Academic Challenge and Engagement Index (CEI): Development and Validation

Debbie E. MᶜGhee, Gerald M. Gillmore, Anthony G. Greenwald
November 2007

INTRODUCTION

In response to expressed concerns regarding the degree of academic challenge posed by UW courses, OEA undertook to develop a single index of challenge and student engagement based on items from *Instructional Assessment System (IAS)* course evaluation forms.[1] Although a set of items specifically directed at this topic had been added to *IAS* forms in 1998, we felt that a single index might provide a simpler and more powerful representation for individual courses. Also, because the *IAS* is used to evaluate a large percentage of courses taught at the UW, the index could provide useful insight to more general student perception of UW educational experiences.

METHOD

All *IAS* forms include six items that ask students to rate the academic challenge of the course and the degree to which they feel involved in or engaged by the course (Table 1). To determine which combination of items might yield an optimum index, we drew raw and median ratings of classes taught autumn 2000 through spring 2004 from the database of UW-Seattle *IAS* responses. Only classes with 5 or more ratings were included, resulting in a final dataset containing student ratings of 37,177 classes.

Table 1. *IAS* Challenge and Engagement Items

Item Abbreviation	Item Text	Response Scale
Evaluative Items		0=very poor — 5=excellent
Course as a whole	The course as a whole was:	
Course content	The course content was:	
Instructor's contribution	The instructor's contribution to the course was:	
Instructor's effectiveness	The instructor's effectiveness in teaching the subject matter was:	
Challenge & Engagement Items		
	Relative to other college courses you have taken,	1=much lower — 7=much higher
Challenge	the intellectual challenge presented was:	
Effort expended	the amount of effort you put into this course was:	
Effort to succeed	the amount of effort to succeed in this course was:	
Involvement	your involvement in this course was:	
Total hours	On average now many hours per week have you spent on this course?	1=under 2 — 12=22 or more
Valuable hours	From the total average hours above, how many do you consider were valuable in advancing your education?	

---

[1] We would like to thank the Faculty Council on Instructional Quality (FCIQ) for their review and suggestions.

After examining the interrelationships among the item ratings, we computed several preliminary Challenge and Engagement Indices (CEIs). We then examined the utility of the indices in relation to external criteria provided by: 1) IAS ratings of teaching award recipients and 2) responses to a survey of UW seniors. Finally, looked at how CEI scores varied as a function of course characteristics (e.g., course level).

RESULTS

Item Inter-relationships

Table 2 displays the correlations among the item ratings at the class level. The rater- and class-level coefficients were highly similar. Because our goal was to associate a CEI with each class, all subsequent analyses involved class-level item medians.

Median ratings of items 24 through 27 (relative challenge, effort expended, effort required to succeed, and involvement) were all strongly intercorrelated ($\bar{r}$ = .77). Ratings of items 24-27 were moderately correlated with Item 28 (total hours spent, $\bar{r}$ = .54) and Item 29 (valuable hours, $\bar{r}$ = .56), but slightly less correlated with the "per credit" versions of those items ($\bar{r}s$ = .31 and .35). As expected, ratings of items 28 and 29 were very strongly correlated (r = .92). Importantly, Items 24-27 were only modestly correlated with the aggregate of Items 1-4 (overall evaluation of the class, $\bar{r}$ = .29).

Table 2. Intercorrelations among ratings on *IAS* Challenge-related items

	Challeng	Expend	Succeed	Involve	Totalhrs	Valubhrs	Tothr/cr	Valhr/cr
24. Challenge	—							
25. Effort expended	.77	—						
26. Effort to succeed	.78	.87	—					
27. Involvement	.66	.85	.78	—				
28. Total hours	.45	.59	.58	.52	—			
29. Valuable hours	.51	.60	.56	.52	.91	—		
Total hours/credit	.27	.35	.31	.28	.49	.42	—	
Valuable hours/credit	.35	.38	.32	.31	.44	.53	.92	—
Overall Evaluation	.39	.29	.19	.28	.05	.25	.03	.21

Note. n = 37,175.

A series of principal components analyses (PCA) further illuminated the interrelationships among the item medians. First, items 24-29 resolved into a single component that accounted for 71% of the variance in the six variables. Second, when items 28 and 29 were replaced by their "per credit" versions, two components emerged. The first accounted for 54% of the variance and was comprised of Items 24-27, and the second component (akin to a price-to-earnings ratio) was comprised of the remaining two items and accounted for 33% of the variance. The third PCA included items 24-29 and 1-4. This resulted in two components: challenge and engagement (items 24-29: 42% of the variance) and overall evaluation (items 1-4: 38% of the variance). Thus, the challenge and engagement items measured something different than the overall evaluation items.

