Chapter 6

STATISTICAL ANALYSIS

Most of the data in this study were analyzed using AMOS© 3.61, a program distributed by SmallWaters Corporation for solving structural equations with latent variables. AMOS© is a relatively new graphical SEM analysis tool that can fit multiple models in a single analysis by constraining parameters within the models.

Given that the essence of this dissertation involves the analysis of task complexity as a moderating variable in the Social Cognitive model for computer training and that this moderator is introduced only for the behavior modeling and lecture-based instruction levels of training, the following sections describe tests performed on data in only these two groups. Analysis of the CAI training method is presented separately, as an empirical analysis, at the end of this chapter.

6.1 AMOS© ASSUMPTIONS

Inherent in AMOS's programming are the following assumptions:

1. For any value pattern of the fixed variables, the remaining (random) variables have a (conditional) normal distribution.
2. The (conditional) variance-covariance matrix of the random variables is the same for every pattern of the fixed variables.
3. The (conditional) expected values of the random variables depend linearly on the values of the fixed variables. (Arbuckle, 1997)

One caveat of using AMOS© is that in merely meeting the traditional requirements of normality and independence of observations, only asymptotic conclusions can be reached. That is, actual parameter values are valid for only large samples. In this study, no inferences are intended for parameter values. Instead, of primary importance are the significance levels of path coefficients; therefore, minor deviations of these assumptions are not expected to produce erroneous conclusions.

In addition to AMOS's inherent assumptions, the traditional assumptions of causal modeling are expected. That is, the latent exogenous variables are assumed to be uncorrelated with each other and the zetas (i.e. the variance in the endogenous latent variables that is unexplained by the variance in the causal linkages from the exogenous latent variables) also are uncorrelated with each other. These assumptions are enforced in the path analysis by constraining the covariances between these items with a parameter value of zero, indicating no correlation between these components.

6.2 TEST OF NORMALITY ASSUMPTION

Using AMOS©, multivariate normality is determined by using the "$normalitycheck" command. This command produces statistics on skewness and kurtosis. Table 6.1 presents the results of this analysis.

As shown in Table 6.1, skewness and kurtosis do not appear to be significant problems in the data set. Using the benchmark ± 2.0, no items exhibited significant skewness, and only two items on the Outcome Expectations questionnaire (#2 and #4) and one item on the Self-efficacy instrument demonstrated kurtosis.

It should be noted that deviations from multivariate normality may or may not affect the results of analysis. Arbuckle (1997) advocates, "A departure from normality that is big enough to be significant could still be small enough to be harmless." It is argued here that because no inferences about actual population parameters are intended, and rather, only the significance of relationships between variables in the model are important, these minor departures may pose no threat to the conclusions reached in this study. Further, given that the questionnaire items that exhibit kurtosis were used in previous studies that used structural equation modeling (Compeau and Higgins, 1995, 1995b), and that these studies were concerned with only the significance levels of the path coefficients, it is assumed that these departures from normality can be justifiably disregarded.

Table 6.1 Tests of Normality: Skewness and Kurtosis

Assessment of Multivariate Normality			
Skewness			**Kurtosis**
Indicator			**Indicator**
Training	-0.069	-1.995	**Training**
Self-Reported Expertise	1.268	-0.02	**Self-Reported Expertise**
QCA	-0.096	-0.04	**QCA**
Prior Performance Index	-0.093	-0.206	**Prior Performance Index**
Final Performance	-0.679	-0.759	**Final Performance**
Outcome Expectations #6	-0.837	0.727	**Outcome Expectations #6**
Outcome Expectations #5	-1.192	1.714	**Outcome Expectations #5**
Outcome Expectations #4	-1.49	2.661	**Outcome Expectations #4**
Outcome Expectations #3	-0.756	0.638	**Outcome Expectations #3**
Outcome Expectations #1	-0.889	1.117	**Outcome Expectations #1**
Outcome Expectations #2	-1.511	3.234	**Outcome Expectations #2**
Self-Efficacy #10	-1.245	1.477	**Self-Efficacy #10**
Self-Efficacy #9	-1.374	2.437	**Self-Efficacy #9**
Self-Efficacy #8	0.075	-0.931	**Self-Efficacy #8**
Self-Efficacy #7	-0.493	-0.466	**Self-Efficacy #7**
Self-Efficacy #6	-0.386	0.236	**Self-Efficacy #6**
Self-Efficacy #5	-0.689	0.204	**Self-Efficacy #5**
Self-Efficacy #4	-0.11	-0.602	**Self-Efficacy #4**
Self-Efficacy #3	-0.17	-0.443	**Self-Efficacy #3**
Self-Efficacy #2	0.529	-0.584	**Self-Efficacy #2**
Self-Efficacy #1	0.214	-0.629	**Self-Efficacy #1**

Assuming a multivariate normal distribution, shaded items in the table above exhibit significant skewness and/or kurtosis using a benchmark of +/- 2.0.

