**
D	**-0.44**	**<0.001**
E	**-0.21**	**<0.001**
F	**-0.14**	**0.001**
G	**-0.14**	**0.001**
H	**-0.08**	**0.047***
I	**-0.14**	**0.001**
J	**-0.08**	**0.048***
K	-0.01	1.00
Q	**-0.37**	**<0.001**
R	**-0.35**	**<0.001**
S	**-0.28**	**<0.001**
T	**-0.27**	**<0.001**
U	**-0.27**	**<0.001**
V	**-0.43**	**<0.001**
W	**-0.43**	**<0.001**
X	-0.04	0.392

*Table 5. Significance test of the difference in means of Survey results between SIR II and eSIR. Positive mean scores indicate SIR II > eSIR. No mean data provided by ETS for IDs L through P. *Denote marginal significance as defined by Johnson (2013). Bold-faced denotes significant results. *Denote marginal significance as defined by Johnson (2013).*

More substantial disagreements exist between questions concerning course outcomes and student effort and involvement (Question IDs Q – W). Among traditional courses, in all cases the mode (central tendency) response of students surveyed is 5, or 'much more than most courses'. In the case of eSIR survey results, modality of survey response is 3, or 'about the same as others'.

In the extreme case mentioned above, Question ID W (SIR II: 'I was challenged by this course'; eSIR: 'I was academically challenged by this course'), 23.6% of students surveyed responded with '3' in traditional courses, while online students surveyed responded with a '3' 44.2% of the time.

The following topics display noticeable modal differences between traditional and online survey response:

Learning outcomes

- *Q: Increased learning in the course*
- *R: Students made progress toward achieving course objectives*
- *S: Student interest in subject area increased*
- *T: Course helped student to think independently on the subject matter*
- *U: Course actively involved student in learning*

Student effort and involvement

- *V: Student studied and put effort into the course*
- *W: Student was challenged by the course*

The differences between online and traditional students surveyed regarding learning outcomes, while noteworthy, are fairly common among undergraduates' comparisons of the two learning mediums (Mullen and Tallent-Runnels, 2006). More specifically, students perceive support of faculty as lower in online learning. This may result in significant differences in responses to survey prompts like "Was learning increased?" or "Course actively involved student in learning" because both can be associated with the affective support of faculty in the course (Swan, 2001; Mullen and Tallent-Runnels, 2006).

4 CONCLUSIONS

FSW Course-level Assessment Plan includes courses that are offered in all three modalities (campus-based, online, and dual enrollment). The course list includes: AMH2010, AMH2020, PSY2012, BSC1010, BSC1010L, FRE1120, FRE1121, SPC2608, SPN1120, SPN1121, HUM2211, HUM2235, HUM2510, GEB1011, ENC1101, ENC1102, MAC1105, MAC1114, MAC1140, EDF2005, and EDF2085). Since Fall 2015 will initiate a new instruction evaluation tool, this report will serve as a baseline for FSW student evaluation survey results with the new SEI tool.

A Welch's t-test determined that of the courses listed above, for PSY2012, SPN1121, HUM2211, ENC1102, and EDF2085, we must reject the null hypothesis that the difference in the means of the overall evaluation survey question is equal to 0; and we can conclude with a 95% confidence that the differences in scores are not solely due to chance. However, HUM2211 and EDF2085 may be a result of a Type I error (false positive). For the remaining rubric criteria we cannot reject the null hypothesis, meaning the differences in mean scores for those artifacts can be a result of chance.

During Fall 2013, 71.2% of 35,566 students enrolled participated in SIR II surveys conducted in traditional classes. That same semester, 32.1% of 2771 students enrolled participated in eSIR surveys conducted in online classes. Mean survey scores for the courses reviewed in this study range from 4.35 to 3.70 in SIR II evaluations, and 4.08 to 3.00 in eSIR evaluations.

Questions that can be compared reasonably and effectively between SIR II and eSIR surveys were identified and college-wide scores for those questions were compared. Of 24 common questions, only 19 were comparable as means are not reported by ETS for five of the identified questions. Seventeen of the 19 questions exhibited statistically significant differences between eSIR and SIR II survey means. In all cases, significant or non-significant, evaluations were lower for eSIR surveys than for SIR II surveys.

Substantive disagreement exists between the surveys regarding questions concerning course outcomes and student effort and involvement. Survey mean scores from the eSIR report exhibit scores ranging from 0.27 to 0.43 points below SIR II survey means (scoring range: 0 = lowest, 5 = highest). All questions in these areas were statistically significant.

Overall course rating was not significantly different between offering modality. The eSIR report survey mean was 0.04 points below the SIR II mean score. Survey comparison indicates considerable disagreement at a topical level but not with regard to overall course evaluation.

5 REFERENCES

Davis, J.C. 1973. Statistics and Data Analysis in Geology. John Wiley & Sons, New York, New York, 564 pp.

de Winter, J.C.F. 2013. Using the Student's T-Test with Extremely Small Sample Sizes. Practical Assessment, Research, and Evaluation, 18(10), 1-12.

Douglas, I., and Alemanne, N.D. 2007. Measuring student participation and effort. Procedings from the IADIS International Conference on Cognition and Exploratory Learning in Digital Age, Algarve, Portugal, Dec. 7-9, 2007.

Dutton, J., Dutton, M., and Perry, J. 2002. How do online students differ from lecture students? Journal of Asynchronous Learning Networks, 6(1), 2002.

Jamieson, S. 2004. Likert scales: how to (ab)use them. Medical Education, 38(12), 1217-1218.

Johnson, V. 2013. Revised Standards for Statistical Evidence. Proceedings of the National Academy of Science, 110(48), 19313-19317.

McDonald, J.H. 2009. Handbook of Biological Statistics (2nd ed.). Sparky House Publishing, Baltimore, Maryland.

Mullen, G.E. and Tallent-Runnels, M.K. 2006. Student outcmoes and perceptions of instructors' demands and support in online and traditional classrooms. The Internet and Higher Education, 9(4), 257-266.

Siegel, S. 1956. Nonparametric statistics for the behavior sciences. McGraw-Hill Book Company, New York, New York, 312 pp.

Swan, K. 2001. Virtual interaction: Design factors affecting student satisfaction and perceived learning in asynchronous online courses. Distance Education, 22(2), 306-331.

Tallent-Runnels, M.K., Thomas, J.A., Lan, W.Y., Cooper, S., Ahern, T.C., and Shaw, S.M. 2006. Teaching courses online: A review of the research. Review of Educational Research, 76(1), 93-135.

Wilkinson, L. 1999. APA Task Force on Statistical Inference. Statistical Methods in Psychology Journals: Guidelines and Explanations. American Psychologist 54 (8), 594–604.

Yang, Y., YoonJung, C., Mathew, S., and Worth, S. 2011. College student effort expenditure in online versus face-to-face courses: The role of gender, team learning orientation, and sense of classroom community. Journal of Advanced Academics, 22(4), 619-638.