

Technical Paper 3: The Relationship between ETIP Case Essay Scores and Relevancy Scores

Eric Riedel, Ph.D.
Center for Applied Research and Educational Improvement (CAREI)
University of Minnesota

David Gibson, Ph.D.
The Vermont Institutes

Shyam Boriah
Center for Applied Research and Educational Improvement (CAREI)
University of Minnesota

Abstract:

The statistical relationship between scores assigned to ETIP case essays and measures of the relevancy of the user's search to the case question are explored. The sample includes data from 117 student users in ten course sections over two semesters. There was no significant relationship between essay scores and the relevancy index (sum of relevancy weights divided by the number of steps taken in the case). There was a weak relationship between the number of steps taken in a case and essay scores such that highly rated essays reflected at least a minimal search of the case. An alternative measure of relevancy, the number of unique items relevant to the case question accessed, did have a statistically significant relationship to essay scores. This relationship was strongest in later rather earlier cases completed by the user.

Original draft released on January 13, 2004. Final draft released on April 21, 2005. Correspondence regarding this paper can be directed to the first author at the Center for Applied Research and Educational Improvement (CAREI), University of Minnesota, 275 Peik Hall, 159 Pillsbury Avenue SE, Minneapolis, MN 55455, riedel@umn.edu.

Executive Summary

The following paper explores the relationship between relevancy scores and instructor-assigned essay scores derived from students' use of the ETIP cases at test-bed sites in fall 2002 and spring 2003 semesters. The relevancy index (sum of relevancy weights of items accessed by students / number of steps taken by students) was used as the main relevancy score measure. With fall 2002 data, essay quality was measured by an additive index of six instructor-assigned scores. With spring 2003 data, essay quality was measured both by an additive index of three instructor-assigned scores and an automatically calculated global score which weighted the "decision" criterion more heavily than the other two scores.

In neither semester did the relevancy index have a statistically significant relationship to essay quality. Three possible explanations were explored to explain this finding: (1) the case search is unrelated to essay quality; (2) access to relevant information positively predicts essay quality but this relationship is obscured by the current relevancy index which values efficient searches; and (3) both relevant and irrelevant is needed by students to produce high quality essays.

In both semesters, high essay scores were associated with at least a minimal search of the case. Users who made typically 30 or fewer steps in a case, even if they scored well on the relevancy index, wrote lower quality essays than those users who made a more extensive search. Essay quality was not related to the number of steps taken in a case after this threshold was met.

An alternative measure of relevancy, the relevancy total (the sum of relevancy weights of items accessed by students not discounted by number of steps taken in a case), did have a modest relationship to essay quality, especially in the third case of spring 2003. Other alternative relevancy indices were explored using the third case in spring 2003. It appeared that a simple count of the number of different relevant items accessed was the strongest predictor of essay quality. Other measures, which were more directly influenced by the number of steps taken in the case overall, did not perform as well.

The results lend support to the second and third hypotheses. The search of the case does matter to essay quality – both in terms of users having an overall understanding of the case and having access to relevant information. A two-stage model is proposed which posits that first users must gain a general knowledge of the case through exploration of both relevant and irrelevant information in order to write a high quality essay. In the second stage, having gained a general understanding of the problem space, users can focus on relevant information in answering the case challenge.

Introduction

The Educational Theory into Practice Software (ETIPS) originated with a grant in 2001 from the U.S. Department of Education's Preparing Tomorrow's Teachers to Use Technology (PT3) program. Since its inception these online cases were designed to provide a simulated school setting in which beginning teachers could practice decision-making regarding classroom and school technology integration guided by the Educational Technology Integration and Implementation Principles. In each case, users are given a case challenge based on one of these six principles about how they would use educational technology in the specific scenario¹. They then can search out information about the school staff, students, curriculum, physical setting, technology infrastructure, community, and professional development opportunities. After responding to the case challenge in the form of a short essay, users are given feedback about their essay and case search. (View cases at <http://www.etips.info/>.)

The present paper draws on research and evaluation data gathered on the actual use of the cases during part of the test phase of the ETIP Cases project. It is part of a series of technical papers aimed at informing project staff, users of these cases, and researchers of educational technology more generally. This paper focuses on two measures of user performance within the cases – how expertly the user searches for information in the case (relevancy scores) and scores assigned to students' essay scores based on rubrics given to instructors in the instructional supports for cases.

The purpose of this paper is to examine the relationship between relevancy 1 scores assigned to student searches of ETIP cases and measures of students' essays written in response to those cases. It builds on analyses of both sets of measures provided in Technical Reports Numbers 1 and 2 which examine these sets of measures separately. The question underlying the analyses in this paper is whether a students' search of relevant information in a case (as defined by the project staff) is related to the quality of the essay they write after searching through the case. More specific questions guiding the analyses include:

- How strong is the relationship between essay scores and relevancy?
- Where is there a strong relationship?

¹ These six principles state the conditions under which technology use in schools has been demonstrated to be most effective. Case 1: Learning outcomes drive the selection of technology. Case 2: Technology provides added value to teaching and learning. Case 3: Technology assists in the assessment of learning outcomes. Case 4: Ready access to supported, managed technology is provided. Case 5: Professional development targets successful technology integration. Case 6: Professional community enhances technology integration and implementation. See Dexter, S. (2002). eTIPS-Educational technology integration and implementation principles. In P. Rodgers (Ed.), *Designing instruction for technology-enhanced learning* (pp.56-70). New York: Idea Group Publishing.

Description of Relevancy Scores

Relevancy scores were developed by the ETIP Cases project as a way of summarizing student searches within the cases. Specifically they were intended to show to what degree the student accessed case information necessary for answering the questions in the case challenge and thus serve as one measure of technology integration expertise. Each case challenge contained questions related to one of educational technology and integration principles selected by an instructor for the assigned case. Thus relevancy is a concept defined by the specific questions asked.

Relevancy scores were previously assigned by ETIP cases project staff acting as technology integration experts. Each piece of case information was rated as “0 Not Relevant”, “1 Somewhat Relevant”, or “2 Relevant” to answering the question posed in the case prologue. A relevancy index was subsequently calculated as the sum of relevancy scores from the items accessed by a user in a case divided by the number of steps taken in the case. A “step” is defined as accessing an individual piece of information in the case. Returning to the same item later in a search would count as an additional step. (See Appendix A for an example of a case and calculation of the relevancy index.)

The relevancy index had a range of 0 to 2. Based on a large sample of test-bed users, the index had an approximately normal distribution with the mean between .86 and .93 and a standard deviation between .29 and .35 depending on the semester and case. Relevancy index scores generally increased over time. There was slight evidence of increases in the relevancy index over multiple cases by users. The relationship between number of steps taken in a case and the relevancy index was generally negative. As the number of steps taken increased the relevancy index score decreased.

Description of Essay Scores

In fall 2002, instructors were asked to score each student essay with six scores covering validation of the case question, evidence used in support of their decision, and the decision responding to the case question. (See Appendix B for the specific rubric used.) A seventh score, representing an overall judgment of the essay, is used only briefly here. An essay scale was constructed by adding together each of the six scores. The scale had a Cronbach’s alpha equal to .87 which was not improved by removing any of the scores. It ranged from 0 to 12. The scale had a mean of 7.06, a median of 7, and a standard deviation of 3.48. It was skewed slightly towards higher values (skewness=-.285, standard error of skewness=.299).

In spring 2003, instructors were asked to score each student essay with three scores covering validation of the case question, evidence used in support of their decision, and the decision responding the case question. (See Appendix B for the specific rubric used.) Two outcome measures of essay quality were used. The first (global score) was automatically calculated from the three instructor assigned scores. It had a range of 0 to 4, a mean of 2.8, a median of 3, and a standard deviation of 1.08. The global score

was moderately skewed towards the higher values (skewness=-.725, standard error of skewness=.369). The second outcome measure was a scale constructed by adding the values of the three instructor assigned scores. The essay scale had a Cronbach's alpha equal to .55 and a range from 0 to 6. The scale had a mean of 4.68, a median of 5, and a standard deviation of 1.29. It was skewed significantly towards higher values (skewness=-1.12, standard error of skewness=.369).

Individual criterion scores were highly correlated within cases while scores across multiple cases for the same individual were only moderately correlated. There was little evidence for individual growth in essay scores over multiple cases.

