Measuring Student Variables Useful in the Study of Performance in an Online Learning Environment

Cathleen A. Kennedy
March, 2000

Much of the research into performance differences between distance and classroom students relies on analysis of observable explanatory variables such as student age, time spent on group activities, and amount of homework completed. The familiar statistical approach relies on analysis of variance and regression of these observed measures, assuming item independence and an equal error distribution across the observations. Frequently, the variables that are observed and manipulated are used simply because they are convenient and easy to measure.

This paper explores the measurement of unobservable, or *latent*, student characteristics that are likely to contribute to success in online courses and which can be used to investigate differences between student groupings. Since most current research into online learning focuses on student demographics, satisfaction, and achievement, new instruments are needed to measure student characteristics relevant to cognitive effects. Two instruments were developed to measure different aspects of student attitudes and behavior toward college and toward computers. This paper addresses construct validity of the two questionnaires, one measuring aspects of a student's college experience, the *College Experience Survey*, and the other measuring aspects of a student's experience with computers, the *Computer Experience Survey*.

The purpose of this paper is to determine whether the two instruments can be used to accurately discriminate five student variables that are expected to be associated with performance in online courses: the student's purpose in taking the course, the student's customary level of interaction with teachers, study habits, attitude about using computers, and experience with online computing technologies such as email and asynchronous conferencing.

Two models will be explored to gain insights to improve the design of the instruments. Since it is my hypothesis that each instrument measures multiple student variables, I will compare a unidimensional IRT analysis with a multidimensional analysis. These models will be used to determine (1) how well the instruments represent my hypothetical construct that they measure different things, and (2) that each variable measure provides an unbiased representation of linear magnitude on a specific dimension.

The focus of this study is to identify significant student variables that can be used to improve control of the comparison of online and classroom student learning and achievement. It is my conjecture that we will have greater validity when we control for more than just "maturity" as represented by age and gender. To this end, I developed student questionnaires to assess student characteristics that take into account the multidimensional nature of learning in an online environment. These instruments will then be used in experimental models to determine appropriate control groups for use in comparisons between online and classroom students, or subgroups of online students. The multidimensional random coefficients multinomial logit (MRCML) model can then be used to determine person positions in the multidimensional latent space, effectively generating student profiles on unobservable characteristics that can be used for comparison purposes.

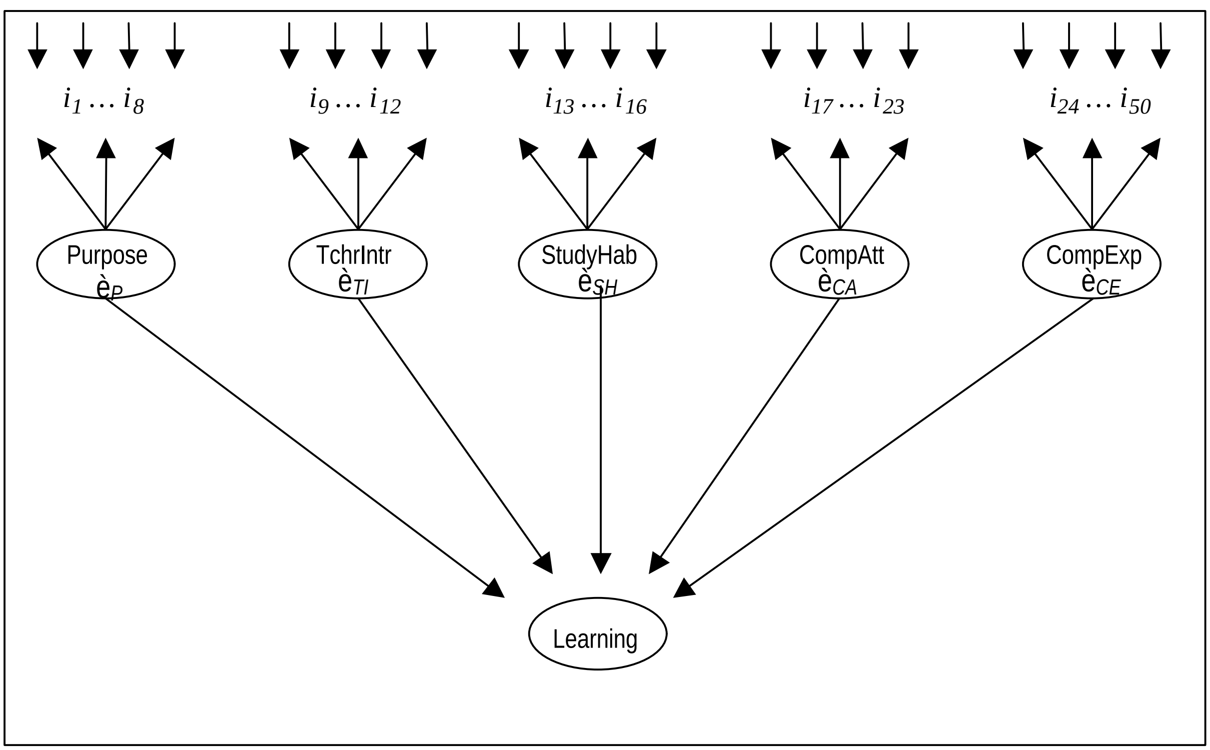
Given that this is an intermediate stage in a larger research project, I expect to fall short of actually establishing strong construct validity; instead, I hope to find interpretable indications of how to improve the two instruments. Ultimately, the instruments will be used to develop student profiles on the variables of interest. This paper begins with a description of the model and methods of data collection. The second section examines the results of the Item Response

Theory analysis. The paper concludes with a summary of findings and implications for further study.

Model

The diagram below (Figure 1) shows the theory underlying this study. The source variables under investigation are 50 items from the two instruments. The model assumes that the five student characteristics influence performance on the 50 items, as well as influence the learning outcome. An important part of the model is assumption of measurement error on each item.

Figure 1 – Student Variables that Influencing Learning



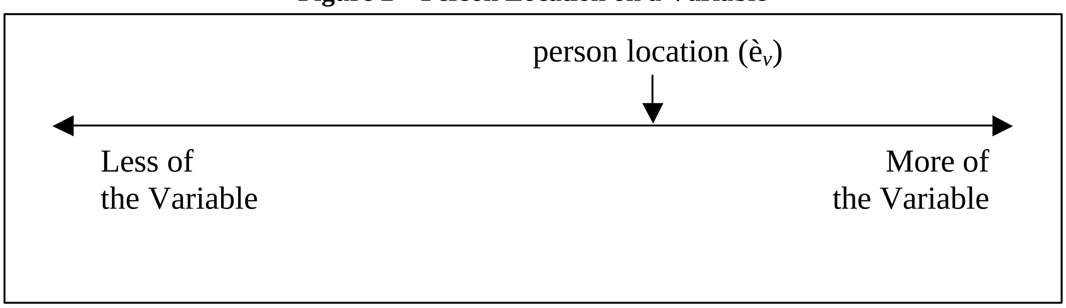
Student Variables

In developing the student variables, each must meet the following requirements for measurement:

1) The reduction of experience to a one dimension abstraction;
2) more or less comparisons among persons and items;
3) the idea of linear magnitude inherent in positioning objects along a line; and,
4) a unit determined by a process which can be repeated without modification over the range of the variable.

For each variable, then, a person will have a position along a continuum of the form shown in Figure 2 below:

Figure 2 – Person Location on a Variable



Purpose in Taking the Course (è$_{P}$)

A person with *more* of this variable has more reasons, or more important reasons, for taking the course, such as needing it for a major requirement, needing to take it because of its schedule, and interest in the subject. A person with *less* of this variable has fewer reasons for taking the course, and/or less important reasons. Measured through six dichotomous items, $i_1 \ldots i_6$.