6.3 MISSING DATA

The computer program used in the experiment automatically recorded responses from the questionnaires; therefore, no missing observations resulting from typing errors were possible. Instead, the major source of missing data resulted from subjects who did not elect to use Solver. For these subjects, no final performance score could be attributed; therefore, the dependent measure in the analysis would have been missing. Although AMOS© provides excellent analysis tools for circumventing missing data, omission of this single dependent measure would have caused an error in the analysis. Thus, observations for which subjects did not use Solver to perform their final task were omitted from the analysis.

Because the Visual Basic macro used to record responses to questionnaires prevented users from continuing to the next screen until all responses were completed, no missing data were attributable to incomplete questionnaires. Thus, a total of sixty-two observations were removed from the analysis for failure to use Solver.

6.4 OUTLIERS

Outliers were determined as part of the analysis (Hair et al., 1995). Table 6.2 presents the results of AMOS's test of outliers using the Mahalanobis distance statistic. This statistic represents the squared distance from the centroid of a data set. As part of the "$normalitycheck" command, AMOS© produces a listing of the top one hundred observations, ranked in order of their Mahalanobis distances. Additionally, AMOS© presents two additional statistics, p1 and p2. The p1 column shows the probability of any observation exceeding the squared Mahalanobis distance of that observation. The p2 column shows the probability that the largest squared distance of any observation would exceed the Mahalanobis distance computed. A heuristic for determining which observations may be outliers is given by Arbuckle (1997): "Small numbers in the p1 column are to be expected. Small numbers in the p2 column, on the other hand, indicate observations that are improbably far from the centroid under the hypothesis of normality."

To determine which, if any, observations were outliers in the original data set, all observations listed on Table 6.2 with p2 values less than .1 were individually examined. This analysis revealed thirty-two observations with two or more “0” responses to the questions on the self-efficacy or two or more “1” responses on the outcome expectations instrument. A “0” response on the self-efficacy questionnaire indicated that the person did not feel that he/she could perform a certain computer task in the situation described. These self-efficacy scores, though low, were still valid responses and, thus, did not warrant removal from the data set. Likewise, low scores on the outcome expectations questionnaire were equally acceptable for the same reason. Eleven participants answered with unusually high, albeit valid, responses to the self-efficacy questions. Five observations were examined as possible outliers as they included perfect final performance scores. Conversely, seven participants were initially considered as outliers as their final performance scores were zero. In total, fifty-five observations were initially identified as possible outliers, but upon closer inspection, proved to be valid data points and, therefore, were retained in the data set.

Table 6.2 Analysis of Outliers

Observations farthest from the centroid (Mahalanobis distance)