Sample

The samples are restricted to those students included in the analyses of essay and relevancy scores previously (see ETIP Case Technical Reports Numbers 1 and 2). In fall 2002, the sample is limited to include data from the first case only. It encompasses a range of different cases however. In spring 2003, the sample includes data from instructors who scored three cases. Information from a user was included if that user returned a pre-semester survey, completed each of the cases assigned in the correct order, and made use of at least four separate steps in each case. These criteria assured that the data utilized met human subjects' protection requirements, the user made a reasonable attempt to follow course instructions, and that the user did not encounter insurmountable technical problems. Tables 1 and 2 summarize the samples for fall 2002 and spring 2003 respectively.

Table 1. Sample of Essay Scores for Fall 2002

Instructor	Course	eTIP	Level	Number of Scored Essays Case 1
Instructor A	Foundations	2	Elementary	9
Instructor G	Foundations	6	Elementary	12
Instructor H 1	Methods	1	Elementary	16
Instructor H 2	Methods	1	Elementary	9
Instructor I	Foundations	2	Elementary	5
Instructor L	Foundations	2	Secondary	13

Table 2. Sample of Essay Scores for Spring 2003

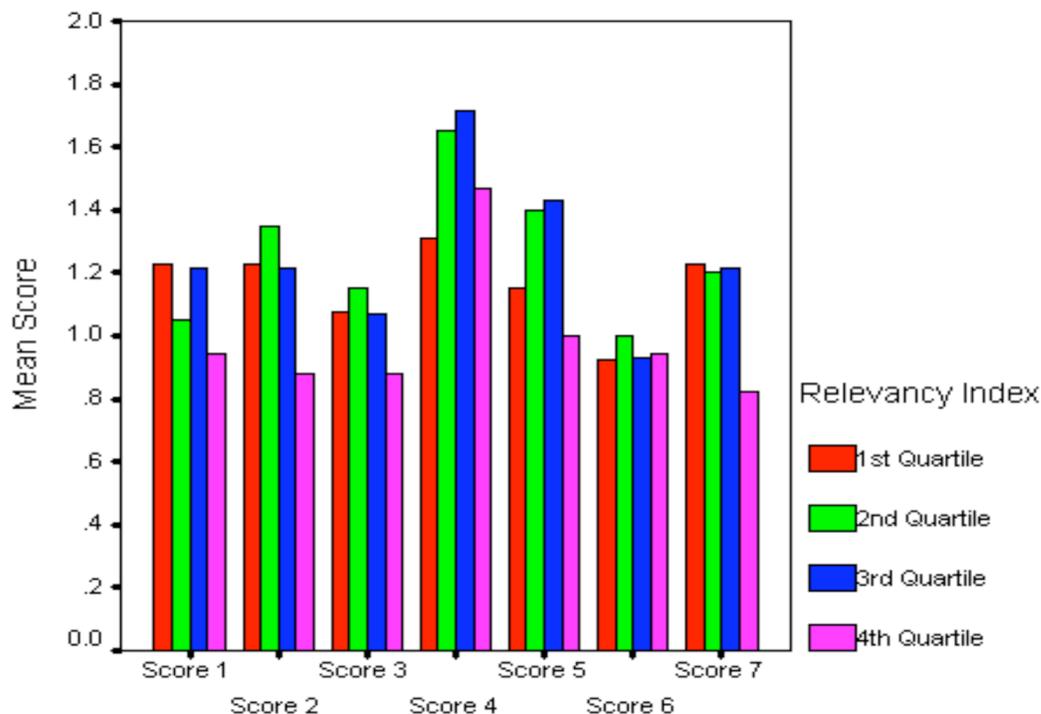
Instructor	Course	eTIP	Level	Number of Scored Essays Case 1	Number of Scored Essays Case 2	Number of Scored Essays Case 3
Instructor J	Ed Tech	2	Secondary	11	11	11
Instructor K	Ed Tech	2	Secondary	11	11	11
Instructor P 1	Methods	2	Elementary	6	6	5

Fall 2002 Results

To simplify the analysis, students were assigned to quartiles on the relevancy index. The first quartile includes those students who scored equal to or below the 25th percentile of scores. The second quartile includes students who scored in the 26th to 50th percentile range. The third quartile includes students who scored in the 51st and 75th percentile range. The fourth quartile includes students who scored in the 76th to 100th percentile range.

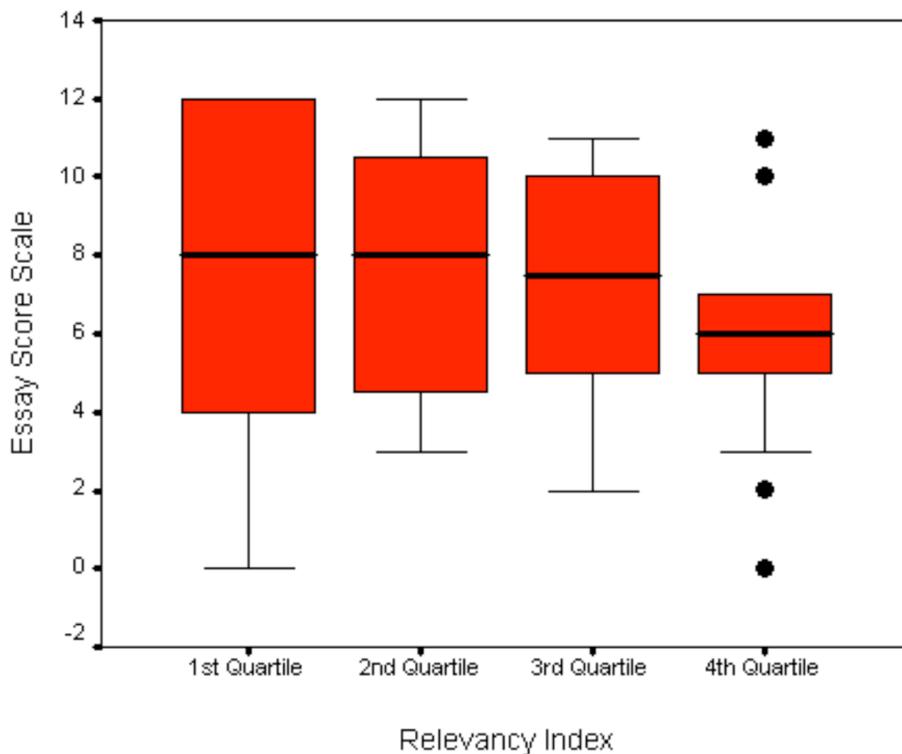
There appear to be only slight differences among the relevancy index quartiles in their means for each of the seven instructor assigned scores. A series of Kruskal Wallis tests run with each of the scores as a dependent variable and the relevancy index quartiles as the independent variable revealed no significant differences for any score among the four quartiles (Score 1: $X=1.265$, $p=.738$; Score 2: $X=3.446$, $p=.328$; Score 3: $X=1.657$, $p=.646$; Score 4: $X=2.275$, $p=.517$; Score 5: $X=3.545$, $p=.315$; Score 6: $X=.111$, $p=.990$; Score 7: $X=3.143$, $p=.370$).

Figure 1. Relationship between Essay Scores and Relevancy Index on First Case (Fall 2002)



It is noteworthy that the fourth quartile generally has the lowest mean for each score and that for some scores, the relationship between the relevancy index and mean essay score appears curvilinear with those in the middle ranges of relevancy having the highest mean score. To further explore the relationship between the relevancy index and essay score, the figure below shows a box plot of the essay scale score by relevancy (Figure 2). Those having the highest relevancy scores have a lower median essay score than other respondents. There is also greater variance in the essay scales for the lowest quartile of relevancy scores. A median test, which doesn't involve an assumption of equality among the variances, found no statistically significant differences among the median values of the four relevancy groups ($X=4.519$, $p=.211$).

Figure 2. Box Plots for Mean Essay Scale Score by Relevancy Index for First Case (Fall 2002)



These results appear problematic. If the relevancy index measures access to information relevant to answering the case question, then one would expect measures of the quality of those answers to reflect that. But in the analyses above, they do not. Three hypotheses are posed to explain why this is so:

- (1) The essay scores mainly measure student writing ability. They bear little relation to how the student searched the case.

- (2) While students need to access relevant information in the case to provide high quality essay answers, the relevancy index emphasizes efficiency to a degree which obscures this relationship. For example, a student could receive a high relevancy index score by accessing an incomplete set of relevant items but nothing else.
- (3) Students need access to relevant and irrelevant information in the case to produce high quality essays. The relevancy weights assigned to individual pieces of information may be incorrect insofar as specifying what is needed to answer the case question.