Level of Interaction with Teachers (è$_{TI}$)

A person with *more* of this variable interacts with the teacher more frequently through asking questions in class and in the teacher's office, and even chatting with the teacher on topics unrelated to the course. A person with *less* of this variable has fewer interactions with the teacher. Measured through four polytomous items, $i_7 \ldots i_{10}$, each with four steps.

Study Habits (è$_{SH}$)

A person with *more* of this variable does more than the assigned work and may even meet with other students in ad hoc study sessions. A person with *less* of this variable may only do the required homework, or may do less homework than required. Measured through four polytomous items, $i_{11} \ldots i_{14}$, each with four steps.

Attitude About Using Computers (è$_{CA}$)

A person with *more* of this variable is comfortable with the idea of using computers, enjoys using computers, and may want to learn more about them. A person with *less* of this variable may feel awkward using computer or may specifically dislike the idea of using them. Measured through seven dichotomous items, $i_{15} \ldots i_{21}$.

Ability to Use Computers (è$_{CE}$)

A person with *more* of this variable uses the computer more frequently, for more types of uses and more complex tasks, and has a better understanding of when to use specific applications. A person with *less* of this variable uses the computer less frequently and for a limited number of uses. Measured through twenty-seven dichotomous items, $i_{22} \ldots i_{48,}$ and two polytomous items, $i_{49,}$ $i_{50.}$

Instrument Design and Coding

Two instruments were used to assess student characteristics and abilities:

1. ***College Experience Survey*** **–** Respondents with higher scores on this instrument have higher levels of engagement in their college experience, i.e. they report more reasons for enrolling in a course, they have higher levels of interaction with their teachers, and they have better study habits than those with lower scores on the instrument. The survey is designed to assess three variables:
    a. The reason the student enrolled in the course, or purpose;
    b. The types of interactions the student has had in the past with teachers; and
    c. The student's study habits.

    The *Purpose for taking the course* variable was assessed through a series of six dichotomous items such as "This is my major" and "I am interested in this subject;" and students could select multiple reasons. The other two variables, *Teacher Interaction* and *Study Habits*, were assessed through Likert-type items such as "I ask questions in class" and "I meet with other students outside of class to study" with choices of "never," "seldom," "sometimes," and "often."

2. ***Computer Experience Survey*** -- Respondents with higher scores on this instrument see themselves as more experienced computer users than those with lower scores. The survey is designed to assess two variables:
    a. Attitude about using computers; and
    b. Experience using online technologies like email and list servers.

    The *Computer Attitude* variable was assessed through a series of seven dichotomous items such as "I feel awkward using a computer" and "I would like to learn more about computers;” and students could select multiple reasons. The *Computer Experience* variable was assessed through a series of several dichotomous items plus five polytomous, partial credit-type items asking questions about best uses of specific technologies and how long the student had been using email and conferencing.

Both instruments are comprised of dichotomous and polytomous items. Coding student responses to some items proved problematic because the instruments were originally intended for regression on various response variables. The process of simply scoring “no” responses as 0 and “yes” responses as 1 lead to problems when items were to be aggregated either as a total test score or into subscales. On the *College Experience* questionnaire, students who received a higher total score should have more college experience, or a more mature response to college, than those who received lower total scores. For example, a student who is taking the course for his major, *and* because he wants to study under a particular teacher, *and* he likes the schedule would naturally score higher than a student who is *only* taking the course because he is interested in the subject, if all other responses are equivalent. Similarly, for the *Computer Experience* questionnaire, students who received a higher total score should have more computer experience than those who received lower scores. Unfortunately, when scoring the dichotomous items as 0 for “no” and 1 for “yes,” the desired ranking was not achieved. This is because the items were not designed to provide any kind of aggregate information. Examining a couple of items should make this clear:

Item Set 1 – 8:

Why are you taking this course (select all that apply):

__ This is my major

__ This course is required by my major

__ This course is an elective in my major

__ This is a General Education elective

__ I am interested in this subject

__ I wanted to take a course from this instructor

__ I like the schedule for this course

__ Other

Clearly, students taking a course because it is a major requirement will rank higher than other students on this item set because they will respond affirmatively to at least two items. This sort of problem was resolved by eliminating the second and third items from the set (if students selected the first, second, or third item, they were coded with selecting the first item).

Item Set 17 – 21:

Please describe your attitude about using computers (select all that apply):

__ I prefer not to use a computer

__ I don't mind using a computer

__ I feel awkward using a computer

__ I enjoy using a computer

__ I would like to learn more about computers

Selecting the first and third items from this set might indicate a negative attitude about using computers, but students making such a response would rate higher on the item set, and therefore on the instrument overall, than students who select only the fourth item. This sort of problem was resolved by reversing the scoring on “negative” items. Selecting the first item would get a score of 0 while not selecting it would get a score of 1, since that is a more positive response.

These modifications to the original scoring plan were expected to improve the overall validity of the instruments, as they allowed me to adjust the item scoring to more consistently represent *more* or *less* on the attitude variables. Since the items were originally intended to simply represent yes or no at the item level for regression analysis, the construct did not naturally lend itself to a more-or-less representation when multiple items are aggregated to create subscales.

Another complicating factor was incomplete responses, or indecipherable responses. Cases where the student could not be identified, and hence no final grade information could be obtained, were removed from the data set. Of the original 321 cases, 16 were removed for this reason. Other cases where information was indecipherable were coded as “missing”.

Data Collection Methods

Two methods of data collection were used. For classroom students, the two questionnaires were administered at the same time at the beginning of class sessions early in the semester. Students completed the questionnaire immediately and handed it in to the administrator or to the instructor. The purpose of the study was described to the students with varying degrees of detail and enthusiasm, depending on who administered the survey in a particular class. There is some evidence that completeness of the surveys was associated with whether or not the purpose of the study was described in detail to the students; most of the responses with missing data came from classes that received minimal instruction about the purpose of the study. Online students were invited to complete the questionnaires online. Each student received an email explaining the purpose of the study and a link to the questionnaires. The link went to an online CGI form equivalent to the first questionnaire. When students completed that form, they were to select the SUBMIT button on the screen. A response page would appear indicating that the form had been received and providing a link to the second questionnaire. A large number of online students only completed the first questionnaire and did not go on to the second link for the next one.

Analysis of differences in responses from online and classroom students must take test administration differences into account. Online students had a much lower response rate due to the voluntary nature of the request. Although classroom students could certainly “opt out” by not filling out the questionnaires, online students had to take the initiative to “opt in.” The result of this administration bias is that online responses represent mostly “good” students who are responsive to authority while classroom responses represent a wider mix of student attitudes. In the future, I will design data collection to be more comparable for the two groups. Although I would like to be able to give students the option of completing the forms online, this may significantly bias the results if many students choose not to participate.

Item Response Theory (IRT) Analysis

The multidimensional random coefficients multinomial logic (MRCML) model (Adams, Wilson & Wang, 1997) was used for this analysis because it is a generalized model that can be applied to a number of special cases. I used the ConQuest program (Wu, Adams & Wilson, 1998) for the computations. The MRCML model uses marginal maximum likelihood (MML) estimators to determine item difficulties and person abilities across one or more latent variables. MML is used to estimate parameters by assuming that persons have ability estimate vectors sampled from a population in which the distribution of abilities comes from the multivariate density function g(è; á), where á is a vector of parameters that characterize the distribution G(è; á) (for example, mean and standard deviation).

The MRCML assumes a set of D latent variables, with person-positions in the D-dimensional latent space represented by the vector è = (è$_1$, è$_2$, … , è$_D$). The scoring function b$_{ikd}$ represents a response in category *k* on item *i* in dimension *d*. The scoring matrix, B, is comprised of vector elements for a given item and response category across the range of dimensions. The item parameter vector, î, is usually comprised of item and step difficulties, and the design matrix, A, is comprised of one or more rows for each item, with a column for each item parameter.