Observation number	Mahalanobis d-squared	p1	p2
76	63.47	0	0.000
37	62.813	0	0.000
237	59.238	0	0.000
234	53.857	0	0.000
126	48.84	0.001	0.000
242	45.068	0.002	0.000
215	44.738	0.002	0.000
9	44.525	0.002	0.000
95	44.134	0.002	0.000
104	43.504	0.003	0.000
41	42.227	0.004	0.000
36	39.758	0.008	0.000
190	39.642	0.008	0.000
219	39.615	0.008	0.000
101	38.858	0.01	0.000
39	38.069	0.013	0.000
186	37.852	0.013	0.000
61	37.518	0.015	0.000
14	35.921	0.022	0.000
174	35.579	0.024	0.000
157	35.159	0.027	0.000
255	34.821	0.03	0.000
98	34.779	0.03	0.000
229	33.632	0.04	0.000
254	33.208	0.044	0.000
259	33.048	0.046	0.000
152	30.76	0.078	0.000
4	30.526	0.082	0.000
70	30.429	0.084	0.000
42	64.559	0	0.001
2	32.115	0.057	0.001
114	32.064	0.058	0.001
121	31.686	0.063	0.001
177	31.637	0.064	0.001
136	31.299	0.069	0.001
108	31.297	0.069	0.001
154	30.929	0.075	0.001
139	30.849	0.076	0.001
130	30.788	0.077	0.001
257	30.047	0.091	0.001
204	29.719	0.098	0.002
60	28.73	0.121	0.032
197	25.839	0.213	0.037
44	28.201	0.135	0.048
245	25.844	0.212	0.050
118	25.923	0.209	0.052
233	25.994	0.207	0.056
203	28.261	0.133	0.058
124	25.615	0.222	0.058
109	28.372	0.13	0.060
23	27.846	0.145	0.065
45	26.027	0.205	0.067
127	25.436	0.229	0.076
94	27.86	0.144	0.085
208	26.042	0.205	0.085
194	26.056	0.204	0.106
183	25.17	0.24	0.127
46	24.999	0.247	0.127
171	24.906	0.251	0.130
261	26.169	0.2	0.131
17	26.066	0.204	0.133
201	26.225	0.198	0.145
189	26.304	0.195	0.152
176	25.006	0.247	0.156
29	26.898	0.174	0.163
112	24.72	0.26	0.169
48	27.309	0.161	0.175
83	26.334	0.194	0.178
260	26.959	0.172	0.180
30	26.44	0.19	0.219
227	26.339	0.194	0.219
67	26.962	0.172	0.224
97	27.066	0.169	0.227
149	26.525	0.187	0.227
213	26.568	0.186	0.256
56	24.36	0.276	0.312
222	24.227	0.282	0.344
43	23.924	0.297	0.387
22	23.999	0.293	0.389
258	23.845	0.301	0.389
230	23.766	0.305	0.390
132	24.028	0.292	0.423
40	23.537	0.316	0.496
115	23.404	0.323	0.537
28	23.265	0.33	0.583
224	23.101	0.339	0.594
50	22.95	0.347	0.599
199	23.007	0.344	0.609
244	22.562	0.368	0.623
191	23.122	0.338	0.630
32	22.254	0.385	0.646
86	22.732	0.358	0.651
231	22.591	0.366	0.651
20	22.649	0.363	0.660
225	22.776	0.356	0.669
195	22.263	0.385	0.687
91	22.093	0.394	0.709
96	22.264	0.384	0.730
12	22.279	0.384	0.762
16	21.897	0.405	0.787

6.5 SAMPLE SIZE

Hair, et al. (1995) recommend a sample size of at least one hundred observations to achieve adequate power in structural equation modeling:

> Maximum likelihood estimation (MLE) has been found to provide valid results with sample sizes as small as 50, but a sample this small is not recommended. It is generally accepted that the minimum sample size to ensure appropriate use of MLE is 100. As we increase the sample size above this value, the MLE method increases in its sensitivity to detect differences among the data. As the sample size becomes large (exceeding 400 to 500), the method becomes "too sensitive" and almost any difference is detected, making all goodness of fit measures indicate poor fit.

Two hundred forty-five valid observations remained in the data set for behavior modeling and lecture based training after all eliminations were made due to subjects failing to use Solver in the final task. Because AMOS© uses the Maximum likelihood estimation method in its analyses, this sample size should prove sufficient for obtaining adequate power based on the Hair, et al. recommendations.

6.6 MEASUREMENT MODEL FIT

Tests of measurement model fit first were conducted to ensure the reliability and validity of the multiple indicators used for the two endogenous variables, computer self-efficacy and outcome expectations.

Reliability refers to the internal consistency exhibited by the manifest indicators of each construct in Structural Equation Modeling. Validity, on the other hand, reflects the degree to which these indicators represent the constructs they purport to measure. Although intricately related, reliability is a necessary, but not sufficient, condition for validity (Hair, et al., 1995). Conversely, the reliability of manifest indicators is founded on the assumption of unidimensionality, a concept related to construct validity. Due to this dependency, Hair, et al. (1995) suggest testing data for unidimensionality before performing tests of reliability. Thus, a discussion of validity is presented below, followed by an analysis of reliability.

6.6.1 Tests of Validity

Confirmatory factor analysis is a technique frequently used to determine the construct validity, or unidimensionality, of indicators in Structural Equation Modeling. Its purpose is to assess the degree to which manifest indicators represent the constructs they are intended to measure.

6.6.1.1 Factor analysis

Factor analysis using a varimax rotation was performed for the computer self-efficacy and outcome expectations constructs. Determination of the number of factors to include was based on the percentage of variance method. Using this method, the researcher ascertains the number of factors to include by examining the variance accounted for by each factor in the unrotated factor loadings. Hair, et al. (1995) state that in the natural sciences, researchers include factors until the total variance explained equals ninety-five percent or until the incremental contribution of the last factor is less than five percent. In the social sciences, however, a much lower percentage is typically acceptable (often as low as sixty percent).

In the current analysis, two factors were used to determine the loadings on the self-efficacy and outcome expectations constructs. The inclusion of a third factor contributed only an additional eight percent of explained variance to the model. Evidence of this variance explained is presented in the unrotated factor loadings in Table 6.3.