The next analyses investigate support for each of these hypotheses by using a measure of case exploration – the number of steps taken by the student in a case. The graph below shows the essay scale score by the number of steps taken (Figure 3). It appears there are a minimum number of steps needed to produce a high quality essay. None of those students taking less than 12 steps scored had an essay scale score above 5. Those students taking 12 to 28 steps produced a range of essay scale scores. They could do equally poorly or well. Those students taking more than 28 steps in the case generally did moderate to well on the essay. Only a handful scored below 5. There do not appear to be increases with each increase in the number of steps after 29 steps.

Figure 3 shows that the search of the case does matter to essay quality and is not merely measuring writing ability apart from student analysis of the case. But number of steps taken may be better conceived as a categorical variable indicating whether or not the student made a substantial search of the case (defined as less than 29 steps here). Those students making less than 29 steps had a mean scale score of 6.09 while those making more than 29 steps had a mean scale score of 8.24 ($t=2.573$, $p=.012$).

The first hypothesis posed, that the search is unrelated to essay quality, is not supported by the evidence. There appears to be a minimum search needed to produce a high quality essay. Even those respondents who score highly on the relevancy index but access only a few items from the case overall tend to do poorly on the essay. This provides some evidence for the second hypothesis, the value of efficiency built in to the relevancy index can work against conducting the search necessary for answering the case challenge. This is shown in Figure 4 in that the relevancy index did not distinguish among those respondents taking less than 29 steps in the case – all quartiles of the relevancy index did equally poorly on the essay

Figure 3. Number of Steps Taken by Essay Scale Score in First Case (Fall 2002)

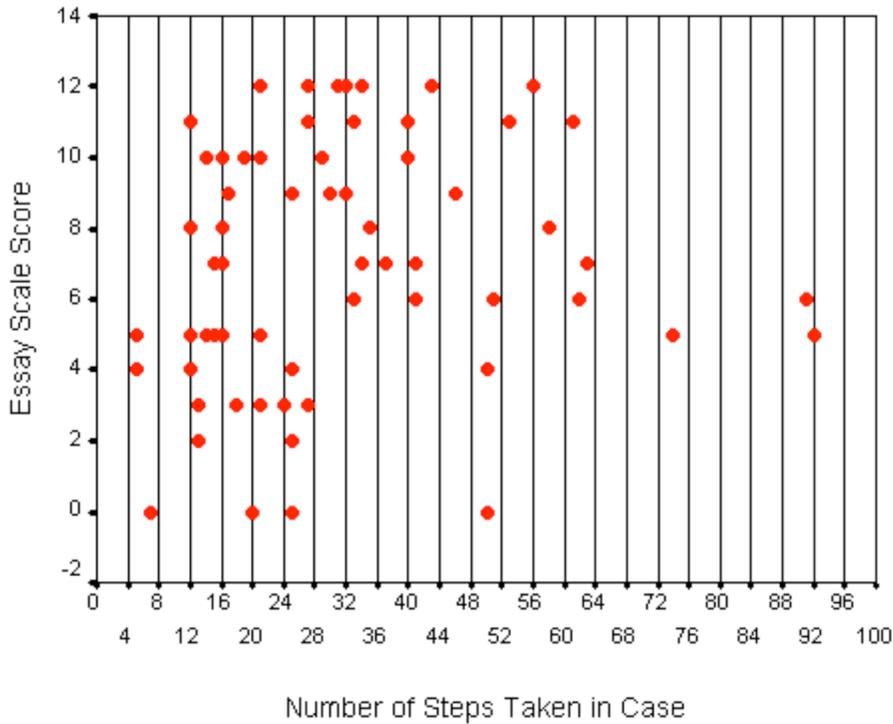
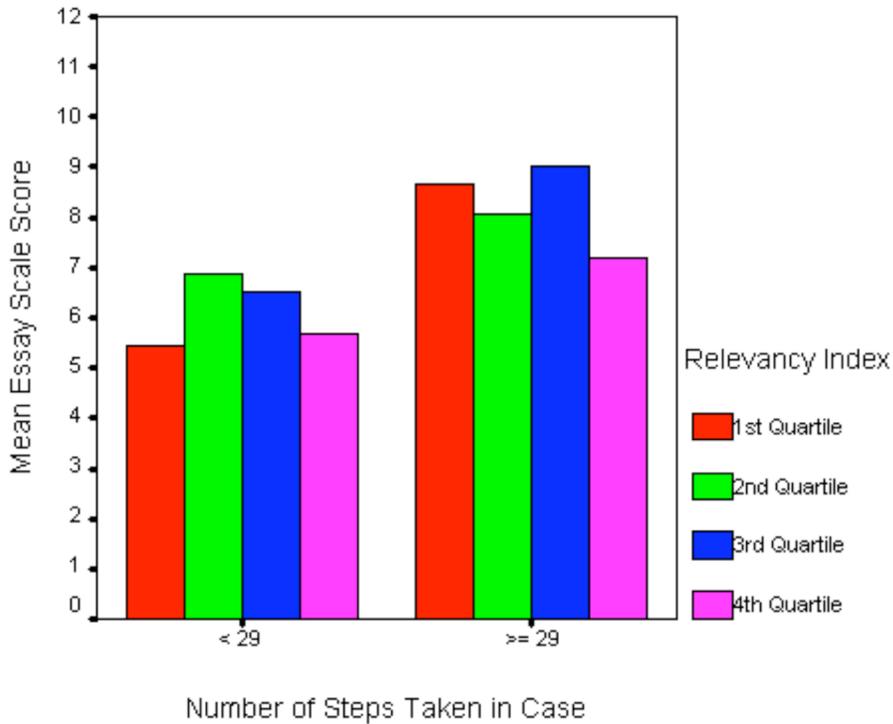


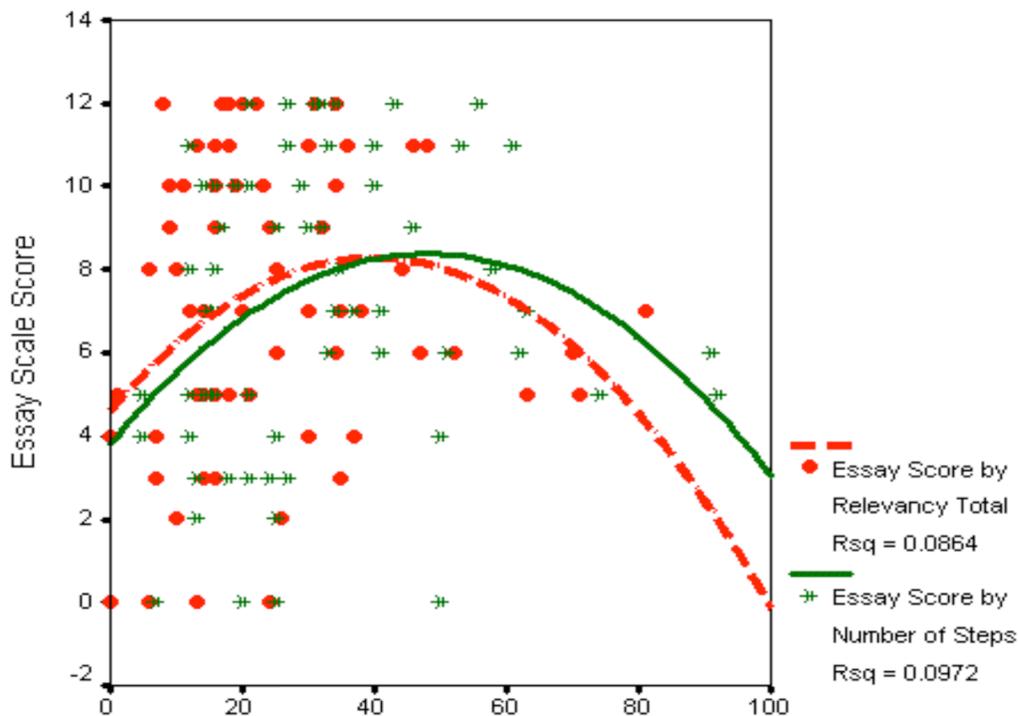
Figure 4. Mean Essay Scale Scores by Relevancy Index and Number of Steps (Fall 2002)



The third hypothesis essentially states that the case search matters but relevancy, as it assigned here, does not. To test the third hypothesis the relevancy total is used. Relevancy weights are summed together to create this score but not discounted by the number of steps taken in the case. It is similar to the count of steps taken in case (each step is assigned a value of 1), in that users who do an extensive search of the case will also tend to have a high relevancy total.