Candidate indices

Based on these results, the next step was to create several indices using various combinations of items and to examine the internal consistency of those indices. We chose to include valuable hours spent (item 29) rather than total hours spent (item 28) because we felt the latter included hours that were

perceived as not useful (so-called "busy work"). The four indices were the: a) mean of items 24-26, b) mean of items 24-27, c) mean of items 24-26 and 29, and d) mean of items 24-27 and 29. For the latter two, item medians were first converted to z-scores. As expected from the simple correlations, the internal consistency of each of the indices was very high (α = .88-.94), and the indices were all highly inter-correlated ($r \geq .96$, Table 3). Importantly, each of the indices correlated only modestly with the overall evaluation index ($\bar{r}$ = .31).

Table 3. Intercorrelations among candidate CEIs and overall evaluation.

	CEI1	CEI2	CEI3	CEI4
CEI1 (Mean of 24-26)	—			
CEI2 (Mean of 24-27)	.99	—		
CEI3 (Mean of 24-26, 29 — standardized)	.97	.96	—	
CEI4 (Mean of 24-27, 29 — standardized)	.97	.98	.99	—
Overall Evaluation	.31	.32	.32	.32

2004 Senior Survey Data[2]

Every other spring quarter OEA surveys all UW seniors regarding their university experiences. Two questions in the 2004 survey asked students to identify up to three courses they had found to be the most and least intellectually challenging during their tenure at the UW.

A total of 2132 students responded to the survey, and they nominated 2106 distinct courses as either the most or least challenging or both. A number of courses were incompletely specified (e.g., no course number given) and were not included in the above counts. Of those 2106 courses, 1479 had IAS ratings data on file. The number of courses identified as most challenging was much higher than the number identified as least challenging (1176 and 711, respectively). We computed a nomination score for each course using the following formula:

$$\textit{Senior Nomination Score} = \frac{n_{\text{most}} - n_{\text{least}}}{\text{TotalEnrollment} \times .23}$$

where n_{most} and n_{least} were the number of nominations for most and least challenging course, total enrollment was the total number of students enrolled in all sections of a course during AU 2000 – SP 2004, and .23 was the Senior Survey response rate. The 54 courses ranked as most challenging and the 42 courses ranked as least challenging are listed in the Appendix.

We should note that these data contain a considerable amount of noise with regard to the research question. Specifically, much of the challenge of a course is dependent on the professor who teaches it, but for the analyses reported below, the unit of analysis was the course. Thus, a senior survey respondent may have had in mind Professor X's version of a course, but — assuming more than one faculty member had taught the course during this time period — nomination scores and course evaluation ratings would have been aggregated over Professor X's, Professor Y's, and Professor Z's versions.

---

[2] The Senior Survey is described at http://www.washington.edu/oea/pdfs/reports/OEAReport0501.pdf.

Table 4. Comparison of *IAS* ratings of courses designated by seniors as low- (*n*=525) and high-challenge (*n*=947).

	Low		High		Regression Results		
	Mean	SD	Mean	SD	$F(1,1470)$	R	R^2
Overall Evaluation							
Average of 1 – 4 (raw)	4.0	.47	4.1	.46	43.3**	.17	.03
Challenge and Engagement Items							
Item24. Challenge	5.0	.47	5.5	.50	382.9**	.46	.21
Item25. Effort expended	4.9	.50	5.3	.53	223.0**	.36	.13
Item26. Effort to Succeed	5.0	.47	5.5	.49	263.8**	.39	.15
Item27. Involvement	5.1	.47	5.4	.50	125.8**	.28	.08
Item29. Valuable hours	6.3	2.13	7.7	2.29	135.6**	.29	.08
Grade Items							
Item23. Relative grade	5.0	.47	4.7	.50	40.0**	.61	.03
Item30. Absolute grade	3.4	.31	3.4	.29	1.6	.03	.001
Indices							
CEI1 (Mean of 24-26)	5.0	.45	5.4	.47	328.7**	.43	.18
CEI2 (Mean of 24-27)	5.0	.45	5.5	.46	284.7**	.40	.16
CEI3 (Mean of 24-26, 29 standardized)	-.2	.62	.4	.63	315.5**	.42	.18
CEI4 (Mean of 24-27, 29 -- standardized)	-.2	.61	.3	.64	228.1**	.40	.16

Notes.　1. Low- and High Challenge groups were formed from a median split of *Senior Nomination Score* and are for display purposes only. See text for computation of *Senior Nomination Score*.