The probability of a response in category *k* of item *i* is then

$$P(X_{ik} = 1; A, B, \text{î}|\text{è}) = \frac{\exp(b_{ik}\,\text{è} + a'_{ik}\,\text{î})}{\text{Ó}\exp(b_{ik}\,\text{è} + a'_{ik}\,\text{î})}$$

and the response vector is

$$f(x;\ \text{î}|\text{è}) = \text{Ø}(\text{è}, \text{î})\exp[x'(B\text{è} + A\text{î})], \text{ where } \text{Ø}(\text{è}, \text{î}) = \{\text{Ó}\exp[z'(B\text{è} + A\text{î})]\}^{-1}$$

The multivariate density function, g(è; á), is applied to obtain the marginal density of the response pattern x_j for person *j*. The likelihood equations of the item and population parameters are derived as cumulative products of the relevant response vectors.

College Experience Survey

The *College Experience Survey* was designed to assess three variables: (1) purpose in taking the course, (2) past interactions with teachers, and (3) study habits. Analysis began by comparing item ranking for the unidimensional and multidimensional models. Some expected patterns became apparent. For example, it was very easy for students to report that they do their homework regularly, but more difficult to report that they met outside of class with other students to study.

Figure 3 – Comparison of Unidimensional and Multidimensional IRT Estimates of College Experience Survey Items (in Logits)

Unidim		Multidim			
Item	CollExp		Purpose	TchrIntr	StudyHab
4 – Want Instructor	1.742	12 – Meet w/ students			1.279
5 – Like Schedule	1.735	13 – More HW			0.736
6 – Other Purpose	1.665	14 – Explore subject			0.678
10 – Chat w/ prof	0.302	4 – Want Instructor	0.670		
8 – Online dis	0.263	5 – Like Schedule	0.593		
2 – GE	0.247	6 – Other Purpose	0.574		
3 – Interest	0.048	10 – Chat w/ prof		0.551	
12 – Meet w/ students	0.031	8 – Online dis		0.439	
1 – Major	-0.710	2 – GE	-.409		
13 – More HW	-0.453	9 – Talk w/ prof on course		-.411	
14 – Explore subject	-0.497	7 – Ques in class		-.579	
9 – Talk w/ prof on course	-0.593	3 – Interest	-.621		
7 – Ques in class	-0.730	1 – Major	-.807		
11 – Regular HW	-3.620	11 – Regular HW			-2.692

In the unidimensional sense, we are looking at the variable *College Experience*. From this analysis, we can differentiate students who have a high degree of college experience from those with average or lower levels primarily from their responses about why they enrolled in the course (Figure 3). Students with higher levels of college experience are more likely to say that they enrolled in the course because they wanted to study under a particular instructor, because they wanted a particular course schedule, or they identified a personal reason for taking the course. On the other hand, students with lower levels of college experience were not likely to mention any reason other than that the course was in their major. In addition, students with lower than average levels of college experience were not as likely as other students to mention that they chat with their instructors about things other than the course, that they participate in online course discussions when they are available, that they are taking the course to meet a general education requirement or because they are interested in the subject, or that they meet with other students outside of class to study.

When we look at each dimension separately in the multidimensional analysis we find that the ranking of items from less difficult to agree with to more difficult to agree with are the same as in the unidimensional model, but that characteristics of high, average, and low levels of the characteristics are slightly different (Figure 3). For example, when looking at *Purpose* alone it's difficult to define an "average" respondent, but we would describe someone with a higher level of *Purpose* in the multidimensional analysis exactly as we would describe someone with a high level of *College Experience* in the unidimensional analysis. If we look at *Study Habits* alone, though, we can see a clearer pattern of high, average, and low levels. A person rated as "high" on study habits is more likely to report that they meet with other students outside of class to study, while a person rated as "low" only mentions that they do their homework regularly. "Average" students do their homework regularly and also are more likely to do more homework than required and to study the subject in more depth than required by the instructor.

Overall, we find no discrepancy in the ranking of items from most difficult to agree with to easiest to agree with in the unidimensional and multidimensional analysis of college experience. The items in the *Purpose* dimension are ranked in the same order as they appear in the unidimensional ranking (4, 5, 6, 2, 3, 1). The same is true for the *Teacher Interaction* (10, 8, 9, 7) and *Study Habits* (12, 13, 14, 11) dimensions. In both cases, the Monte Carlo integration method was used with a convergence criterion of .005 and 500 nodes. When we look at the goodness of fit of the two models, however, we find that making the model more complex by adding dimensions significantly reduces the fit of the model to the data (Figure 4). A possible explanation is that the model has local dependence problems. This possibility should be studied more closely.

Figure 4 – Comparison of Unidimensional and Multidimensional Goodness of Fit

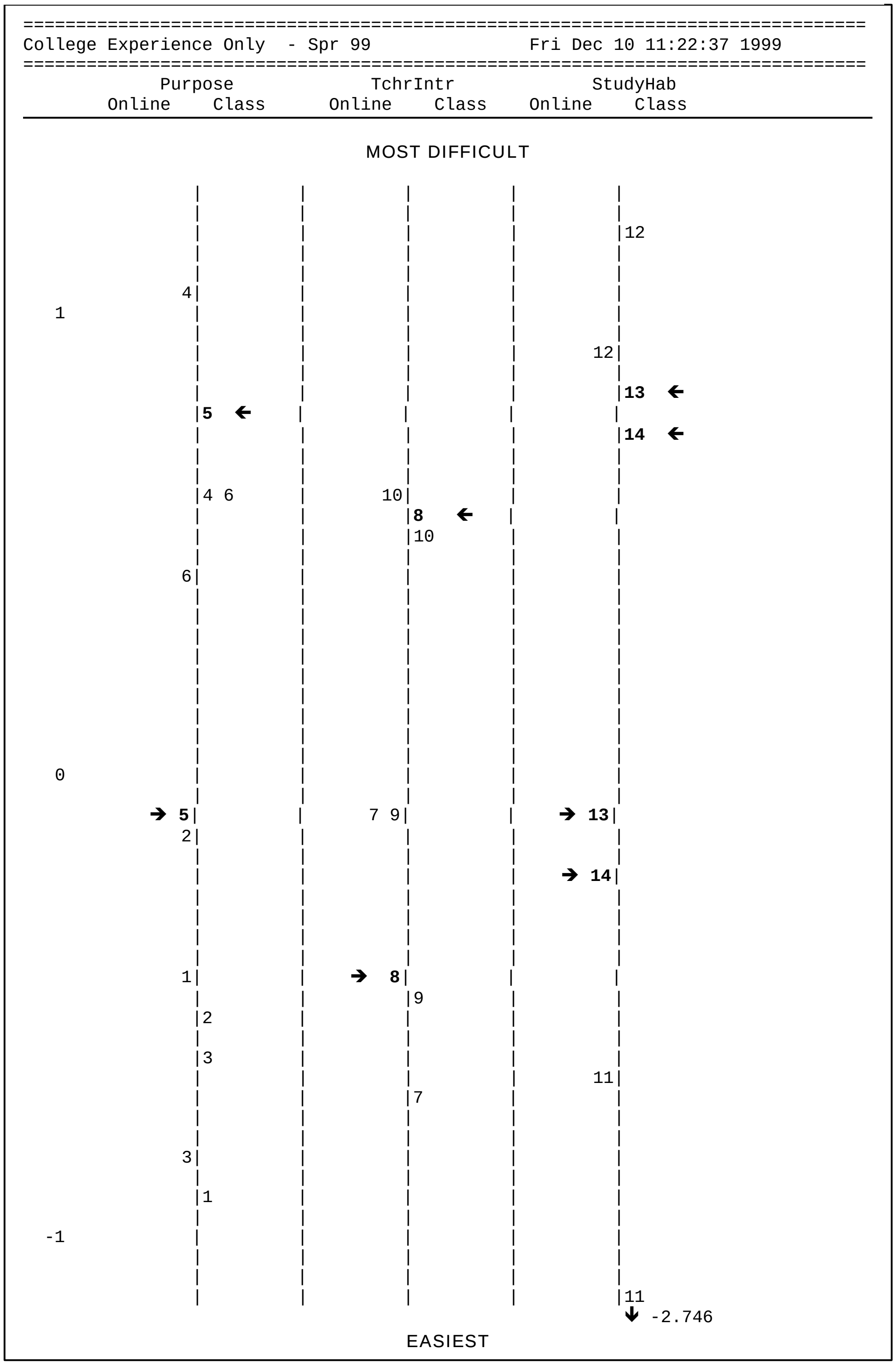
	Unidimensional	Multidimensional	Diff.
Deviance	7263.276	7350.934	87.658
Est. Parameters	30	35	5
÷2 for 5 df at .04			11.07