The scatter graph (Figure 5) below compares the values of relevancy total and number of steps against the essay scale score. Both have a very similar relationship to the essay scale score. This relationship is best modeled as a quadratic function so that the essay scale scores increase with increasing relevancy total and number of steps at the lowest levels of relevancy total and number of steps. This relationship disappears towards higher values of relevancy total and number of steps and even reverses at the highest levels. Relevancy weights, as they are defined in these cases, do not appear to make much of a difference in essay quality beyond serving as a rough proxy for the extensiveness of the case search. For the first case in fall 2002, the evidence lends some support to the third hypothesis – that relevancy, as it is currently defined, does not matter to essay quality.

Figure 5. Scatter Graph of Essay Scale Scores by Relevancy Total and Essay Scale Scores by Number of Steps (Fall 2002)



In general, the impact of the relevancy total is indistinguishable from that of the number of steps taken in a case. Students need to undertake a certain minimal search of the

case to score well on their essay. Students conducting an exhaustive search of the case, however, do not appear to do better and perhaps do even worse than students who do a moderate but complete search. If access to relevant items mattered, we would expect to see a stronger relationship between relevancy total and the essay scale score than between the number of steps and the essay scale score. That is not the case and relevancy total functions mainly as a proxy for the extensiveness of the search rather than qualities of that search. The preceding deals only with the first case that students did and it may be that relevancy does not begin to matter until students have become more comfortable with the assignment and software. The following analyses examine this possibility by including data from sequential cases.

Spring 2003 Results

To aid analysis, students were assigned into relevancy index quartiles. This was undertaken separately for each of the three cases. Figure 6 shows how the mean essay scores for each of those quartiles on the first case. Figures 7 and 8 repeat this for the second and third cases respectively. There are no apparent differences in how each quartile did on the essay for any score on any case.

For the first case, a series of Kruskal Wallis tests run with each of the scores as a dependent variable and the relevancy index quartiles as the independent variable revealed no significant differences for any score among the four quartiles (Score 1: $X=2.972$, $p=.396$; Score 2: $X=2.323$, $p=.508$; Score 3: $X=4.799$, $p=.187$). For the second case, a series of Kruskal Wallis tests run with each of the scores as a dependent variable and the relevancy index quartiles as the independent variable revealed no significant differences for any score among the four quartiles (Score 1: $X=1.601$, $p=.659$; Score 2: $X=.321$, $p=.956$; Score 3: $X=1.885$, $p=.597$). For the third case, a series of Kruskal Wallis tests run with each of the scores as a dependent variable and the relevancy index quartiles as the independent variable revealed no significant differences for any score among the four quartiles (Score 1: $X=3.770$, $p=.287$; Score 2: $X=2.816$, $p=.421$; Score 3: $X=.741$, $p=.863$).

Figure 6. Essay Scores Means by Relevancy Index Quartiles for First Case (Spring 2003)

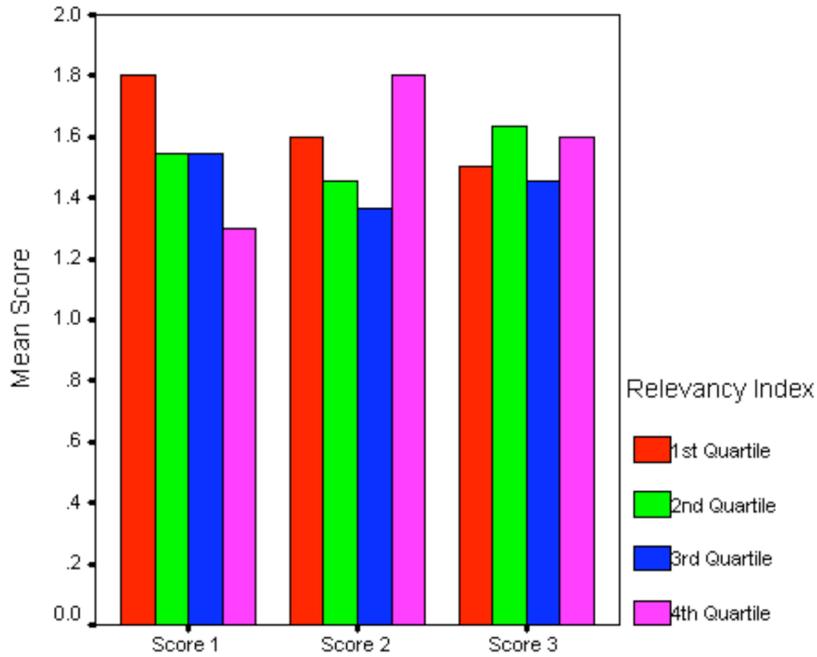


Figure 7. Essay Score Means by Relevancy Index Quartiles for Second Case (Spring 2003)

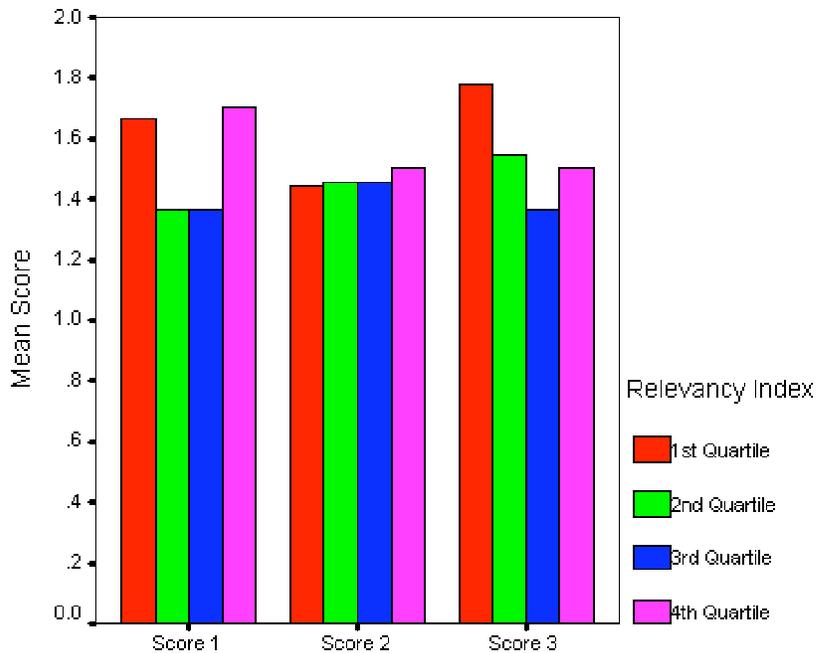
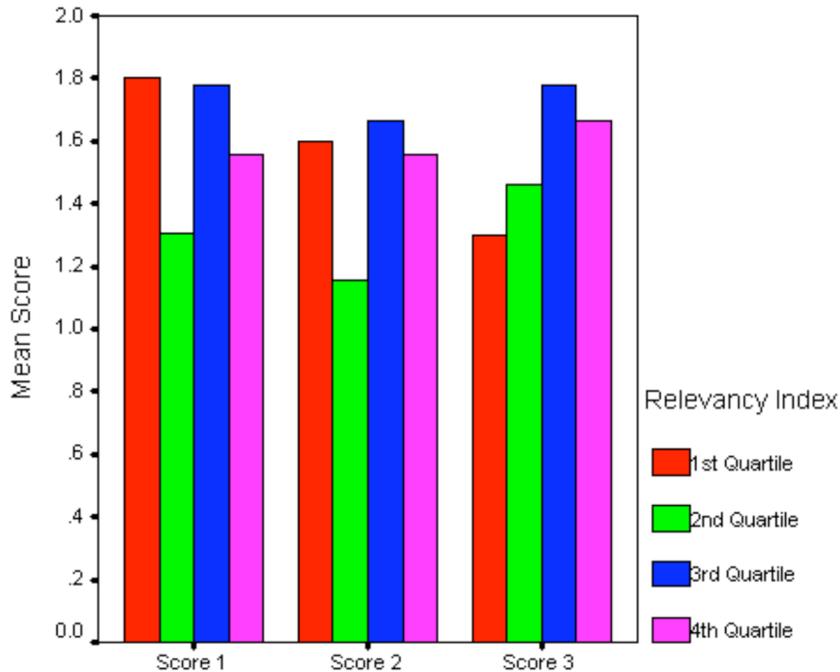


Figure 8. Essay Score Means by Relevancy Index Quartiles for Third Case (Spring 2003)



The preceding tests do not reveal any statistically significant differences among essay scores means by the relevancy index for any of the cases. We repeat the analyses conducted for fall 2002 to explore this lack of relationship, using the three possible explanations as a guide. Both the automatically calculated global score and an additive scale of the three individual scores are used as the outcome variables.