2. *F*- and *R*- values are from regression analyses predicting nomination score from the *IAS* rating.

* $p < .005$, ** $p < .0001$.

Regression analyses showed a slight but statistically significant relationship between the evaluative rating and nomination score (Table 4). In contrast, all of the relationships between nomination score and the challenge and engagement items were large in magnitude, with the most challenging courses showing higher means on all items. Especially large were the effects for *intellectual challenge presented*, *effort expended*, and *amount of effort to succeed* — each of which accounted for at least 13% of the variance in nomination score. The relationships between nomination score and each of the two grade items were also large and significant, but for those items it was the *least* challenging courses that obtained higher means. That is, courses with the highest grades also tended to be rated as least challenging.

Finally, there were large and significant relationships between nomination score and each of the candidate CEIs. The relative magnitudes of the coefficients were very similar, which is not surprising since the indices were highly correlated.

Distinguished Teaching Awards Data

Course evaluation data were available for 77 instructors who had received the Distinguished or Excellence in Teaching Award between the years 1972-2004. The mean class-level ratings of award recipients on each of four candidate CEIs were compared to those of non-recipients; the dataset was limited to evaluations collected between 2001-2003.

Table 5 shows that award recipients had higher means than non-recipients on the evaluative index, items 24-27, and item 29. The greatest single-item difference between the groups was on item 24 (*Ms* = 5.58 vs. 5.25). The weakest challenge-related item was item 28, total hours; valuable hours was a much stronger differentiator of the two groups. Recipients and non-recipients also differed on average expected grade.

Table 5. Comparison of *IAS* Ratings of Classes Taught by Distinguished Teaching Award Recipients (*n*=77) and Non-Recipients (*n*=3865)

Item or Index	Non-Recipient Mean	SD	Recipient Mean	SD	$F(1,3940)$	d
Evaluative Indices						
Average of 1 – 4 (raw)	3.9	.55	4.5	.34	78.8**	1.6
Challenge and Engagement Items						
Item24. Challenge	5.2	.54	5.6	.49	33.6**	.74
Item25. Effort expended	5.1	.55	5.4	.51	16.0**	.50
Item26. Effort to Succeed	5.2	.54	5.4	.48	11.0*	.42
Item27. Involvement	5.3	.50	5.5	.47	12.0*	.43
Item29. Valuable hours	6.8	2.08	7.8	1.93	16.6**	.50
Grade Items						
Item23. Relative grade	4.8	0.51	4.8	.46	.65	.10
Item30. Absolute grade	3.4	0.20	3.5	.16	8.8*	.41
Indices						
CEI1 (Mean of 24-26)	5.2	.51	5.5	.47	21.7**	.58
CEI2 (Mean of 24-27)	5.2	.49	5.5	.46	20.2**	.56
CEI3 (Mean of 24-26, 29 — standardized)	.02	.77	.45	.69	23.5**	.62
CEI4 (Mean of 24-27, 29 — standardized)	.02	.76	.43	.69	22.0**	.59

Note. * $p < .005$, ** $p < .0001$.

Course Characteristic Effects

Analyses of variance were conducted on each of the four potential indices to ascertain whether mean challenge rating varied with certain course characteristics. The fixed factors were course level (100, 200, 300, 400, and 500-800), course credits (1-2 vs. 3-5 vs. 6-16), and instructor rank (tenure- and non-tenure track faculty vs. TA's and "Others"). Figure 1 illustrates the pattern of findings for all four indices.