Figure 5 – Correlations of Dimensions

	1 Purpose	2 Tchr Inter	3 Study Habits
1 Purpose	1.000		
2 Teacher Interactions	-.229	1.000	
3 Study Habits	-.531	0.775	1.000

When we look at the correlations between the dimensions in the multidimensional analysis (Figure 5), we find modest negative correlations between *Purpose* and the other dimensions (r = -.23 and -.53 respectively), and a larger correlation between *Teacher Interactions* and *Study Habits* (r=0.78). This may suggest that students who are more purposeful in taking courses rely somewhat less on interacting with their teachers and much less on studying activities. Also, as study habits improve, so do interactions with teachers. Our intuition might suggest that more interaction with teachers helps students improve their study habits; that is, that teachers can help individual students improve their study habits. Conducting follow-up student interviews would be a good way to explore further interpretation of these findings.

Figure 6 – Comparison of Online and Classroom Item Difficulty Ranking



Next, I looked at item ranking from separate estimates of online and classroom students. Figure 6 provides a graphical representation of the differences. In the *Purpose* dimension, we see a distinct difference between the item rankings from online and classroom students. Online students found it most difficult to say that they enrolled in a course because they wanted to take a course from a specific instructor (item 4). Classroom students found this difficult, also, but it was not the most difficult item for them. Classroom students found it most difficult to say that they enrolled in a course because of the course schedule (item 5). It is not surprising to find that online students found this item fairly easy to agree with. This item difference is actually what I hoped the instrument would find, so it may be considered impact rather than bias. This will be discussed in more detail later in the paper.

In the *Teacher Interaction* dimension, one item (item 8) stands out as having a very different response for online and classroom students. Classroom students found it very difficult to agree with the statement "I participate in online discussions for class when available" while online students found it very easy. Again, this is what our theory suggests, and is more likely an item impact rather than an item bias.

Finally, in the *Study Habits* dimension, both groups found it most difficult to agree that they meet with other students outside of class to study (item 12) and easiest to agree that they do their homework regularly (item 11). The big difference in this dimension is that online students found it easy to agree that they explore the subject in more depth (item 14) and do more homework than is required (item 13) while classroom students found it more difficult to agree with these two statements. This finding may suggest that students who choose to enroll in online courses are already accustomed to doing "extra" work in their classes.

Next, I looked at how online and classroom students differed when item calibration was held constant for the multidimensional analysis. Item difficulties were generated in the multidimensional IRT for all students, then the difficulties were used as parameter anchors in separate runs for online and classroom students. The resulting expected a-posteriori person (EAP) estimates of ability on each dimension were then mapped on a logistic scale where persons and items use the same unit of measure.

Figures 7, 8 and 9 are plots of the student theta (ability) estimates on each dimension. As can be seen, online student estimates are quite different from classroom students. It is imperative that we understand why. Are the students truly different, or is this a manifestation of item bias or a valid impact difference? Terry Ackerman defines valid impact difference as “a between-group difference in test performance caused by a between-group difference on a valid skill” (1992). He goes on to explain than an instrument may have different dimensionality from one group of respondents to another and that this is an important construct validity issue. The challenge is to determine whether the observed differences in performance in the two groups is due to the expected difference we intended to measure or to some other extraneous, *nuisance* abilities.

These plots show that online students tend to have a higher level of purpose for taking the course, are accustomed to more interaction with their teachers, and have a higher and more uniform level of study habits than classroom students. Part of this difference can be explained by the test administration technique in which online students had to take the initiative to participate in the study. The possibility of Differential Item Functioning (DIF) causing the effect is discussed later in the paper.

Figure 7 – Plot of student estimates on Purpose dimension (EAP)

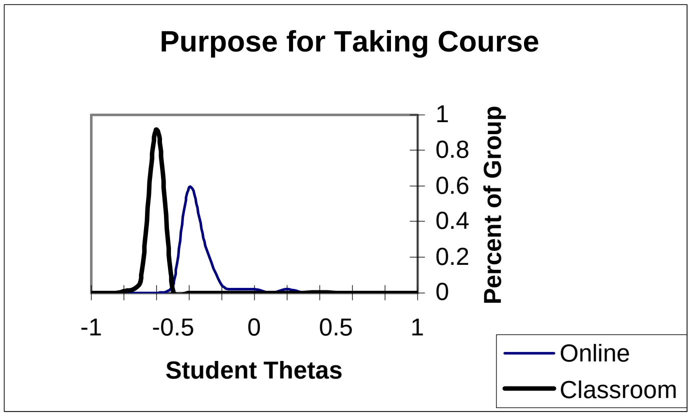


Figure 8 – Plot of student estimates on Past Teacher Interactions dimension (EAP)

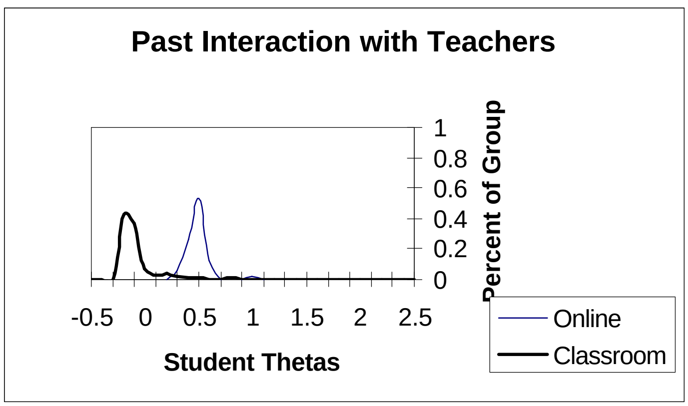
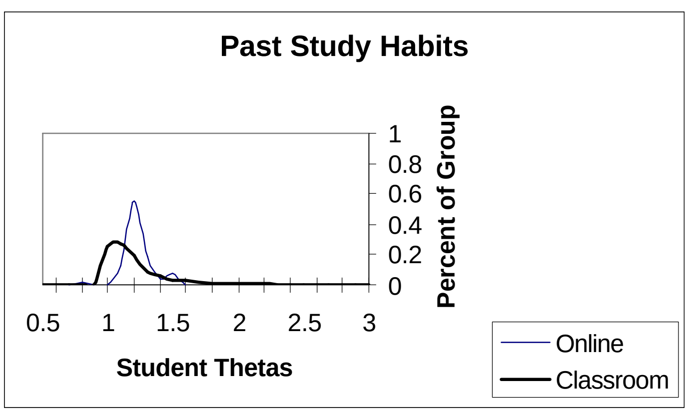


Figure 9 – Plot of student estimates on Past Study Habits dimension (EAP)



When developing instruments that are specifically intended to measure latent variables that may differentiate population subgroups, we need to make sure that the content of the items doesn't bias the findings or artificially create the subgroups we are hoping to see. Wilson investigated item parameter invariance using IRT on individual attitude scales (1994). Here, I use MRCML as a generalized extension of RCML (Adams, Wilson & Wang, 1997) to test for Differential Item Functioning (DIF) in the multidimensional case. I also compared mean ability estimates for the two groups with and without considering a course type (online vs. classroom) interaction (e.g. the interaction, or item DIF parameter, of interest is item*online):

model 1: item – online + item*online + online*item*step

model 2: item – online + item*step

Figure 10 shows the differences in item rankings for the two estimates. As can be seen, using the interaction does not produce significantly different results in the rank order of the items within each dimension, but it does produce important differences in the relative difficulty of the *Teacher Interaction* and *Study Habits* items. When the course type interaction is considered, item 8, "I participate in online discussions for the course when available," is much easier to agree with. In addition, when the interaction is not considered items 13 and 14, "I do more homework than required" and "I explore the subject in more depth than required" are almost equally difficult to agree with. When the interaction is considered, it becomes much <u>more difficult</u> for students to agree that they explore the subject in more depth than required and <u>much easier</u> to agree that they do more homework than required.

