IDEA Research Report #6

JULY 2003

Understanding The IDEA System's Extraneous Variables

Donald P. Hoyt and Eun-Joo Lee
The IDEA Center

Research has confirmed the common belief that instructional outcomes are influenced by "extraneous variables."[1] Although it is always important to estimate the amount of student learning in a given course, when the focus is on the *evaluation of teaching effectiveness*, it is important to separate the contributions of the teacher from the contributions of extraneous factors to student learning. That is the purpose of the IDEA system's **adjusted ratings**.

To adjust scores on The IDEA Center's Diagnostic Form, five extraneous variables are assessed and taken into account. These include measures of course motivation (CM), work habits (WH), student effort (E_N), disciplinary difficulty (D_N), and size of class (N). This report is intended to improve understanding of these variables and their implications both for interpreting IDEA reports and understanding the dynamics of learning.

Following a brief definition of the extraneous variables considered in the IDEA System, the relationships among the variables and their impact on course outcomes is discussed. Conclusions and recommendations are presented at the end of the paper.

Extraneous Variable Definitions

1. CM is the average response to Item 39. *I really wanted to take this course regardless of who taught it.*

2. WH is the average response to Item 43. *As a rule, I put forth more effort than other students on academic work.*

3. N is the number of students enrolled in the class.

4. E_N is based on Item 37. *I worked harder on this course than on most courses I have taken.* Because responses to this "effort" item are affected by decisions the instructor makes, it cannot serve as a *bona fide* extraneous variable. E_N modifies responses to Item 37 by taking into account the effects of the three most relevant instructor requirements – amount of reading (Item 33); amount of work in other assignments (Item 34); and intellectual challenge (Item 8. *Stimulates students to intellectual effort beyond that required by most courses*). The portion of Item 37 that remains after these "teacher influences" have been removed (the residual, called "E_N") is considered a true "extraneous variable" (a variable that is beyond the control of the instructor). Classes with a high E_N average contain a disproportionately large number of students who worked harder than expected on the basis of the instructor's requirements.

5. D_N is intended to measure the inherent difficulty of the course. It is a modification of the average response to Item 35, which asks students to compare the *difficulty of the subject matter* with that for other courses. Responses to this question are influenced by the same three "instructor characteristics" cited in the definition of E_N; therefore, it cannot be considered a true "extraneous variable" until these effects have been statistically removed. D_N is the result of that operation.

Understanding E_N and D_N. CM, WH, and N are all unmodified measures; results involving them can be interpreted in a straightforward manner. E_N and D_N, on the other hand, are complex measures. To interpret them requires a more detailed study of the measures that are involved.

As noted above, instructors influence ratings of how hard students worked (effort, Item 37) and of student perceptions of course difficulty (Item 35). Both of these ratings are affected chiefly by the amount of required reading (Item 33), the amount of other (non-reading) work (Item 34), and the degree to which the instructor stimulates student intellectual effort (Item 8). To understand E_N and D_N, it is helpful to examine the inter-relationships among ratings on these items.

First, as expected, the direct measures of "Difficulty" and "Effort" (Items 35 and 37) were highly correlated (r=.67). In courses perceived to be "difficult," students reported that they worked harder than in courses perceived to be "easy."

Second, the teacher requirements reflected in Items 8, 33, and 34 were only slightly related to each other. There was a very slight tendency for instructors who required much reading to also make above average demands in other work (r=.17). The degree to which the instructor stimulated intellectual effort (Item 8) had only small positive relationships to amount of reading and amount of other work (r's of .24 and .33); apparently, such stimulation is more than a matter of making demanding assignments.

Third, although the three measures of how the instructor managed the course (Items 8, 33, and 34) were positively related to both "Difficulty" and "Effort," the relationships were stronger for "Effort" than for "Difficulty," and the pattern of correlations was quite different. "Effort" was related principally to "Amount of other (non-reading) work" (r=.68), and to the stimulation of intellectual effort (r=.56). Correlations of "Amount of other work" and "Stimulation of intellectual effort" were substantially higher for "Effort" than for "Difficulty," but "Amount of reading" was more closely related to "Difficulty" than to "Effort."