Table 6.3 Unrotated Factor Loadings for Percentage of Variance Method

Principal Component Factor Analysis of the Correlation Matrix
Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Factor4	Commnlty
Outcome Expectations #1	-0.257	-0.536	-0.202	-0.647	0.812
Outcome Expectations #2	-0.362	-0.75	0.046	-0.169	0.724
Outcome Expectations #3	-0.309	-0.689	0.124	0.131	0.603
Outcome Expectations #4	-0.319	-0.779	0.114	-0.032	0.723
Outcome Expectations #5	-0.418	-0.684	0.055	0.182	0.679
Outcome Expectations #6	-0.355	-0.55	0.033	0.345	0.549
Self-Efficacy #1	-0.686	0.18	-0.383	-0.119	0.664
Self-Efficacy #2	-0.623	0.217	-0.407	-0.033	0.602
Self-Efficacy #3	-0.632	0.095	-0.369	0.122	0.56
Self-Efficacy #4	-0.681	0.223	-0.186	0.059	0.551
Self-Efficacy #5	-0.658	0.15	-0.26	0.021	0.524
Self-Efficacy #6	-0.712	0.253	0.154	-0.1	0.605
Self-Efficacy #7	-0.723	0.285	0.158	-0.012	0.629
Self-Efficacy #8	-0.693	0.075	0.108	0.326	0.603
Self-Efficacy #9	-0.589	0.3	0.591	-0.164	0.813
Self-Efficacy #10	-0.584	0.289	0.557	-0.135	0.753
Variance	5.0484	3.1876	1.3461	0.8128	10.3949
% Var	0.316	0.199	0.084	0.051	0.65

Table 6.4 presents the results of the factor analysis and reveals that the six outcome expectations questions loaded almost exclusively on one factor, and the ten self-efficacy questions loaded predominantly on the other factor. Given these results, it was concluded that the items on each questionnaire adequately represented the constructs to which they were associated.

Table 6.4 Factor Analysis

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Commnlty
Outcome Expectations #1	0.034	0.593	0.353
Outcome Expectations #2	0.049	0.831	0.694
Outcome Expectations #3	0.023	0.755	0.571
Outcome Expectations #4	-0.002	0.842	0.709
Outcome Expectations #5	0.126	0.792	0.643
Outcome Expectations #6	0.119	0.644	0.429
Self-Efficacy #1	0.703	0.095	0.503
Self-Efficacy #2	0.659	0.037	0.435
Self-Efficacy #3	0.621	0.153	0.409
Self-Efficacy #4	0.714	0.053	0.513
Self-Efficacy #5	0.665	0.112	0.456
Self-Efficacy #6	0.755	0.038	0.571
Self-Efficacy #7	0.777	0.012	0.604
Self-Efficacy #8	0.669	0.194	0.485
Self-Efficacy #9	0.659	-0.053	0.437
Self-Efficacy #10	0.65	-0.044	0.425
Variance	4.7785	3.4576	8.2361
% Var	0.299	0.216	0.515

6.6.2 Tests of Reliability

Three of the most common statistics for assessing reliability in Structural Equation Modeling are Cronbach's alpha, construct reliability, and variance extracted. These measures are presented in the sections that follow.

6.6.2.1 Cronbach's (Coefficient) alpha

As discussed previously, alpha coefficients were computed for the outcome expectations and computer self-efficacy questionnaires. The typical benchmark for assessing this coefficient is .70 (Hair, et al., 1995). Results indicated that the reliabilities of all manifest indicators were adequate. As shown in Table 6.5 Cronbach's alpha was 0.88 and 0.84 for computer self-efficacy and outcome expectations, respectively.

Table 6.5 Computation of Reliability (Cronbach's Alpha)

Cronbach's Alpha

Formula: $\frac{k}{k-1} \times 1 - \frac{EF^2}{EF^2 + 2(EF)}$

Outcome Expectations

	OE1	**OE2**	**OE3**	**OE4**	**OE5**	**OE6**	**Total**	**Composite**
Sum	972.00	1065.00	941.00	1074.00	1019.00	967.00	6038.00	6038.00
Mean	3.72	4.08	3.61	4.11	3.90	3.70	23.13	23.13
Std. Deviation	0.94	0.94	1.05	1.00	1.05	1.10	6.07	4.54
Variance	0.88	0.89	1.09	0.99	1.11	1.20	6.17	20.62
n=261								