Figure 10 - DIF Comparison of Interaction with "Online format" in Multidimensional Analysis

```
================================================================================
College Experience Only  - Spr 99            Fri Dec 10 11:22:37 1999
================================================================================
             Purpose             TchrIntr             StudyHab
       No Inter.   Inter.   No Inter.  Inter.   No Inter.  Inter.
---------------------------------------------------------------------------------
                                                         12
                                 MOST DIFFICULT

                 |         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |       12|
                 |         |         |         |         |
                 |         |         |         |         |
   1             |         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |         |14  ←
                 |         |         |         |         |
                 |4        |         |         |         |
                 |         |         |         |         |
                4|         |         |10       |         |
                 |         |         |         |         |
                5|         |         |         |         |
                6|         |       10|         |         |
                 |6        |         |         |  →    13|
                 |         |         |         |  →    14|
                 |5        |         |         |         |
                 |         |  →     8|         |         |
                 |         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |         |
   0             |         |         |         |         |
                 |         |         |         |         |
                 |         |         |8  ←     |         |13  ←
                 |         |         |         |         |
                 |         |         |         |         |
                 |         |         |9        |         |
                 |         |         |         |         |
                 |         |         |         |         |
                 |2        |         |         |         |
                 |         |         |         |         |
                 |         |        9|7        |         |
                2|         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |         |
                 |         |        7|         |         |
                 |         |         |         |         |
                3|1        |         |         |         |
                 |3        |         |         |         |
                 |         |         |         |         |
                 |         |         |         |         |
                1|         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |         |
  -1             |         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |         |
                 |         |         |         |       11|11
                                                 -2.236    -2.343
                                     EASIEST
```

Figure 11 shows that the model with interaction is a significantly better fit than the model without interaction, with an improvement in the deviance of 108.5 for 25 degrees of freedom. The mean difference in performance of classroom and online students is 1.18 on the model with interaction and 0.988 on the model without interaction. These two findings suggest that the difference in the models is most pronounced in the step structure, while overall performance of classroom and online students is similar for the two models.

Figure 11 – Comparison of Models with and without Class Type (Online) Interaction

	No Interaction	With Interaction	Diff
Deviance	7188.4	7079.9	**108.5**
Parameters	36	61	**25**
÷2 for 25 df at .05			37.6
Mean difference	**.988**	**1.180**	

I then looked at how individual items fit the two models by examining the item fit parameter estimates, mean square and t-value (Figure 12). The mean square has an expected value of 1 and is considered within a reasonable range between .75 and 1.33; the t-value should fall within the range –2.0 to 2.0 for items that fit the model well. I found items 5, 12 and 13 to be outliers for the model that does not consider the course type interaction. The fit of item 5 was improved in the model with interaction, but items 12 and 13 continued to be problematic. The improved fit of item 5 is not surprising, as students who agree with item 5 were saying that they enrolled in the course because of its schedule. This is a more likely response from online students than from classroom students.

Item 12, "I meet with other students outside of class to study" has a high, positive t-value (2.9), suggesting that it is less discriminating than other items in the model. Since it's effect is slight, we can safely keep it in the model. Item 13, however, is more problematic. The extremely negative t-value of –3.1 suggests that the item is more discriminating than other items in the model and should probably be removed from the instrument.

Figure 12 – Comparison of Item Fit With and Without Class Type (Online) Interaction – Unweighted Fit

Item	No Interaction		Interaction	
	MnSq	t	MnSq	t
Purpose				
5	**0.74**	**-2.5**	**0.85**	**-1.3**
Study Hab				
12	**1.35**	**2.8**	**1.37**	**2.9**
13	**0.67**	**-3.1**	**0.67**	**-3.1**

If we focus on the results of the model that considers the course type interaction, we see that the mean difference in performance between online and classroom students was 1.18 (Figure 13). This means that online students performed 1.18 "higher" on the instrument than classroom students. Since the standard deviations for the three variables are 0.20, 0.65, and 1.91, this difference looks significant.

Figure 13 – Means and Standard Deviations with Class Type (Online) Interaction

	Mean
Classroom Students	-.590
Online Students	.590
Difference	1.180
	Std. Dev.
Purpose	0.20
Tchr Interaction	0.65
Study Habits	1.91

Figure 14 shows the amount that must be added to the difficulty of each item for classroom and online students to account for the interaction of the course type on performance on the item. These results indicate that classroom students found it relatively easier to agree with items 1, 2, 7, 9, 10, 12 and 14 than did online students, while online students found it relatively easier to agree with items 5, 6, 8, 11 and 13 than did classroom students. That is, classroom students were more likely than online students to report they were taking the course for their major or for a general education requirement; that they ask questions in class and meet with their instructors outside of class; and that they explore the course subject in more depth than required. Online students were more likely than classroom students to report that they enrolled in the course because of the schedule, or for some special reason; that they participate in online discussions about the course when they are available; and that they do more than is required on homework assignments. Only items 3 and 4 were relatively neutral; classroom and online students were equally likely to report that they enrolled in the course because they were interested in the subject matter or wanted to take a class from the instructor.

Figure 14 – Estimate Differential for Interaction between Course Type and Items

Item	Classroom	Online
1	-.327	.327
2	-.280	.280
3	.051	-.051
4	-.095	.095
5	.516	-.516
6	.135	-.135
7	-.219	.219
8	.637	-.637
9	-.267	.267
10	-.151	.151
11	.560	-.560
12	-.560	.560
13	.551	-.551
14	-.551	.551

There is clearly evidence of DIF in the *College Experience Survey;* some items function differently for online and classroom students. However, this difference is most likely an expected impact rather than a bias that favors one group over another.

A final analysis of fit was completed with item 13, "I do more homework than required," removed. This item appeared as an outlier in every analysis (Figures 6, 10, 12). Removal of item 13 resulted in a number of other items becoming extreme outliers. Items 4, 5, and 8 had mean squares below 0.75 and t-values below –2.0, suggesting that without item 13 these three items stop fitting the model well.

Figure 15 – Item 11: I do my homework regularly

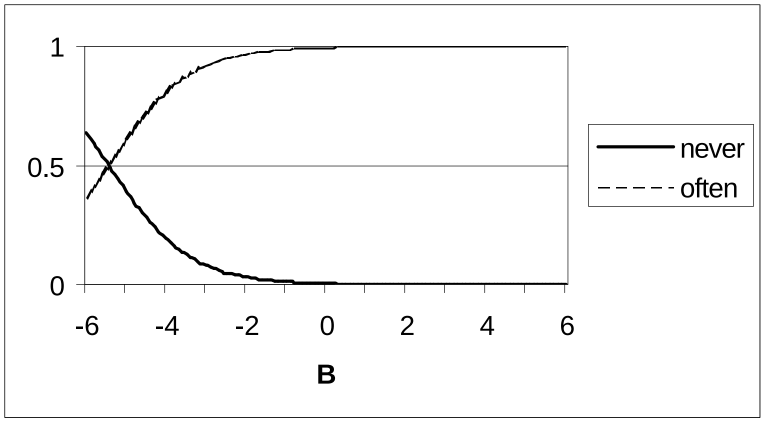


Figure 16 – Item 12: I meet with other students outside of class to study.