The three box plots (Figure 9-11) below show the distribution of global scores and essay scales scores for each of the quartiles on the relevancy index. Again, there is no statistically significant difference in the levels of either the global score or essay scale score for the first case (global score: $X=.929$, $p=.818$; essay scale score: $X=.815$, $p=.846$), second case (global score: $X=.746$, $p=.862$; essay scale score: $X=.646$, $p=.886$), or third case (global score: $X=3.100$, $p=.376$; essay scale score: $X=2.806$, $p=.422$) based on Kruskal Wallis test. The third case, however, shows a hint of a positive relationship between essay quality and the relevancy index.

Figure 9. Box Plots of Relevancy Index Quartiles by Global and Essay Scale Scores for First Case (Spring 2003)

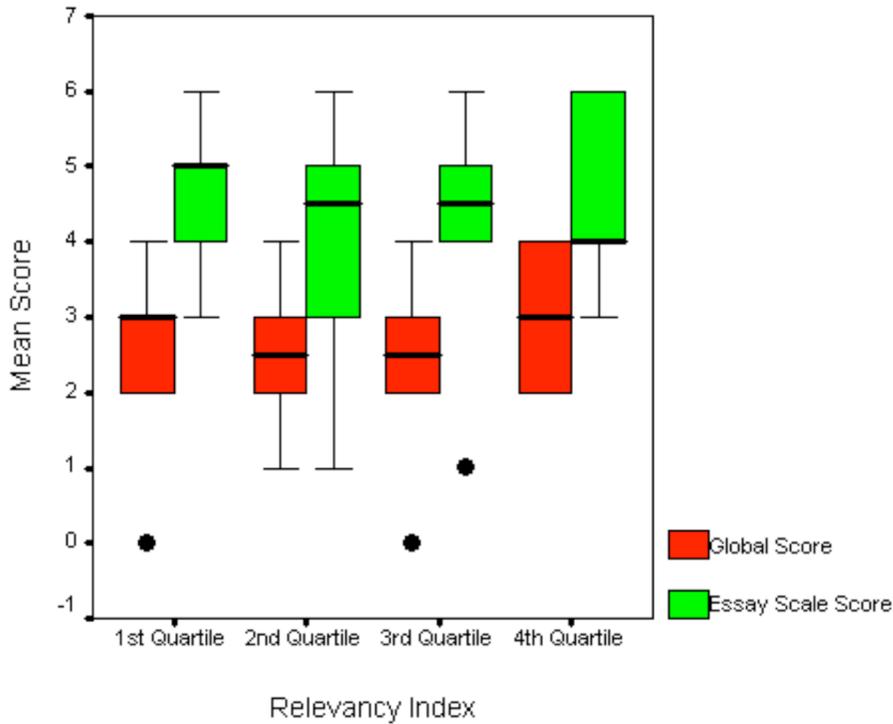


Figure 10. Box Plots of Relevancy Index Quartiles by Global and Essay Scale Scores for Second Case (Spring 2003)

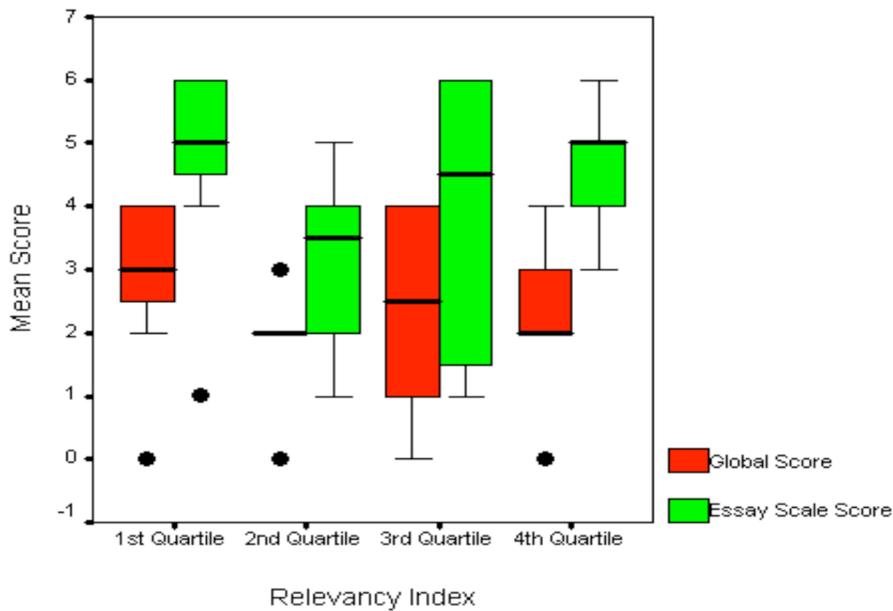
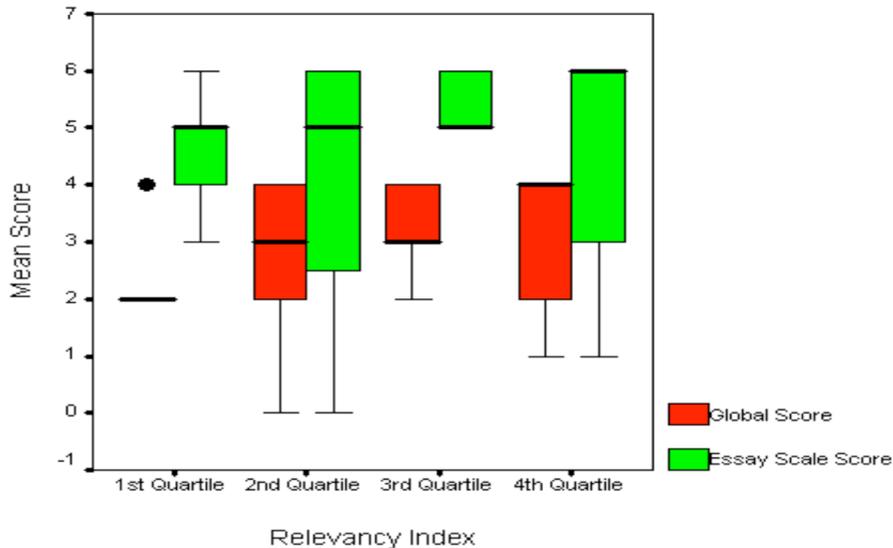


Figure 11. Box Plots of Relevancy Index Quartiles by Global and Essay Scale Scores for Third Case (Spring 2003)



A comparison of the relevancy total and number of steps taken in the case with global and essay scale scores is used to investigate whether the case search has an impact on essay quality and whether relevancy has an impact. Each of the scatter graphs below (Figures 12-14) overlays the global score and essay scale score by both the number of steps taken in the case and the relevancy total. For each of the three cases completed, the pattern is similar. The relevancy total and number of steps appear to have a similar impact on the outcome variables. They are, in fact, highly correlated (first case Pearson $r=.82$; second case Pearson $r=.87$; third case Pearson $r=.87$). A certain minimum number of steps are needed to assure moderate to high level essay quality as measured by both the global score and essay scale score. For the first and second cases, it is approximately 40 steps. For the third case, it is somewhat lower. Beyond this level there is no apparent relationship between the numbers of steps taken or relevancy total and essay quality.

Similar to the analyses conducted for fall 2002, curvilinear regression functions are displayed on the graph. The first and second cases specify that global scores and essay scale scores are a quadratic function (inverted U) of the relevancy total or number of steps. At the lowest levels of the independent variable the outcome increases with subsequent increases in the independent variable but this pattern disappears and even reverses at the higher ends of the independent variable. For the third case, a cubic function (S) is used (which fits better than a quadratic function). This function is similar to the quadratic function except that the pattern reverses again at the highest ends of the relevancy total.

For the first case, essay quality, whether measured by the global score or essay scale, simply does not appear to be strongly related to the number of steps or relevancy total. For the second and third cases, however, there are curvilinear relationship between the number of steps, relevancy totals, and the outcome variables. In general, relevancy totals are more strongly associated with essay quality than number of steps taken.