Mean ratings varied significantly by course level. For each of the four indices, the mean rating of 300-level course was significantly higher than those of courses of all other levels. The lowest mean ratings went to 200- and graduate-level courses. The relationship between course credits and challenge rating was fairly linear: 1-2 credit courses had the lowest means, and 6+ credits courses had the highest means. The one-way effect of rank was not significant. The highest mean rating overall went to 300-level courses worth six or more credits.
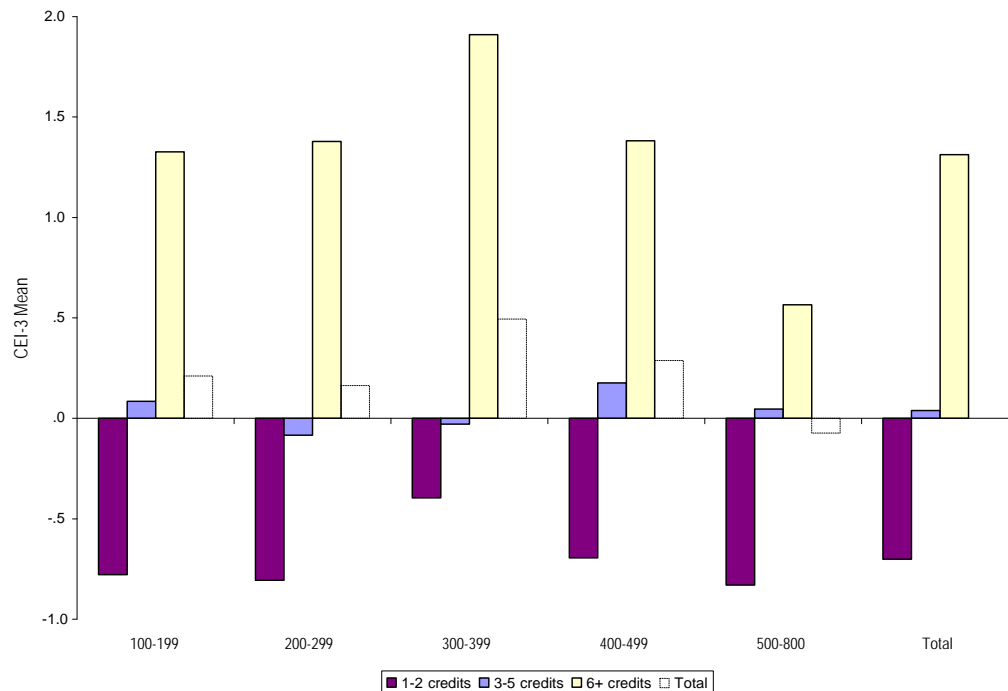
Figure 1. Mean challenge ratings by course level and course credits (CEI3).

Index Selection

The goal of this research was to construct an index from existing course evaluation items that encapsulated students' perceptions of the challenge and engagement level of a course. Statistically, none of the candidate indices proved exceptionally better than the others. All performed with approximately equal reliability and discrimination. Therefore, we chose the set of items that seemed to offer the best content validity and which seemed most straightforwardly to address the concept of challenge and engagement. Thus, we included *valuable hours* (item 29) because it provided information not provided by the other items, and we excluded *involvement* (item 27) because the two indices that included this item performed no better than the two that did not. Therefore, we selected CEI3 as the index that will be included in course evaluation reports.

DISCUSSION AND CONCLUSIONS

The development of an index relating to challenge and engagement from the existing system by which students rate courses would appear to be both an efficient and effective way to inform faculty about students' perception of the challenge and engagement of their courses. Of first importance is to know whether such an index would accurately reflect what we intended. In other words, would the index be valid? Determining scale validity is not a cut-and-dried matter, and it is seldom one that can be accomplished with a single study. However, the results of the present series of analyses provide strong evidence for validity of a Challenge and Engagement Index (CEI) that we constructed from existing IAS items.

First, we have shown that the challenge index measures something different than students' general liking for the course. The correlation between the CEI and the overall evaluation score was .32. A correlation of that magnitude is commonly regarded as small, and it indicates that the two constructs share only nine

percent of variance. Furthermore, a comparison of the ratings given to courses that were retrospectively nominated by seniors as most and least challenging revealed a significant difference for the CEI but no difference in the overall evaluation rating. Thus, although we cannot say that the CEI and overall evaluation are independent, we can assert that the CEI is indicative of a course attribute different from students' rating of its overall quality.

Second, we have shown that the CEI works together with the overall evaluation rating to differentiate between recipients and non-recipients of distinguished teaching awards. That is, in contrast to the senior survey findings — where respondents were asked to focus on challenging *courses* — examination of the mean ratings of distinguished *teachers* showed that their courses were perceived as more challenging, higher in overall quality, and required more hours of valuable work.

Third, we have shown that the CEI has a complex relationship with grades. The senior survey data show that students associate *low challenge* with *higher relative grade*; however, the two groups of courses they nominated did not differ on absolute grade. In contrast, distinguished teachers' classes — relative to the classes of others — were rated as more challenging, average in terms of relative grade, yet more generous in terms of absolute grade. Other analyses revealed no relationship between CEI score and *relative* grade but an inverse relationship between CEI and *absolute* grade. Taken together, these results suggest that seniors used a "personal best/worst" scheme for recalling and nominating most and least challenging courses. However, when one analyzes ratings data that is free of the noise introduced by collapsing over instructors, one finds that the most challenging classes tend to award grades that are slightly lower than average.