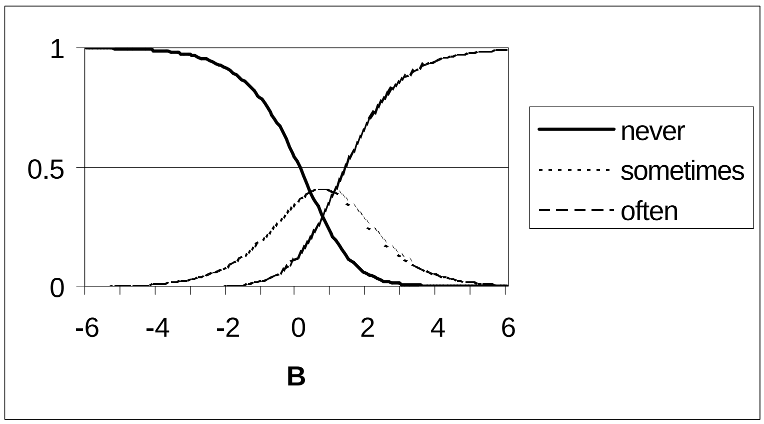
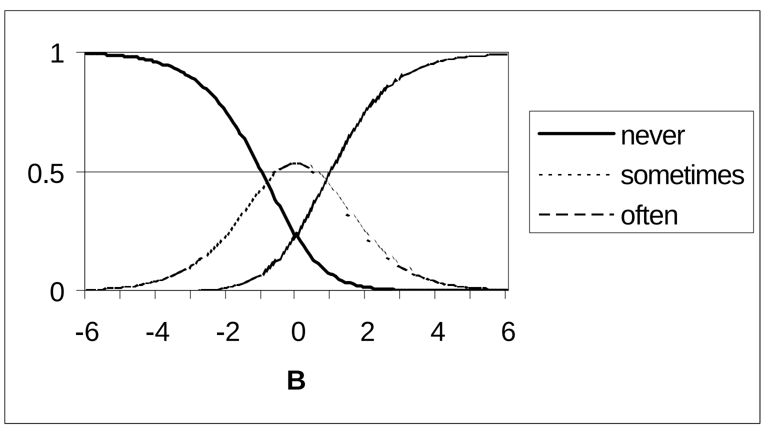


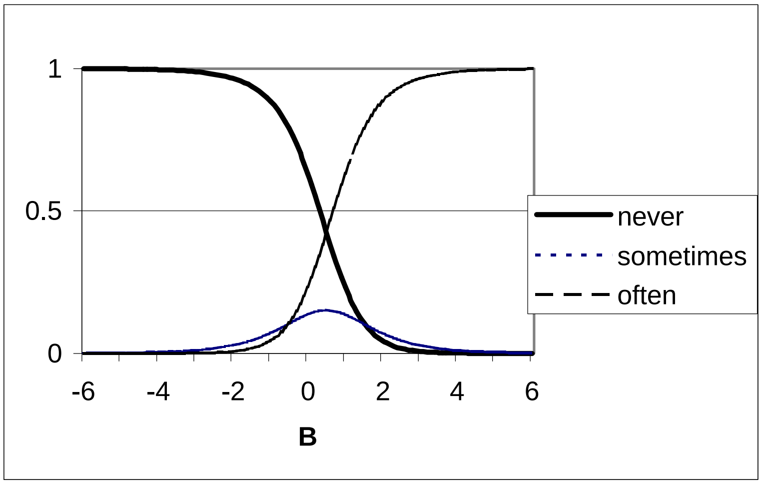
Figure 17 – Item 13: I do more than is required on homework.



If we look at category curves on the polytomous items, we find that some of the other items on the *Study Habits* subscale have more unusual response probabilities than item 13 (Figures 15, 16, 17). For example, item 11, “I do my homework regularly” acts more like a dichotomous item than a Likert-type item, and item 12, “I meet with other students outside of class to study” has a very small range for the middle response, “sometimes.” From this perspective, item 13 appears quite normal.

When we look at the category curves for the *Teacher Interaction* variables, item 8, “I participate in online discussions for the class when available,” is much more problematic (Figure 18). The category “sometimes” is never the most likely response, regardless of respondent’s ability level.

Figure 18 – Item 8: I participate in online discussions



Instead of simply removing item 13, a number of items should be rewritten, additional items should be added to each subscale, and the choices on the Likert-type items should be reconsidered.

Computer Experience Survey

The *Computer Experience Survey* was designed to assess two variables: (1) attitude about using computers; and (2) experience using online technologies like email and list servers. As with the *College Experience Survey*, analysis began by comparing item ranking for the unidimensional and multidimensional models (Figure 19). For the most part, items rank in about the same order in both models. Items in the *Attitude* dimension rank in exactly the same order on both instruments. Generally speaking, students found it easier to respond in negative ways about computer use and attitude; “I don’t use computers”, “I don’t like to use computers”, and “I feel awkward using computers” were less difficult items than “I am experienced with online conferencing” and “I value the results of using search engines.”

When we consider the instrument as a whole, measuring *Computer Experience* as a single variable, we find that students who rate higher on the variable are more likely than other students to use computers and be knowledgeable about selecting appropriate uses, such as conferencing instead of email or a list server instead of a conference. Students who rate lower on the variable are still likely to have an email account and use computers at home, but tend to mention that they are fairly new users, use the computer less than once a week, and feel awkward using the computer or actually don’t like to use it. If we consider the *Attitude* and *Experience* dimensions separately, we have similar characterization of students who rank high or low on the variables. Average respondents mentioned that they enjoy using computers and email and want to learn more; students with a higher attitude ranking mentioned that they enjoy conferencing, while lower ranking students mentioned that they don’t like computers or feel awkward using them. As we would expect, students with a lower ranking on *Experience* tended to report that they use the computer infrequently, usually in only one place, like at home or at work, are fairly new computer users, but do have an email account. Students who ranked high on the variable tend to use the computer from a variety of places, and use list servers and online conferences more than those who ranked at an average level of experience.

Figure 19 – Comparison of Unidimensional and Multidimensional IRT Ranking of Instrument 2 Items

Unidim		Multidim		
Item	CompExp	Item	Attitude	Exper
26 – conf exper	4.124	26 – conf exper		4.090
25 – conf wkly	3.080	25 – conf wkly		2.699
18 – use in other places	2.948	18 – use in other places		2.641
13 – other exper	2.459	7 – enjoy conferencing	2.560	
30 – lstsrv vs. conf.	2.195	13 – other exper		2.266
29 – value srch eng	2.075	29 – value srch eng		1.955
27 – conf vs email	1.818	2 – don't mind computers	1.691	
7 – enjoy conferencing	1.850	17 – use in library		1.352
17 – use in library	1.709	30 – lstsrv vs. conf.		1.173
34 – diff in srch engines	1.459	28 – subscribe to lstsrv		1.140
28 – subscribe to lstsrv	1.392	36 – yrs exper conf		0.914
2 – don't mind computers	1.245	27 – conf vs email		0.877
36 – yrs exper conf	0.822	34 – diff in srch engines		0.856
16 – use at school	0.633	33 – experienced srcher		0.519
33 – experienced srcher	0.610	24 – part in confs		0.179
24 – part in confs	0.533	16 – use at school		0.168
32 – use srch wkly	0.237	6 – enjoy using email	0.011	
12 – exper w/ comp	0.121			
11 – exper w/ web	0.117	32 – use srch wkly		-.019
		4 – enjoy computers	-.035	
23 – exper emailer	-.002	11 – exper w/ web		-.255
22 – daily email	-.019	5 – want to learn comp	-.267	
10 – use email	-.023	12 – exper w/ comp		-.274
9 – use some apps	-.068	10 – use email		-.399
31 – know how to srch	-.398	9 – use some apps		-.434
15 – use at work	-.387	23 – exper emailer		-.460
6 – enjoy using email	-.426	22 – daily email		-.493
4 – enjoy computers	-.462	31 – know how to srch		-.827
35 – yrs exper email	-.701	15 – use at work		-1.035
5 – want to learn comp	-.744	35 – yrs exper email		-1.196
21 – have email acct	-1.259	3 – feel awk w/ comp	-1.780	
8 – new to comp	-1.737	21 – have email acct		-1.979
14 – use at home	-1.771	1 – don't like comp	-2.181	
19 – use less once wk	-3.028	8 – new to comp		-2.269
3 – feel awk w/ comp	-3.228	14 – use at home		-2.692
20 – never use comp	-3.755	19 – use less once wk		-3.893
1 – don't like comp	-11.421	20 – never use comp		-4.604