Figure 12. Global Score and Essay Scale Score by Number of Steps and Relevancy Total on First Case (Spring 2003)

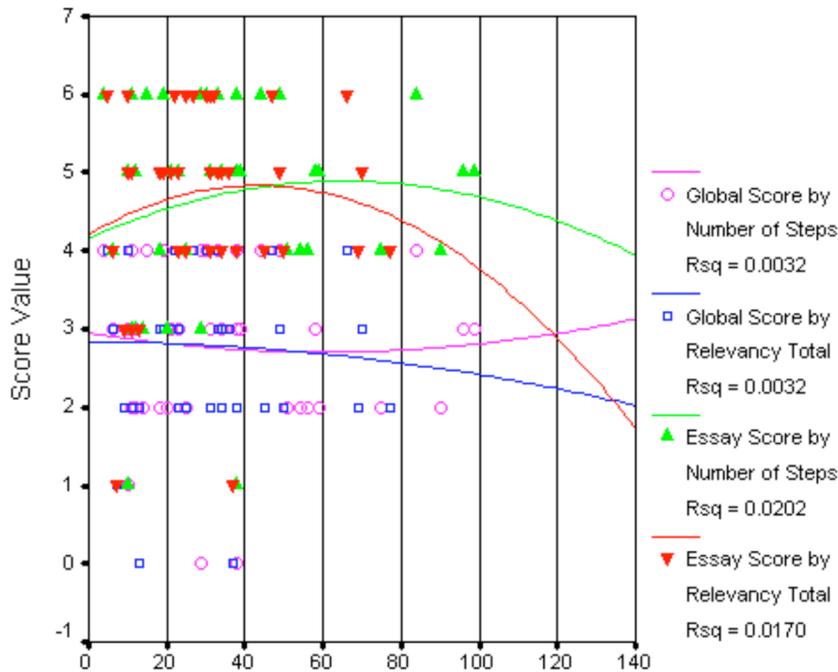


Figure 13. Global Score and Essay Scale Score by Number of Steps and Relevancy Total on Second Case (fit for quadratic function of independent variable)

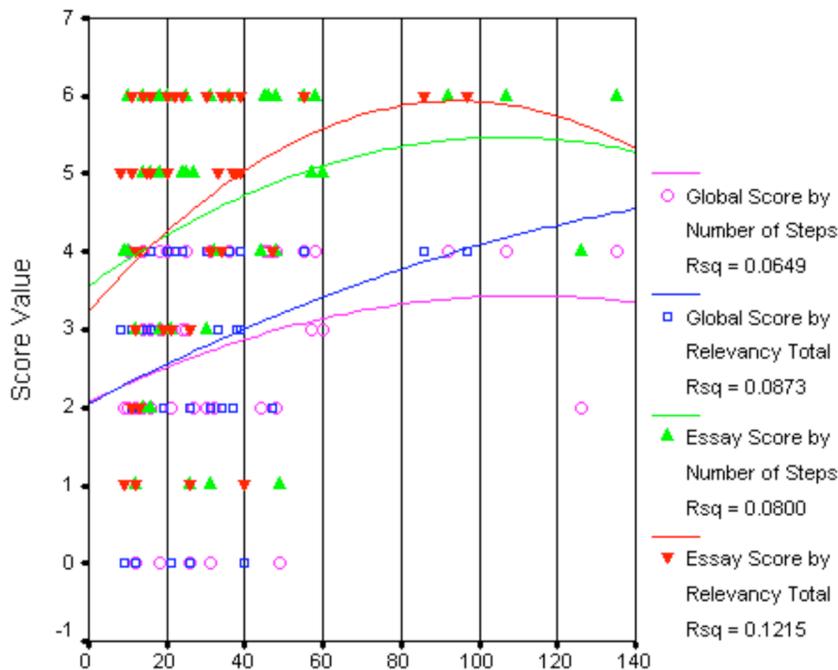
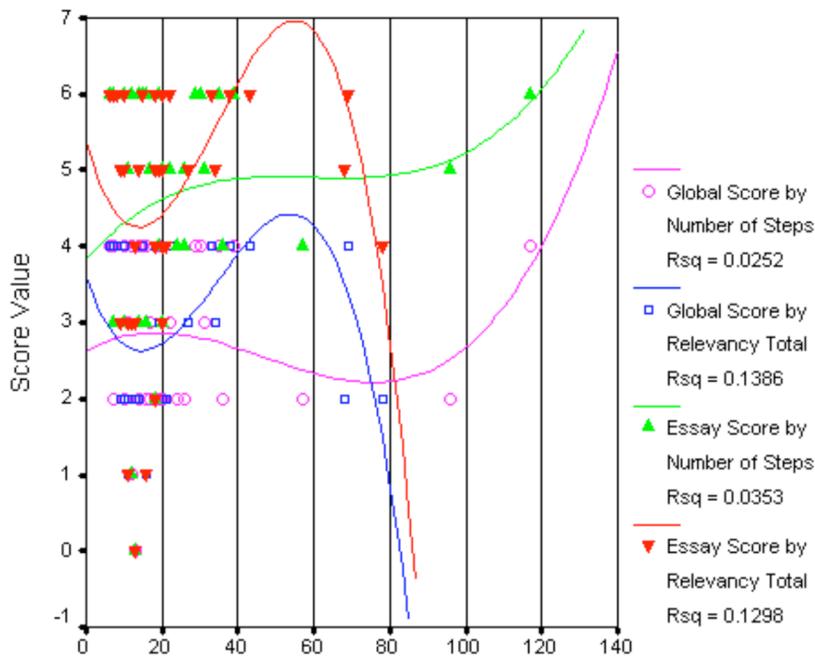


Figure 14. Global Score and Essay Scale Score by Number of Steps and Relevancy Total on Third Case (Spring 2003)



In some ways, the patterns on these graphs replicate those found in fall 2002. Those students who take only a few steps in the case tend to do poorly on the essay – no matter which case. For students who have low relevancy total, the result is the same, suggesting that low relevancy totals reflect an incomplete search of the case for these students. However, on the second and third cases, essay scores – whether measured by the global score or essay scale score – tend to increase with increasing access to relevant information. For these later cases, relevancy total is a better predictor of essay quality than extent of the search. Like the results for fall 2002, these results do not lend support to the first hypothesis – that the case search is unrelated to essay quality. Unlike the results for fall 2002, these results lend modest support to the second hypothesis that relevancy does matter but it the relationship is apparent only when the student is not penalized by the number of steps taken in a case.

Alternative Relevancy Indices

The preceding analyses suggest that the current version of the relevancy index, which divides the relevancy total by the number of steps taken, is problematic because it weights efficient searches too highly. The result is that individuals who access only a few relevant items and nothing else receive a high score on the relevancy index. But the same individual has accessed too little of the case overall to write a high quality essay.

The graphs below present a comparison of four different measures of relevancy in relation to the mean scores on the global and essay scale scores. Since this is a preliminary exploration, the data for the third case in spring 2003 data is used. This is where the strongest relationship between essay quality and relevancy is suspected to occur. The first two measures are alternative constructions of a relevancy index. The first alternative index (B) weights information by proportion of search within each level of relevancy and then treats these levels equally. The second alternative index (C) weights information by proportion of search within each level of relevancy and then additionally weights the levels of relevancy in the total index.

Relevancy Index B =
(Number of Irrelevant Items Selected / Total Number of Irrelevant Items) +
(Number of Semi-Relevant Items Selected / Total Number of Semi-Relevant Items) +
(Number of Relevant Items Selected / Total Number of Relevant Items)

Relevancy Index C =
(Number of Irrelevant Items Selected / Total Number of Irrelevant Items) +
((Number of Semi-Relevant Items Selected / Total Number of Semi-Relevant Items)*2))
+ ((Number of Relevant Items Selected / Total Number of Relevant Items) * 3))

The third and fourth alternative measures of relevancy are simply the count of number of different relevant items accessed and the number of irrelevant items accessed. This last measure is simply to assure we are not mistaking effects of relevancy for a simple count of the number of items accessed.

Neither Relevancy Index B (Figure 15) nor C (Figure 16) function very well as predictors of essay quality. A simple count of the number of different relevant items accessed (Figure 17), by contrast, seems to have a linear relationship to essay quality as measured by both the global score and essay scale score. The exception is a single user to who accessed only two relevant items but did very well. This user only made 10 steps altogether in the case. A comparison with a count of irrelevant items shown in Figure 18 suggests that this pattern is not simply a proxy for the relationship between the number of steps and essay quality.