About 64 percent of the variation in how hard students worked was accounted for by differences in instructors' requirements; the other 36 percent of this variation was attributable to factors beyond the instructor's control. Interpretations of E_N, the measure intended to assess this 36 percent of "effort," are offered in a later section of this paper.

[1] Variables that influence student ratings but that are beyond the control of the instructor.

Only about 38 percent of the variation in D_N scores was attributable to differences in instructors' requirements. The remaining 62 percent of variations in "course difficulty" ratings could not be accounted for by class management decisions. Whereas student effort was determined mostly by instructor requirements, course difficulty appeared to be determined more by the inherent difficulty (complexity, abstruseness) of the concepts introduced by the course and/or the discipline than by the instructor's requirements.

Relationships Among the Extraneous Variables

To what extent did the five measures of extraneous variables overlap? Did they seem to be assessing the same, or different, influences on student ratings? A table of inter-correlations was prepared and is the basis for the following observations.

First, D_N and E_N correlated .47. This was substantially below the .67 found for the unadjusted means for Items 35 (difficulty) and 37 (effort); once instructor requirements were taken into account, the relationship between effort and difficulty was diminished. But it was still substantial; students worked extraordinarily hard in courses that were inherently difficult.

Second, CM and WH were positively correlated (r=.30). Classes with a disproportionate number of motivated students (CM) were likely to contain a disproportionate number of students who typically worked harder at their academic studies than their friends (WH). A possible explanation for this relationship is that students who are highly motivated to obtain a college degree are likely to rate themselves above average on both measures.

Third, D_N was essentially unrelated to either CM (Item 39) or WH (Item 43). Whether a course was inherently easy or difficult had no bearing on motivation to take the course or on academic conscientiousness.

Fourth, E_N had low positive correlations (.30 and .21) with CM and WH, respectively. After the impact of the instructor's requirements was taken into account, student effort was still greater in classes that enrolled a disproportionate number of highly motivated students. This provides statistical confirmation that academic motivation is related to (and probably responsible for) academic effort. The tendency for effort to increase as the enrollment of conscientious students increased was markedly reduced after the impact of instructor requirements were taken into account; WH correlated .43 with Item 37 (unadjusted "effort") compared to the .21 correlation between WH and E_N. Although this was relatively low, it is consistent with our previous speculation that high WH ratings and high E_N scores may both be characteristic of students who are motivated to complete a college degree.

Finally, size of class was essentially unrelated to the other four extraneous variables. A slight negative relationship was found between CM and class size, probably because upper division classes in one's major tend both to attract highly motivated students and to be smaller than average.

In general, inter-correlations among the five measures of extraneous variables were relatively low. They appear to be assessing relatively unique qualities.

Impact of Extraneous Variables on Course Outcomes

A step-wise multiple regression procedure was used to determine how these five measures of extraneous variables could be combined so as to "explain" the maximum amount of variation in measures of outcomes. A total of 15 outcome measures were considered: progress ratings on the 12 learning objectives included on the student ratings form[2] and three global ratings derived from individual items (*As a result of taking this course, I have more positive feelings about this field of study*; *Overall, I rate this instructor an excellent teacher*; and *Overall, I rate this course as excellent*).

The five measures of variables not under the instructor's control accounted for an average of 18.1 percent of the variation in student ratings of progress on the 12 learning objectives. This percentage varied from 11.9 (on the *Critical analysis* objective) to 23.8 (on the objective related to *Professional skills, competencies, and points of view*). These measures accounted for even more variation in two of the three global ratings (*Increased positive attitude toward the subject matter*, 36 percent, and *Excellence of course*, 29 percent); but they accounted for only 8.8 percent of the variation in ratings of *Excellence of the teacher*. It can be concluded that, although most of the differences in course outcomes can be attributed to the quality of the instructor's methods and strategies, a significant portion was due to factors beyond the control of the instructor. There is obviously a need to "level the playing field" when interpreting student ratings of an instructor's effectiveness.