***a* = 6/(6-1) * (1 - (6.17/(20.62))) = 0.84**

Computer Self-Efficacy

	SE1	**SE2**	**SE3**	**SE4**	**SE5**	**SE6**	**SE**	**SE8**	**SE9**	**SE10**	**Totals**
Sum	840.00	619.00	1260.00	1380.00	1847.00	1535.00	1630.00	1126.00	2042.00	1951.00	14230.00
Mean	3.22	2.37	4.83	5.29	7.08	5.88	6.25	4.31	7.82	7.48	54.52
Std. Deviation	2.44	2.20	2.39	2.24	2.23	2.00	2.70	2.85	2.08	2.31	23.44
Variance	5.95	4.83	5.71	5.04	4.98	4.00	7.29	8.11	4.31	5.36	55.58
n=261											

***a* = 10/(10-1) * (1 - (55.58/(262.64))) = 0.88**

6.6.2.2 Construct Reliability

To fairly assess the reliability of the manifest indicators in the study, additional methods of determining measurement model fit were used. The first of these tests, Construct Reliability, is presented below:

Construct Reliability = (Σstandardized loadings)^2 / (Σstandardized loadings)^2 + Σε_j

where the standardized loadings are the factor loadings resulting from the factor analysis and ε_j is the measurement error for each indicator. Hair, et al. (1995) suggest that these reliabilities should exceed .50, "which roughly corresponds to a standardized loading of .70." Table 6.6 presents the measurement error matrix used in the computation of construct reliability and variance extracted (discussed below).

Table 6.6 Measurement Error Matrix from Factor Analysis

Measurement Error Matrix

	OE1	OE2	OE3	OE4	OE5	OE6	SE1	SE2	SE3	SE4	SE5	SE6	SE7	SE8	SE9	SE10
OE1	0.647															
OE2	-0.026	0.306														
OE3	-0.140	-0.097	0.429													
OE4	-0.095	0.006	-0.053	0.291												
OE5	-0.139	-0.088	-0.052	-0.067	0.357											
OE6	-0.102	-0.127	-0.081	-0.138	-0.030	0.571										
SE1	0.093	-0.015	-0.034	-0.049	-0.035	0.008	0.497									
SE2	0.045	0.001	0.031	-0.063	-0.015	-0.018	0.152	0.565								
SE3	0.016	-0.044	-0.064	-0.029	0.000	-0.001	0.045	0.007	0.591							
SE4	-0.032	0.036	-0.028	0.011	0.029	-0.061	-0.054	0.004	-0.006	0.487						
SE5	0.024	-0.019	-0.064	0.017	-0.008	-0.035	-0.026	-0.064	0.076	0.041	0.544					
SE6	0.021	0.023	0.031	0.019	-0.039	-0.012	-0.086	-0.111	-0.139	-0.060	-0.065	0.429				
SE7	-0.002	-0.001	0.063	0.025	-0.022	-0.019	-0.089	-0.058	-0.135	-0.086	-0.116	0.038	0.396			
SE8	-0.094	-0.040	0.004	0.006	-0.020	0.042	-0.096	-0.088	-0.057	-0.081	-0.079	-0.029	0.021	0.515		
SE9	-0.034	0.060	0.049	0.061	0.016	-0.011	-0.194	-0.238	-0.147	-0.072	-0.140	0.015	-0.026	-0.058	0.563	
SE10	-0.024	0.016	0.022	0.041	0.057	0.037	-0.144	-0.1497	-0.155882	-0.1826834	-0.135456	-0.051465	-0.009521	-0.04071	0.310	0.575

OE=Outcome Expectations
SE=Self-Efficacy

As indicated in Table 6.7, Construct Reliability for the self-efficacy questionnaire was 0.90 . For the outcome expectations instrument, it was 0.88. Both of these values, greatly exceeding the .50 benchmark, indicate sufficiently high reliability for the indicators of these constructs.

Table 6.7 Computation of Construct Reliability

Construct Reliability

Formula:

$$\frac{(\text{E standardized loadings})^2}{(\text{E standardized loadings})^2 + \text{E (measurement error)}}$$

Sum of Standardized loadings: (from Table 6.4)

	SE		OE
SE:	0.703	**OE:**	0.593
	0.659		0.831
	0.621		0.755
	0.714		0.842
	0.665		0.792
	0.755		0.644
	0.777		4.457
	0.669		
	0.659		
	0.65		
	6.872		

Sum of Measurement Error (from Table 6.6)

	SE		OE
SE:	0.497183	**OE:**	0.646636
	0.564549		0.306461
	0.591408		0.42943
	0.486991		0.291187
	0.544495		0.356816
	0.428625		0.571341
	0.395839		2.601871
	0.514587		
	0.563152		
	0.575236		
	5.162065		

Reliability Computation:

SE: 0.901462 **OE: 0.88419**

6.6.2.3 Variance Extracted

Another test of reliability, variance extracted, “reflects the overall amount of variance in the indicators accounted for by the latent construct.” (Hair et al., 1995, p. 642). The formula for this measure is presented below:

Variance Extracted = Σ(standardized loadings)^2 / Σ(standardized loadings)^2 + $\Sigma\varepsilon_j$

where the standardized loadings are the factor loadings resulting from factor analysis, and ε_j is the measurement error for each indicator. Hair, et al., (1995) suggest that this measure should exceed .50 for a construct.