Figure 15. Relevancy Index B by Mean Global and Essay Scale Score for Third Case (Spring 2003)

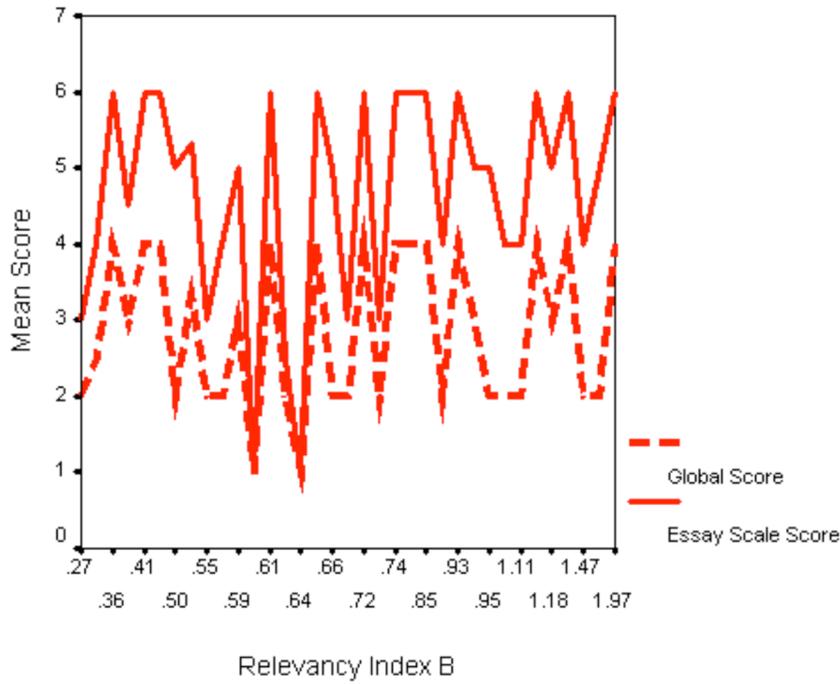


Figure 16. Relevancy Index C by Mean Global and Essay Scale Score for Third Case (Spring 2003)

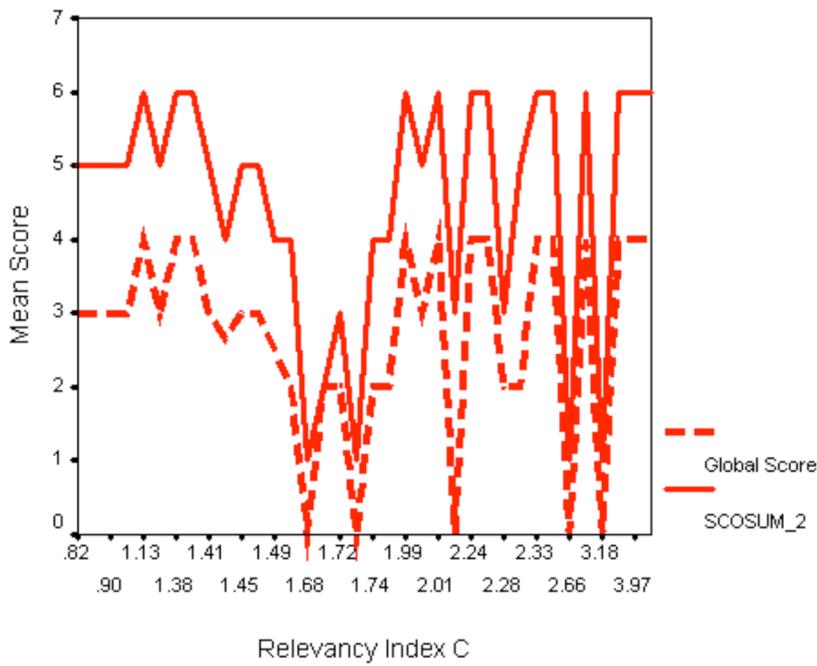


Figure 17. Number of Relevant Items Accessed by Mean Global and Essay Scale Score for Third Case (Spring 2003)

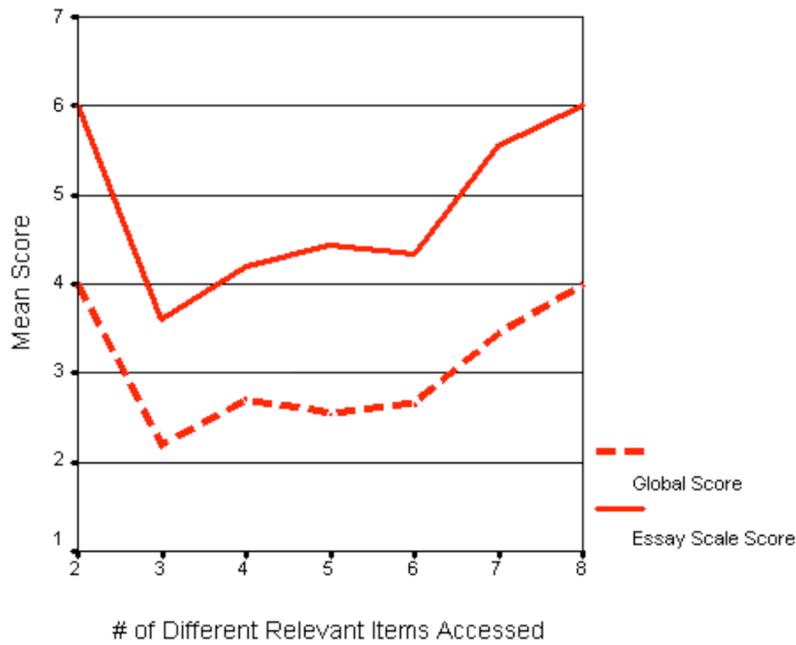
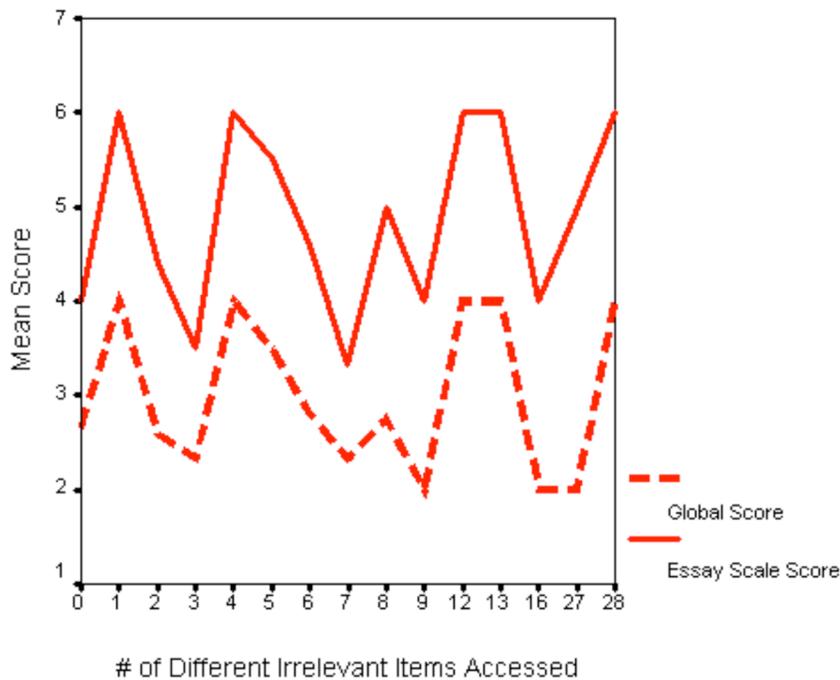


Figure 18. Number of Irrelevant Items Accessed by Mean Global and Essay Scale Score for Third Case (Spring 2003)



Discussion

The analyses in this paper were driven by the underlying question of whether students' access of relevant information in a case predicts their performance on essays following case searches. In fall 2002, where only student performance on the first case was examined, it did not. Three hypotheses were explored to account for this lack of a relationship: (1) the case search was unrelated to essay performance; (2) access to relevant information was related to essay performance but not when penalizing for the extent of the search; (3) the case search was related to essay performance but the relevancy as defined by the project staff was not. It was found for both semesters that students who completed a minimal exploration of the search did better on the essay than those who did not.

The relevancy index appeared to obscure this relationship by penalizing students who did extensive case searches. When this penalty was removed by looking at the relevancy total, however, there was a modest and positive relationship between access to relevant information and essay quality. This relationship only appeared for the second and third cases in spring 2003 however. Overall, these results lend support to the second hypothesis – relevancy, as assigned by the project staff, matters in later cases but trying to account for efficiency of search obscures this relationship. The strongest relationship between relevancy and essay quality was found when using a simple count of the number of different relevant items accessed as the measure of relevancy.