The five extraneous variables differed in terms of their influence on specific outcomes. The unique significance of each is described in the following paragraphs.

1. Ratings on **Course Motivation** (CM, Item 39. *I really wanted to take this course regardless of who taught it*) were positively correlated with all 15 outcomes. It was the most potent predictor for all 3 global ratings and for 5 of the 12 ratings of progress on individual objectives. Four of the objectives for which it was most relevant were those most frequently chosen by instructors-the two cognitively oriented objectives (*Factual knowledge* and *Principles and theories*) and the two that focus on applications of learning (*Applications and Professional skills, viewpoints*). Other extraneous variables were more important in explaining variance on objectives related to intellectual development (*Broad liberal education, Values development*, and *Critical analysis*), expressiveness (*Creative capacities* and *Communication skills*), and *Team skills*. Results on lifelong learning objectives were mixed; CM was the best predictor of progress ratings for the objective concerned with *Acquiring an interest in learning*, but only third most important for the objective concerned with *Finding and using resources*.

In summary, instructors in classes with highly motivated students had a considerable advantage over those teaching classes with poorly motivated students. Across all 12 objectives, the average progress ratings in classes when CM was 4.5 or above was 4.24; for those where CM was less than 3.0, this average was only 3.64.

Findings for the objective of *Developing skill in expressing myself orally or in writing* merit special comment. CM correlated only .09 with progress ratings on this objective; although its regression weight was statistically significant, it was the last of the extraneous variables to enter the regression equation. Other extraneous variables, especially disciplinary difficulty (D_N, see later discussion), were much more relevant to progress ratings on this objective.

CM correlated higher with two of the global criteria (*Increased positive feelings toward the field of study; Excellence of course* – correlations of .51 and .47, respectively) than with any of the other 13 criteria. Students who were strongly motivated to take a course appeared to be predisposed to have a favorable regard for both the course and its discipline.

[2] Analyses of these 12 measures were based only on classes for which the instructor identified the objective as "Important" or "Essential."

2. The second extraneous variable, **Work Habits** (WH, Item 43. *As a rule, I put forth more effort than other students on academic work*) was also positively related to all 15 outcome measures and made a significant contribution to the prediction of each. It was second only to CM in its relevance for ratings on five objectives (the two cognitively-oriented ones, the two concerned with applications, and the one focused on acquiring an interest in learning more) as well as on the objective for which CM was of relatively little importance (*Developing skill in expressing myself orally or in writing*). For the other six objectives, it was the extraneous variable most closely related to progress ratings.

Students who have developed strong academic habits were more inclined than others to report high progress on learning objectives. Progress ratings in classes where the WH average was 4.5 or higher averaged 4.53 across the 12 objectives; for those with WH scores below 3.0, this average was only 3.44.

Classes that contained a disproportionate number of such students were generally rated more favorably than were other classes. But global ratings were less influenced by this extraneous measure than by CM; positive "biases" in overall ratings were due more to differences in course motivation than to differences in academic work habits.

3. The third measure of extraneous circumstances, **Number Enrolled** (N), had low negative correlations with each of the 15 ratings of outcomes. In general, ratings for large classes were less favorable than those for small classes. Although the effects were not large, they made a statistically significant contribution to the explanation of variation in ratings of all criteria except for three individual progress ratings: *Factual knowledge*; *Team skills*; and *Values development*). Some speculative explanations for these exceptions include: (1) Many studies have shown that factual knowledge can be taught nearly as effectively in large lecture classes as in small discussion groups. (2) Classes stressing "team skills" and "values development" often rely on teaching formats that feature small group interactions; regardless of class size, as many small groups as needed can be developed to address these objectives.

Class size was especially important in explaining the variance of ratings of three objectives: the two concerned with applications (*Applications of course material*; *Professional skills and viewpoints*) and the communications objective (*Developing skill in expressing myself orally or in writing*). Although attention to the questions and needs of individual students contributes to learning on almost all of the objectives, such attention appears to be especially facilitative of growth on these three objectives. Findings concerning the relevance of class size reflect the limitations it imposes on the instructor's capacity to offer individualized assistance.