As shown in Table 6.8, the Variance Extracted formulas indicate adequate reliability with estimates of 0.48 and 0.56 for computer self-efficacy and outcome expectations, respectively. Although the Variance Extracted statistic for the computer self-efficacy instrument falls slightly short of the .50 benchmark, the other tests presented above provide enough evidence to suggest that this questionnaire exhibits adequate reliability.

6.6.3 Correlations

An additional test to verify the unidimensionality of each construct’s manifest indicators is an examination of the correlations among questionnaire items. As illustrated in Table 6.9, this common sense approach suggests less unidimensionality than indicated by the tests of reliability and validity presented above. For example, question #1 on the Outcome Expectations instrument exhibits relatively low correlations with the other items on the questionnaire. Similarly, questions 8, 9, and 10 on the Self-efficacy instrument demonstrate weak correlations with questions 1,2,3, and 5 and relatively high correlations with items 6 and 7 on the same questionnaire. In short, examining the correlations between indicators suggests a degree of multi-dimensionality that should be considered when interpreting the results of statistical analyses involving these instruments.

Table 6.8 Computation of Variance Extracted

Variance Extracted

Formula:

$$\frac{E\ (\text{standardized loadings})^2}{E\ (\text{standardized loadings})^2 + E\ (\text{measurement error})}$$

Sum of Squared Standardized loadings: (from Table 6.4)

		Squared			**Squared**
SE:	0.70	0.49	**OE:**	0.59	0.35
	0.66	0.43		0.83	0.69
	0.62	0.39		0.76	0.57
	0.71	0.51		0.84	0.71
	0.67	0.44		0.79	0.63
	0.76	0.57		0.64	0.41
	0.78	0.60			3.36
	0.67	0.45			
	0.66	0.43			
	0.65	0.42			
		4.74			

Sum of Measurement Error (from Table 6.6)

SE:	0.50	**OE:**	0.65
	0.56		0.31
	0.59		0.43
	0.49		0.29
	0.54		0.36
	0.43		0.57
	0.40		2.60
	0.51		
	0.56		
	0.58		
	5.16		

Variance Extracted Computation:

SE: 0.48 **OE: 0.56**

Table 6.9 Correlation Matrix

	PPIndex	QCA	Expertis	O1	O2	O3	O4	O5	O6	SE1	SE2	SE3	SE4	SE5	SE6	SE7	SE8	SE9	SE10
QCA	-0.006																		
Expertis	0.36	-0.052																	
O1	0.137	0.032	0.192																
O2	0.174	0	0.158	0.48															
O3	0.145	0.071	0.144	0.332	0.538														
O4	0.111	0.046	0.135	0.388	0.705	0.549													
O5	0.192	-0.019	0.185	0.322	0.537	0.545	0.567												
O6	0.124	0.017	0.114	0.291	0.403	0.42	0.398	0.492											
SE1	0.291	-0.147	0.284	0.235	0.1	0.072	0.032	0.121	0.106										
SE2	0.195	-0.044	0.133	0.112	0.044	0.063	-0.056	0.058	0.06	0.58									
SE3	0.259	-0.037	0.227	0.169	0.139	0.071	0.103	0.18	0.127	0.52	0.418								
SE4	0.203	-0.112	0.138	0.055	0.123	0.031	0.049	0.151	0.068	0.448	0.469	0.455							
SE5	0.23	-0.149	0.157	0.139	0.103	0.026	0.094	0.155	0.083	0.456	0.379	0.506	0.538						
SE6	0.223	-0.137	0.127	0.133	0.127	0.088	0.068	0.063	0.076	0.47	0.441	0.368	0.501	0.471					
SE7	0.17	-0.057	0.143	0.057	0.065	0.088	0.024	0.059	0.068	0.447	0.481	0.379	0.475	0.419	0.657				
SE8	0.221	-0.154	0.235	0.109	0.181	0.192	0.155	0.214	0.229	0.385	0.385	0.397	0.406	0.374	0.509	0.549			
SE9	0.177	-0.018	0.115	-0.024	0.072	0.044	0.023	0.06	0.051	0.245	0.213	0.259	0.418	0.306	0.516	0.497	0.376		
SE10	0.198	-0.058	0.142	-0.02	0.041	0.028	0.02	0.103	0.101	0.291	0.271	0.243	0.3	0.313	0.441	0.496	0.393	0.728	
Percent	0.244	0.11	0.07	0.046	0.072	0.076	0.072	0.055	-0.032	0.069	0.026	0.086	0.061	0.108	0.112	0.09	0.074	0.163	0.152