A preliminary two-stage model relating case search, relevancy, and essay quality can be formulated out of these results. The first stage involves becoming familiar with the task and problem space. Essay quality at this stage is dependent mainly on those students who choose to engage the task substantially versus those who do not. The second stage assumes a certain familiarity with the problem space and task. Students can now focus on making selection in accessing information that is relevant and not relevant to the task. These choices in turn begin to impact their essays. The first stage still applies – students who choose not to look at the case will still not produce good quality essay responses. But now students can make finer connections between the problem space and their responses.

The data employed here are quite limited in their potential to test the above model – particularly in the number of subjects who were scored on multiple cases. Moreover, they do not rule out an alternative version of the third hypothesis – that the relationship between relevancy and essay quality would be stronger given an alternative set of relevancy weights. Still two conclusions can be drawn from these results. First, there exists a certain minimal level of case exploration that is necessary for answering the case question well. Second, emphasis of efficient case searches (embedded in the relevancy index) obscures the relationship between access to relevant information and essay quality.

Appendix A: Example of Case with Relevant Items Highlighted

The following example illustrates how relevancy is applied in one of the ETIP cases. It is taken from a case with an urban, middle school called Cold Spring in which the instructor assigned questions pertaining to eTIP2 (“added value”). The case challenge reads as follows:

This case will help you practice your instructional decision making about technology integration. As you complete this case, keep in mind eTIP 2: technology provides added value to teaching and learning. Imagine that you are midway through your first year as a seventh grade teacher at Cold Springs Middle School, in an urban location. A responsibility of all teachers is to differentiate their lessons and instruction in order to accommodate for the varying learning styles, abilities, and needs of students in their classrooms and to foster students' critical and creative thinking skills. As a new teacher at Cold Springs Middle School, you will be observed periodically throughout the first few years of your career. One of the focuses of these observations is to analyze how well your instructional approaches are accommodating students' needs. The principal, Dr. Kranz, was pleased with your first observation. For your next observation she challenged you to consider how technology can add value to your ability to meet the diverse needs of your learners, in the context of both your curriculum and the school's overall improvement efforts. She will look for your technology integration efforts during your next observation.

On the case's answer page, you will be asked to address this challenge by making three responses:

1. Confirm the challenge: What is the central technology integration challenge in regard to student characteristics and needs present within your classroom?
2. Identify evidence to consider: What case information must be considered in a making a decision about using technology to meet your learners' diverse needs?
3. State your justified recommendation: What recommendation can you make for implementing a viable classroom option to address this challenge?

Examine the school web pages to find the information you need about both the context of the school and your classroom in order to address the challenge presented above. When you are ready to respond to the challenge, click "submit answer".

After reading the challenge, the user would then search for information relevant to the questions posed. The table below lists all the information categories and individual items in those categories available for searching in all cases. The information items relevant to this particular case (eTIP 2) are highlighted. Relevant information is in bold and semi-relevant information is in bold and italics. Note that this table serves as a key for examination of individuals in two selected classes presented later in the paper.

Table A. 1. Sample Problem Space with Relevant Information Highlighted

CATEGORY	INDIVIDUAL INFORMATION ITEMS
Prologue (1)	Prologue=1

About the School (2-11)	Mission Statement=2; School Improvement Plan=3; Facilities=4; School Map=5; Student Demographics=6; <i>Student Demographics Clipping=7; Performance=8;</i> Schedule=9; Student Leadership=10; Student Leadership Artifact=11
Staff (12-22)	Staff Demographics=12; Staff Demographics Talk=13; <i>Mentoring=14;</i> Staff Leadership=15; Staff Leadership; Talk=16; Faculty Schedule=17; Faculty Meetings=18; Faculty Talk=19; Faculty Meetings Artifact=20; Faculty Contract=21; Faculty Contract Talk=22
Curriculum and Assessment (23-30)	Standards=23; Instructional Sequence=24; <i>Computer Curriculum=25; Classroom Pedagogy and Assessment=26; Teachers=27; Talk=28; Talk 2=29; Clipping=30</i>
Technology Infrastructure (31-42)	School Wide Facilities=31; Library / Media Center=32; Classroom-Based Facilities=33; Classroom-Based Software Setup=34; <i>Community Facilities=35; Technology Support Staff=36;</i> Policies and Rules=37; Policies Clipping=38; Technology Committee=39; Technology Committee Talk=40; <i>Technology Survey Results=41;</i> Technology Plan and Budget=42
School Community Connections (43-48)	Family Involvement=43; Family Involvement Clipping=44; Business Involvement=45; Business Involvement; Clipping=46; Higher Education Involvement=47; Community Resources=48
Professional Development (49-68)	<i>Professional Development Content=49; Professional Development Content Area=50;</i> Resources=51; Professional Development Leadership=52; Professional Leadership=52; Professional Leadership Talk=53 Professional Development Talk=53; <i>Learning Community=54;</i> Learning Community Talk=55; Professional Development Process Goals=56; Professional Development Data=57; Professional Development Data; Artifact=58; Professional Development Evaluation=59; Professional Development Evaluation Talk=60; Professional Development Research=61; Professional Development Research Artifact=62; Professional Development Design=63; Professional Development Design Talk=64; Professional Development Learning=65 Professional Development Learning Artifact=66; Professional Development Collaboration=67; Professional Development Collaboration Artifact=68
Epilogue (69)	Epilogue=69
Essay (70)	Essay=70

Bold items have high relevance. Bold, italicized items have medium relevance.

The path a student took in the case (instructor J's class presented later in this paper) to search through the case is as follows:

- Step 1: Prologue
- Step 2: Student Demographics**
- Step 3: Faculty Contract
- Step 4: Faculty Schedule
- Step 5: Staff Leadership

- Step 6: Standards
- Step 7: Classroom Pedagogy and Assessment**
- Step 8: Teachers**
- Step 9: School Wide Facilities**
- Step 10: Library / Media Center**
- Step 11: Classroom-Based Facilities**
- Step 12: Classroom-Based Software Setup**
- Step 13: Policies and Rules
- Step 14: Technology Survey Results**
- Step 15: School Improvement Plan
- Step 16: (Clicks again on School Improvement Plan – Not Counted)
- Step 17: School Wide Facilities**
- Step 18: Library / Media Center**
- Step 19: Essay

The student had 18 actual steps in the sense that he or she made 18 separate steps within the case. The relevancy index is thus calculated by summing nine items with high relevancy (18 points) each plus one item with medium relevancy (1 point) divided by the total number of steps (18) which are subtracted by one so as not to count the step for writing the essay at the end). The formula is:

$$\begin{aligned} \text{Relevancy Index} &= (\text{Sum of Relevancy Points}) / (\text{Number of Actual Steps}) - 1 \\ 1.12 &= (19 / 17) \end{aligned}$$

Appendix B: Essay Score Rubrics

Table B.1. Summary of Rubric Score Criteria (Fall 2002)

Score	Criterion
1	<u>Validation</u> : Explains central challenge.
2	<u>Evidence</u> : Identifies factors in the case related to the challenge.
3	<u>Evidence</u> : Analyzes range of options for addressing challenge noting their advantages and disadvantages.
4	<u>Evidence</u> : States a decision or recommendation for implementing an option or change in response to the challenge.
5	<u>Decision</u> : Explains a justifiable rationale for the decision or recommendation.
6	<u>Decision</u> : Describes anticipated results of implementing the decision or recommendation.
7	Essay meets or does not meet expectations for all six decision making

criteria.

Table B.2. Summary of Rubric Score Criteria (Spring 2003)

Score	Criterion
1	<u>Validation</u> : Explains central challenge.
2	<u>Evidence</u> : Identifies case information that must be considered in meeting the challenge.
3	<u>Decision</u> : States a justified recommendation for implementing a response to the challenge.

Table B.3. Summary of Decision Rules of Global Score (Spring 2003)

Score	Decision Rule
0	Does not meet expectation because the decision criterion (score 3) equals 0 or rubric is blank.
1	Does not meet expectation because validation (score 1) and evidence (score 3) are both equal to 0.
2	Somewhat meets because other conditions above are not met.
3	Meets expectation because scores in both decision and validation, or decision and evidence equal 2.
4	Exemplary because scores for all three criteria (validation, evidence, decision) equal 2.