In terms of overall ratings, large classes were generally rated less favorably than small ones. However, this was not a very large effect; class size entered the regression equation either last or next to last on all three of the global criteria.

4. **Disciplinary Difficulty** (D_N) was negatively related to the three global ratings and to 10 of the 12 ratings of progress on individual objectives. In general, progress ratings on learning objectives were negatively affected when course concepts and content were inherently complex or abstruse. This was especially apparent for the objective concerned with *Communication skills*; on this criterion, D_N was the most influential of the five extraneous variables. It can be inferred that, when course concepts were unusually difficult, students were more likely to focus on grasping these concepts rather than on improving their communication skills, even if such improvement was one of the instructor's objectives. D_N was chosen second or third in all other regression analyses except for the two concerned with applications (*Applications of course material*; *Professional skills and viewpoints*) where it was the last variable selected.

Results for the two most frequently stressed objectives (cognitively-related objectives *Factual knowledge* and *Principles and theories*) were different than for the other objectives. Zero-order correlations were low positive for these two measures. Furthermore, when all five extraneous variables were optimally combined, the regression weights for D_N were positive and third in importance among the five measures of extraneous variables. Previous studies of the impact of "difficulty" on student ratings concluded that, contrary to popular belief, it had a positive relationship to ratings of effectiveness. This conclusion was confirmed when the measure of effectiveness concerned cognitive outcomes. Since these are the most common outcomes pursued in higher education, the current findings can be regarded as consistent with those previously reported. However, when other types of objectives were considered, disciplinary difficulty consistently had a negative regression weight. Instructors choosing objectives other than cognitively-oriented ones were disadvantaged if their subject matter was inherently difficult.

D_N was correlated negatively with each of the three global ratings and was either the second or third variable chosen in the regression analysis. As with most of the specific learning objectives, instructors teaching highly difficult subject matter were at a disadvantage in obtaining high overall ratings.

5. The last of the five extraneous measures, **Student Course Effort** (E_N), had low correlations with the 15 criteria, about half of which were positive; working harder than expected on the basis of the instructor's requirements had only a minor (and inconsistent) relationship to course outcomes. Its regression weight, however, was significant for all but one of the 15 and, for individual progress ratings, was consistently negative. These findings contrasted with those for uncorrected "effort" (ratings on Item 37) where correlations with progress ratings ranged from +.23 to +.47 on the12 objectives.[3] As expected, working hard generally resulted in higher progress. But once the extraneous variables of course motivation (CM), work habits (WH), and disciplinary difficulty (D_N) were taken into account, students who worked harder than expected on the basis of the instructor's requirements reported lower progress. How can this be explained?

First, it will be recalled that E_N was positively related to D_N (the inherent difficulty of the course or discipline); it makes intuitive sense to expect "extra" effort from those enrolled in inherently difficult courses. Second, course motivation (CM) and academic work habits (WH) were also positively related to E_N; those who were motivated to enroll and who had established strong academic habits tended to make academic effort over and beyond that expected on the basis of the instructor's requirements. These positive relationships are consistent with the previous suggestion that students with strong academic habits and strong course motivation probably have an above average desire to obtain a college degree; such students could also be expected to work unusually hard (have high E_N scores).

After those relationships have been taken into account, progress ratings for classes with high E_N scores were lower than for those with low E_N scores. An obvious possibility is that students who made efforts beyond those accounted for by their academic work habits,

[3] See the Center's *Technical Report #12: Basic Data for the Revised IDEA System*.

their course motivation, the inherent difficulty of the discipline, and the instructor's class requirements may lack the background needed to grasp course content or feel inadequate for some other reason.

Unlike specific progress measures, overall ratings of *Increased positive feelings about the field of study* and *Excellence of the course* were positively related to E_N, although both the correlation and regression coefficients were low. For the other global measure, *Excellence of the teacher*, the correlation and regression weight for E_N were slightly negative. In classes where students made extraordinary efforts, perceptions of the course were positively affected; but those of the instructor were negatively affected. These effects were statistically significant, but relatively slight in their influence.