6.7 MANIPULATION CHECK

As stated in Chapter 5, a final question was posed to all subjects requesting the degree of difficulty (on a scale of 1 to 10) that they attributed to the task they performed. To determine the strength of the manipulations, a one-way analysis of variance was performed. This ANOVA revealed that the subjects who received the high complexity task felt that the task was significantly more difficult than those subjects who received the low complexity task. ($p=.038$) Table 6.10 presents the results of the manipulation check.

6.8 TESTS OF HYPOTHESES

A fundamental goal of this dissertation was to replicate not only the experiment conducted by C&H, but also the method of analysis, Structural Equation Modeling. Using the software chosen for this analysis, AMOS©, hypotheses 4 through 9 were successfully tested using path analysis. Due to the bivariate structure of the training variable, however, separate analyses of variance were necessary to test hypotheses 1 through 3. Tests of the moderator, (hypotheses 10 through 12), were performed using analyses of variance and AMOS© with a multi-group strategy. Finally, hypotheses 13 and 14, which dealt with avoidance behavior, were tested using a chi-square test.

Table 6.10 Task Complexity Manipulation Check

One-Way Analysis of Variance

Source	df	SS	MS	F	p
Complexity	1	44.8	44.8	4.33	0.038*
Error	351	3612.9	10.4		
Total	352	3657.7			

Level	N	Mean	StDev
Low complexity	179	5.143	3.167
High complexity	174	5.858	3.263

6.8.1 Hypotheses 1 Through 3

Because the first nine hypotheses do not address interactions with task complexity, data were collapsed across levels of complexity for these analyses.

Hypothesis 1 predicted that behavior modeling would have a greater positive effect on computer self-efficacy than lecture-based training. As shown in Table 6.11, a univariate Analysis of Variance was performed on the data by collapsing (summing) the ten self-efficacy questions into a single index. This analysis revealed a significant difference between training types ($p = .035$). The difference between training methods, however, was not in the direction predicted by hypothesis 1. That is, lecture-based training resulted in significantly higher self-efficacy scores than behavior modeling. Thus, based on the results of this analysis, hypothesis 1 was not supported in this experiment.

Hypothesis 2 predicted that behavior modeling subjects would exhibit higher outcome expectations than lecture-based subjects. Like the data for hypothesis 1, responses for the outcome expectations questionnaire were collapsed into a single index and analyzed using an ANOVA. Results of this test, presented in Table 6.12, indicate that the null hypothesis (i.e. that no differences in outcome expectations exist between training methods), could not be rejected at the .05 alpha level ($p = .199$). Hypothesis 2 also was not supported in this experiment.

Hypothesis 3, which predicted higher performance scores for subjects who received behavior modeling training, also was tested using a univariate analyses of variance. Results of this test are presented in Table 6.13 and indicate a rejection of the null hypothesis, (i.e. no difference between training methods). Thus, Hypothesis 3 was supported in this study ($p = .056$).

Table 6.11 Test of Hypothesis #1

Hypothesis #1: Subjects who receive behavior modeling training will develop higher perceptions of computer self-efficacy than subjects who receive non-modeling training.

One-Way Analysis of Variance
(Collapsing indicators into a single index)

Source	DF	SS	MS	F	p
Training	1	1138	1138	4.47	.035*
Error	243	61825	254		
Total	244	62962			

```
                              Individual 95% CIs For Mean
                              Based on Pooled StDev
Level    N    Mean    StDev  -------+---------+---------+---------
  1    118   52.03   16.61  (--------*---------)                      Behavior modeling
  2    127   56.35   15.31              (--------*--------)           Lecture training
                              -------+---------+---------+---------

Pooled StDev =   15.95             51.0      54.0      57.0
```

* p < .05

The results in the table above indicate that the null hypothesis, (i.e. that no differences exist between training methods), could be rejected at the .10 alpha level using the univariate analysis of variance. This difference between training methods, however, was not in the direction predicted by hypothesis 1. Lecture-based training resulted in higher self-efficacy scores than behavior modeling.

Table 6.12 Test of Hypothesis #2

Hypothesis 2: Subjects who receive behavior modeling training will develop higher outcome expectations than subjects who receive non-modeling training.