In conclusion, the present research shows that students tend to like challenging, engaging classes somewhat more than less challenging classes, and the most well-regarded teachers are those who challenge their students even while being 'harder' graders. Thus, the CEI is a mechanism by which instructors can monitor and perhaps improve the challenge level of their courses and, indirectly, a mechanism for countering the temptation to give lenient grades in the service of higher evaluation ratings.

APPENDIX
Most and Least Challenging Courses based on Senior Survey Nominations

Most Challenging Courses (80th percentile or greater)

Course	Most	Least	SNS	Course	Most	Least	SNS
ART 205	5	0	.91	MATH 328	6	0	.18
PSYCH 331	12	1	.75	CHEM 242	8	0	.18
MATH 327	14	1	.73	OCEAN 444	7	0	.18
CHID 390	25	0	.64	MICRO 410	16	0	.18
CHEM 436	7	0	.49	ENGL 494	6	0	.17
BIOL 453	5	0	.45	BIOC 441	19	0	.16
COM 306	6	1	.43	MICRO 411	12	2	.16
SIS 495	17	1	.43	ENGL 310	6	1	.15
LSJ 320	4	1	.37	SPHSC 461	11	0	.15
POL S 441	6	0	.37	CSE 421	7	0	.15
PHYS 325	7	0	.37	ACCTG 485	4	1	.15
GEOG 425	5	0	.36	ECON 485	5	0	.14
LSJ 490	5	1	.33	ENGL 370	9	1	.14
LSJ 367	6	0	.33	ENGL 498	15	2	.14
CSE 467	7	0	.29	ENGL 304	8	0	.14
PSYCH 445	6	0	.29	CEE 452	5	0	.14
PHIL 450	6	0	.28	SPHSC 320	5	0	.13
POL S 405	5	0	.26	CSE 451	16	0	.13
MATH 310	11	0	.26	CHEM 455	14	1	.13
MATH 334	5	0	.26	MATH 300	11	2	.13
BIOEN 303	6	0	.26	SISEA 490	6	1	.12
E E 443	10	0	.25	ECON 421	6	0	.12
MICRO 442	5	0	.24	ENGL 367	5	0	.12
NBIO 301	10	0	.23	ECON 483	5	0	.12
CHEM 453	24	0	.23	BIOL 409	4	1	.12
COM 334	7	0	.21	ESS 212	5	0	.12
E E 478	14	0	.19	E E 271	10	0	.12

Least Challenging Courses (15th percentile or lower)

Course	Most	Least	SNS	Course	Most	Least	SNS
ECON 200	29	59	-.03	SOC 287	0	8	-.06
C LIT 272	0	7	-.03	CLAS 205	0	22	-.06
PSYCH 201	0	5	-.03	Q SCI 381	1	13	-.06
MUSIC 160	1	6	-.03	LING 400	0	5	-.06
COM 202	3	12	-.03	SPHSC 100	0	17	-.06
COM 201	2	16	-.03	COM 270	2	10	-.08
PSYCH 101	15	110	-.03	ARCH 151	0	24	-.08
ENGL 212	0	5	-.03	ESC 110	2	85	-.09
FREN 102	1	10	-.03	BIOL 356	0	7	-.09
SOC 110	1	17	-.03	OCEAN 102	0	17	-.09
LING 100	1	15	-.04	PSYCH 303	1	4	-.09
FREN 101	1	11	-.04	ARCH 150	0	36	-.09
PSE 102	0	7	-.04	ANTH 228	0	5	-.12
TC 231	1	24	-.04	ARCHY 212	1	8	-.15
ESS 103	0	5	-.04	ART 124	2	12	-.16
ENVIR 202	1	4	-.04	MUSIC 162	3	41	-.16
DRAMA 101	2	24	-.04	DANCE 101	1	50	-.17
A A 101	0	15	-.04	ANTH 209	1	7	-.18
PSYCH 200	0	7	-.05	ART 123	1	5	-.19
MGMT 300	2	14	-.05	ART 121	2	7	-.32
MUSIC 120	0	15	-.05	ANTH 310	0	5	-.37

Note. SNS is the Senior Nomination Score. See text for computation.

OEA Report 07-05: Academic Challenge and Engagement Index (CEI): Development and Validation