Conclusions

1. Measures of five extraneous circumstances accounted for about 15-20% of the variation in ratings of course outcomes. The supposition that student learning is impacted by circumstances beyond the instructor's control was confirmed.

2. Although each of the measures made an independent contribution to the prediction of most outcomes, the two most important ones were **Course Motivation** (CM) and **Work Habits** (WH). The former was especially influential in objectives concerned with acquiring a cognitive background or with applying that background; it also was the primary predictor of global ratings. WH was the most influential in promoting progress on other types of objectives. Intuitive expectations that ratings of self-learning would be higher if students were highly motivated and if they had established conscientious approaches to academic work were confirmed.

3. Those who taught large classes generally received lower ratings than those who taught small classes; however, for classes that stressed factual knowledge, team skills, and/or values development, size of class was unrelated to progress ratings. Overall ratings were slightly lower for large classes. The belief that learning is enhanced in small classes was generally confirmed, although the effect was not a large one.

4. Classes were made more or less difficult by the instructor's decisions regarding reading assignments, other outside work, and the degree of academic challenge offered to students. After these factors were taken into account, considerable variability in difficulty ratings among classes remained, suggesting large inherent differences among disciplines in the complexity and abstruseness of their subject matter. The measure of disciplinary difficulty employed in this study was positively related to progress ratings on the two most frequently chosen objectives (*Factual knowledge*; *Principles and theories*). However, ratings of progress on other kinds of objectives and all overall ratings were negatively related to this measure of disciplinary difficulty. In many instances, then, instructors teaching subjects whose content is inherently difficult to grasp were disadvantaged when students rated course outcomes; but this was not true in classes where cognitive objectives were stressed.

5. The amount of effort students make in a given class was affected by the demands made by the instructor. It was also affected by the level of student motivation and by their academic work habits. Finally, it was affected by the inherent difficulty of the subject. After all these factors are taken into account, an unexplained difference among classes in student effort remained. Statistical analysis showed that this "residual source of effort" generally had a small negative influence on student ratings. It was inferred that effort unexplained by the instructor's requirements, the course difficulty, or the student's motivation and academic work habits, may be indicative of students whose academic backgrounds were inadequate for the class or who lacked confidence for some other reason. Such students would likely work harder but learn less than their peers.

Recommendations

The evidence cited in this paper underscores the value and importance of adjusting ratings by assessing relevant external circumstances. It seems desirable to continue the search for such variables, both to provide fairer comparisons among classes and instructors and to improve understanding of the dynamics of learning.

This study focused on adjustments that are made on The IDEA Center's Diagnostic (long) Form. Steps should be taken to make comparable adjustments on its Short Form. A preliminary step to this end was taken in the fall of 2002 when the item assessing academic work habits (*As a rule, I put forth more effort than other students on academic work*) was added to the Short Form, replacing an obsolete item. An experimental item was also added in an effort to provide a more direct measure of EN (*My background prepared me well for this course's requirements*).

It is recommended that tests be made of the assumption that D_N is a measure of disciplinary difficulty. This will require comparing scores on this variable within and among disciplines. If such analyses confirm the assumption,[4] an investigation should be undertaken to determine if a simple categorization of disciplines on the basis of the D_N scores would provide an adequate proxy measure. For example, disciplines with the lowest 10% of D_N scores could be categorized "1," those with the highest 10% categorized "5," and those with intermediate scores categorized "2," "3," or "4." If such a measure yielded results comparable to those found for D_N, this would not only simplify the measure but would also make it possible to base adjustments for both the Diagnostic and Short Forms on the same measures of extraneous variables.

The IDEA Center
211 South Seth Child Road
Manhattan, KS 66502
Ph: 800.255.2757 or 785.532.5970
Fax: 785.532.5725
E-Mail: IDEA@ksu.edu
Home page: *www.idea.ksu.edu*

[4] That is, if the differences among classes within a discipline tends to be relatively small compared with the differences among disciplines when all classes are considered.