When we compare the overall fit results for the unidimensional and multidimensional models (Figure 20), we find that the multidimensional model is a much better fit, with an improvement in the deviance of 428 for only 2 degrees of freedom. In addition, the correlation between the two dimensions is –0.372, which is surprising. We would expect a higher correlation and in a positive direction. Instead, the modest negative value suggests that students with more experience using computers may not enjoy the experience. It could be that they feel compelled to use computers, or that using computers is not necessarily a choice. This issue should be explored through individual interviews or focus groups to provide more information about this association.

Figure 20 – Comparison of Unidimensional and Multidimensional Goodness of Fit - College Experience Survey

	Unidimensional	Multidimensional	Diff.
Deviance	10308.253	9879.262	428.991
Est. Parameters	45	47	2
÷2 for 5 df at .05			5.99

As with the *College Experience Survey*, I used the multidimensional model to test for DIF in the *Computer Experience Survey*. The Conquest results are shown in Figure 21 below. These findings suggest significant DIF in the items. First, the model with interaction has a much higher difference in group means than the invariant model, and second, the model with interaction is much better fit than the invariant model (improvement in deviance is 1057 for 42 degrees of freedom). Another compelling finding is that over two thirds of the items (27 of 36) in the invariant model have small mean square values and highly negative t-values, suggesting that the items are more discriminating than the rest of the items. On the model with interaction, however, every item has a large mean square (above 1.33) and a very large positive t-value.

Figure 21 – Comparison of Multidimensional Model with and without Class Type (Online) Interaction – Computer Experience Survey

	No Interaction	With Interaction	Diff
Classroom Mean	-0.277	-1.491	
Online Mean	0.277	1.491	
Difference	0.554	2.982	
Std. Dev. on Attitude	.788	1.05	
Std. Dev. on Experience	.784	2.63	
Deviance	10844.5	9786.7	1057.8
Parameters	48	89	41
÷2 for 40 df at .05			55.76

The DIF analysis also showed some extreme cases of differential item locations between groups. For example, item 1, "I prefer not to use a computer," had a differential estimate of 10.0 between the groups, with classroom students finding this much easier to agree with. Item 14, "I use my computer at least once a week from home," had a differential estimate of 4.0, with online students finding this much easier to agree with, and item 19, "I use a computer less than once a week," had a differential estimate of 6.8, with classroom students finding this item much easier to agree with. Since two of these items are posed in negative terms, we might find that simply restating them in positive terms affects the way respondents answer the question.

I then looked at how online and classroom students differed when item calibration was held constant for the multidimensional analysis. I did not remove any items since so many were problematic. At this point, I wanted to learn as much as possible about the variables and the items so they could be improved. Item difficulties were generated in the multidimensional IRT for all students, then the difficulties were used as parameter anchors to determine individual student thetas in the two latent spaces, *Computer Attitude* and *Computer Experience*. When we look at the student estimates for *Computer Attitude* in Figure 22 we find that online and classroom students are not much different from one another. Figure 23 suggests that, in general, a larger percentage of online students have high levels of computer experience than classroom students and a smaller percentage have low levels of computer experience. These findings match our intuition that students who choose to take courses online are more experienced using online computer technologies than students who do not choose to take online courses.

Figure 22 – Plot of student estimates on Computer Attitude dimension (EAP)

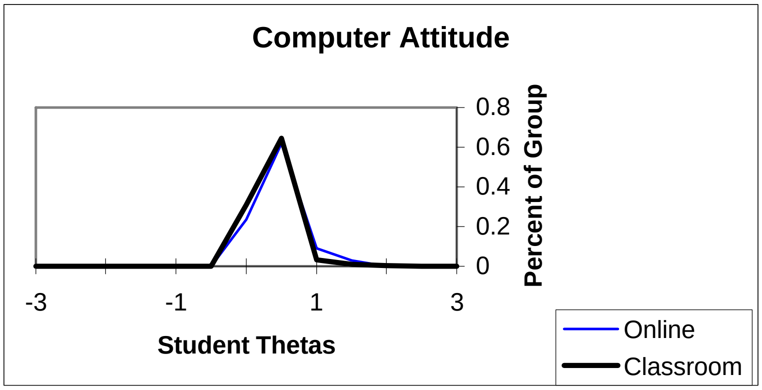
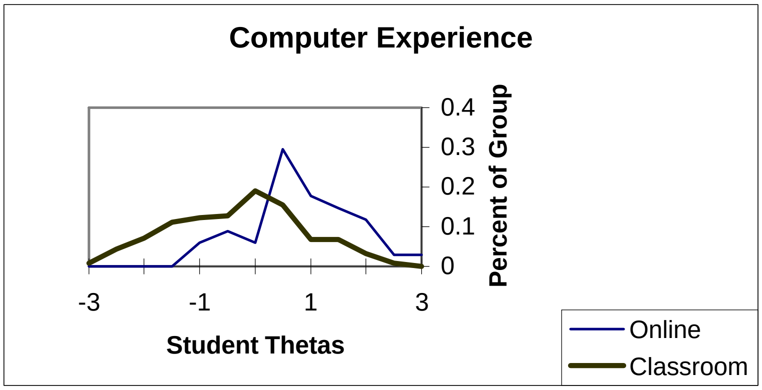


Figure 23 – Plot of student estimates on Computer Experience dimension (EAP)



On the whole, the analysis of the *Computer Experience Survey*, particularly the DIF results, emphasize the need for a thorough review of the items. Student interviews and focus groups, and additional pilot testing of a modified instrument are clearly indicated.

Conclusions

The findings from the analyses conducted thus far support the theory that multiple dimensions can be identified through the two instruments. Several items in the instruments need adjustment, however. On the *College Experience Survey*, items need to be modified to so that one purpose, such as taking the course because it's in the major, is not weighted more heavily than another, and scoring needs to be adjusted so that higher scores always mean a "more positive" attitude or "more of" a variable. In addition, the Likert-type response categories should be changed to include an "always" option, and the "seldom" option should be removed, since students seemed to use "seldom" and "somewhat" to mean the same thing. Finally, more items should be added to each subscale to improve the fit of individual items to the instrument as a whole. This should help ensure that items like 5, 8, and 13 won't be as overly discriminating as they are now.

The *Computer Experience Survey* needs additional analysis, possibly including student interviews, to determine a more valid construct. The extreme DIF could be addressed by reconsidering the structure and content of many of the items. Although we found that online and classroom student attitude EAP estimates were similar, and that online students tend to have more computer experience than classroom students, we are concerned with some of the extreme item location differences for the two groups. We note, though, that this situation may be improved by restating some of the items as positive instead of negative statements.

Administration bias also needs to be considered when the questionnaires are next administered. It will be essential that a more diverse sample of online students respond to the survey, and not just students who are willing to "help out" the researcher. This is problematic whenever respondents need to "opt in" rather than "opt out" of participation. One approach is to make the survey a normal part of online course evaluation instruments, where there is some expectation that all students are to participate.

In addition, focus groups of students should be conducted to better understand how students are interpreting the questions. Difficulty ranking of *College Experience Survey* items 5 (wanted course schedule), 8 (do online discussions) , 13 (do more homework), and 14 (explore subject in depth) were extremely different for the two groups. On the *Computer Experience Survey*, items 1 (prefer not to use computers), 14 (use my computer at least once a week from home), and 19 (use computer less than once a week) are of particular interest.

Once the items are modified to more closely parallel the theoretical model, the *College Experience* instrument should provide higher scores for those with more diverse and successful learning strategies, without bias toward one student group over another. The *Computer Experience* instrument should more accurately measure both attitude and experience so that students with a more positive attitude about using computers and more experience using online technologies will score higher.

Once these two instruments are redesigned , they will be reanalyzed to confirm that I have improved their construct validity. Then, the instruments can be used in further research that compares student groups, with better control for initial differences that I suspect affect student performance in online courses.

Bibliography

Ackerman, T. A. (1992, Spring) A Didactic Explanation of Item Bias, Item Impact, and Item Validity From a Multidimensional Perspective. *Journal of Educational Measurement*, *29, 1,* 67-91.

Adams, R. J., Wilson, M., Wang, W. (1997, March) The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement, 2, 1,* 1-23.

Joreskog, K. & Sorbom, D. (1999) *LISREL 8.30*, Scientific Software International.

Moore, S. (1996) Estimating DIF with the RCML model. In Engelhard, G. & Wilson, M. (Eds.) *Objective Measurement: Theory into Practice, vol 3*. Norwood,NJ:Ablex.

Wilson, M. (1994) Comparing Attitude Across Different Cultures: Two Quantitative Approaches to Construct Validity. In Wilson, M. (Ed.) *Objective Measurement: Theory into Practice, vol 2*. Norwood,NJ:Ablex.

Wu, M., Adams, R. & Wilson, M. (1998) *ACER ConQuest*. Australian Council for EducationalResearch.

Appendix

Demographics

		Classroom N=271	**Online N=49**	**Total N=320**
Age	17 or under	3.0%	4.1%	3.2%
	18-22	40.7%	22.4%	37.9%
	23-25	14.2%	8.2%	13.2%
	26-35	23.9%	32.7%	25.2%
	36-45	11.6%	26.5%	13.9%
	46 or over	6.6%	6.1%	6.6%
		100%	100%	100%
Sex	Male	57.8%	46.0%	55.9%
	Female	42.2%	54.0%	44.1%
		100%	100%	100%
Empl. Status	Full time	46.9%	61.2%	49.1%
	Part time	38.7%	22.4%	36.3%
	Not employed	14.4%	16.4%	14.6%
		100%	100%	100%

Instrument 1 Recoded Items

College Experience Survey

Item Number		
1	Why are you taking this course? Select all that apply: *Any checked box counts as a "yes" answer to item 1.*	❑ This is my major ❑ This course is required by my major ❑ This course is an elective in my major
2	Why are you taking this course? Select all that apply:	❑ This is a General Education elective
3	Why are you taking this course? Select all that apply:	❑ I am interested in this subject
4	Why are you taking this course? Select all that apply:	❑ I wanted to take a course from this instructor
5	Why are you taking this course? Select all that apply:	❑ I like the schedule for this course
6	Why are you taking this course? Select all that apply:	❑ Other ___________
7	Describe your past interactions with teachers: I ask questions in class.	Often – Sometimes – Seldom - Never
8	Describe your past interactions with teachers: I participate in online discussions for class when available.	Often – Sometimes – Seldom - Never
9	Describe your past interactions with teachers: I ask questions outside of class.	Often – Sometimes – Seldom - Never
10	Describe your past interactions with teachers: I communicate (talk, email, etc.) with my teacher about things not related to the specific course I'm taking with him or her.	Often – Sometimes – Seldom - Never
11	Describe your past study habits: I do my homework regularly.	Often – Sometimes – Seldom - Never
12	Describe your past study habits: I meet with other students outside of class to study.	Often – Sometimes – Seldom - Never
13	Describe your past study habits: I do more than what is required on homework assignments.	Often – Sometimes – Seldom - Never
14	Describe your past study habits: I explore the subject are in more depth than required by the teacher (reading, online study, talk to other teachers, etc.)	Often – Sometimes – Seldom - Never

Instrument 2 Recoded Items

Computer Experience Survey

Item Number		
1	Please describe your attitude about using computers (select all that apply):	❑ I prefer not to use a computer (reverse score)
2	Please describe your attitude about using computers (select all that apply):	❑ I don't mind using a computer
3	Please describe your attitude about using computers (select all that apply):	❑ I feel awkward using a computer (reverse score)
4	Please describe your attitude about using computers (select all that apply):	❑ I enjoy using a computer
5	Please describe your attitude about using computers (select all that apply):	❑ I would like to learn more about computers
6	Please select all that apply regarding email:	❑ I enjoy communicating with email
7	Please select all that apply regarding online conferencing (does not include chatting in real time):	❑ I enjoy online conferencing
8	Please describe your experience using computers (select all that apply):	❑ Beginner
9	Please describe your experience using computers (select all that apply):	❑ Regular use of one or two applications
10	Please describe your experience using computers (select all that apply):	❑ Regular use of email
11	Please describe your experience using computers (select all that apply):	❑ Regular web access
12	Please describe your experience using computers (select all that apply):	❑ Experienced computer user
13	Please describe your experience using computers (select all that apply):	❑ Other
14	Check all of the places where you use a computer at least once a week:	❑ Home
15	Check all of the places where you use a computer at least once a week:	❑ Work
16	Check all of the places where you use a computer at least once a week:	❑ School
17	Check all of the places where you use a computer at least once a week:	❑ Public Library
18	Check all of the places where you use a computer at least once a week:	❑ Other place
19	Check all of the places where you use a computer at least once a week:	❑ I use a computer less than once a week (reverse score)
20	Check all of the places where you use a computer at least once a week:	❑ I never use a computer (reverse score)
21	Please select all that apply regarding email:	❑ I have an email account
22	Please select all that apply regarding email:	❑ I access my email at least once a day
23	Please select all that apply regarding email:	❑ I am an experienced email user (can set up an account, forward, add attachments, etc.)
24	Please select all that apply regarding online conferencing (does not include chatting in real time):	❑ I have participated in an online conference

25	Please select all that apply regarding online conferencing (does not include chatting in real time):	❑ I participate in online conferences at least once a week
26	Please select all that apply regarding online conferencing (does not include chatting in real time):	❑ I am an experienced conference participant (can set up a new topic, add attachments, etc.)
27	When would it be better to use online conferencing instead of email?	incorrect answer = 0 correct answer = 1 complete answer = 2
28	Please select all that apply regarding list servers:	❑ I subscribe to at least one list server
29	Please select all that apply regarding list servers:	❑ The information I get from the list server is valuable to me
30	Briefly describe the difference between using a list server and a conference. What is the difference in their purpose?	incorrect answer = 0 correct answer = 1 complete answer = 2
31	Please select all that apply regarding search engines:	❑ I know how to use at least one search engine
32	Please select all that apply regarding search engines:	❑ I use a search engine at least once a week
33	Please select all that apply regarding search engines:	❑ I can make complex searches with multiple search criteria
34	What is the difference between using a search engine like Yahoo! and one like InfoSeek or MetaCrawler?	incorrect answer = 0 correct answer = 1 complete answer = 2
35	How long have you been using email?	No answer = 0 Less than a month = 1 Less than a year = 2 1 – 2 years = 3 More than 2 years = 4