One-Way Analysis of Variance
(collapsing indicators into a single index)

Source	**DF**	**SS**	**MS**	**F**	**p**
Training	1	34.9	34.9	1.66	0.199
Error	243	5110.9	21		
Total	244	5145.8			

```
                             Individual 95% CIs For Mean
                             Based on Pooled StDev
 Level    N     Mean    StDev  --------+---------+---------+--------
   1    118   22.678    4.485  (-----------*-----------)
   2    127   23.433    4.678          (-----------*----------)
                             --------+---------+---------+--------

Pooled StDev =   4.586              22.40     23.10     23.80
```

The results in the table above indicate that the null hypothesis, (i.e. that no differences exist between training methods), could not be rejected at the .05 alpha level.

Table 6.13 Test of Hypothesis #3

Hypothesis 3: Subjects who receive behavior modeling training will score higher than those in non-modeling training on measures of performance.

One-Way Analysis of Variance

Source	DF	SS	MS	F	p
Training	1	0.3665	0.3665	3.68	0.056 **
Error	243	24.2125	0.0996		
Total	244	24.579			

```
                              Individual 95% CIs For Mean
                              Based on Pooled StDev
 Level    N     Mean    StDev  --+---------+---------+---------+----
   1    118   0.6651   0.2854           (---------*--------)          Behavior Modeling
   2    127   0.5877   0.3414  (--------*--------)                    Lecture Based
                              --+---------+---------+---------+----

Pooled StDev =  0.3157         0.540     0.600     0.660     0.720
```

** $p < .10$

Results of this hypothesis indicate a rejection of the null hypothesis (that no differences exist between training methods on measures of performance) at the .10 alpha level.

6.8.2 Hypotheses 4 Through 9

Tests of hypotheses 4 through 9 were conducted by performing a path analysis using AMOS©. Because these hypotheses did not address interactions with task complexity, data were collapsed across levels of complexity.

Results are presented in Table 6.14, and the path diagram used in the analysis is illustrated in Figure 6.1. Table 6.14 reveals that the model exhibited an overall poor fit to the data (p = .000). In structural equation modeling, a significant p-value for a chi-square goodness-of-fit statistic indicates rejection of the null hypothesis and, thus, a poor fit between model and data.

In this study, because the significance of the path coefficients, rather than the overall model fit, is of primary importance in addressing the hypotheses, interpretation of the results is primarily focused on individual paths. Analysis of these results is based on the critical ratio computed for each path coefficient of interest. As this ratio is normally distributed, significance is determined by comparing it to the critical value of the t-distribution for a sample size commensurate with the one in the analysis, at the desired alpha level. In this case, at an alpha of .05, the critical value is 1.96. At an alpha level of .10, the critical value is 1.65.

Table 6.14 indicates that hypotheses 7, 8, and 9 were supported by the analysis in that their critical ratios (3.348, 1.964, and 2.311, respectively) exceeded the 1.96 critical value. These hypotheses were alike in that they addressed relationships between prior performance and the other constructs in the model, self-efficacy, outcome expectations, and final performance. Hypotheses 4, 5, and 6, which predicted significant relationships between self-efficacy, outcome expectations, and performance, were not supported.

Table 6.14 Tests of Hypotheses 4 through 9

Computation of Degrees of Freedom

Number of distinct sample moments:	231
Number of distinct parameters to be estimated:	49
Degrees of freedom:	182

Model Goodness of Fit

Chi-square = 441.133
Degrees of freedom = 182
Probability level = 0.000

Maximum Likelihood Estimates

	Regression Weights	Estimate	S.E.	C.R.	
H7	Self_Efficacy <--- Prior_Performance	0.469	0.140	3.348	*
	Self_Efficacy <------------ Training	0.327	0.209	1.568	
H4	Outcome_Expectations <- Self_Efficacy	0.031	0.044	0.704	
H8	Outcome_Expectations <- Prior_Performanc	0.121	0.062	1.964	*
	Outcome_Expectations <----- Training	0.063	0.107	0.586	
H5	Performance <--------- Self_Efficacy	-0.002	0.017	-0.121	
H9	Performance <----- Prior_Performance	0.060	0.026	2.311	*
H6	Performance <-- Outcome_Expectations	0.000	0.029	-0.004	
	Performance <-------------- Training	-0.089	0.040	-2.251	*

Items that are marked with an asterisk (*) are considered significant at the .05 alpha level, which corresponds to a critical value of 1.96. Results above indicate support for hypotheses 7,8, and 9, which predicted significant relationships between prior performance and the other constructs in the model.

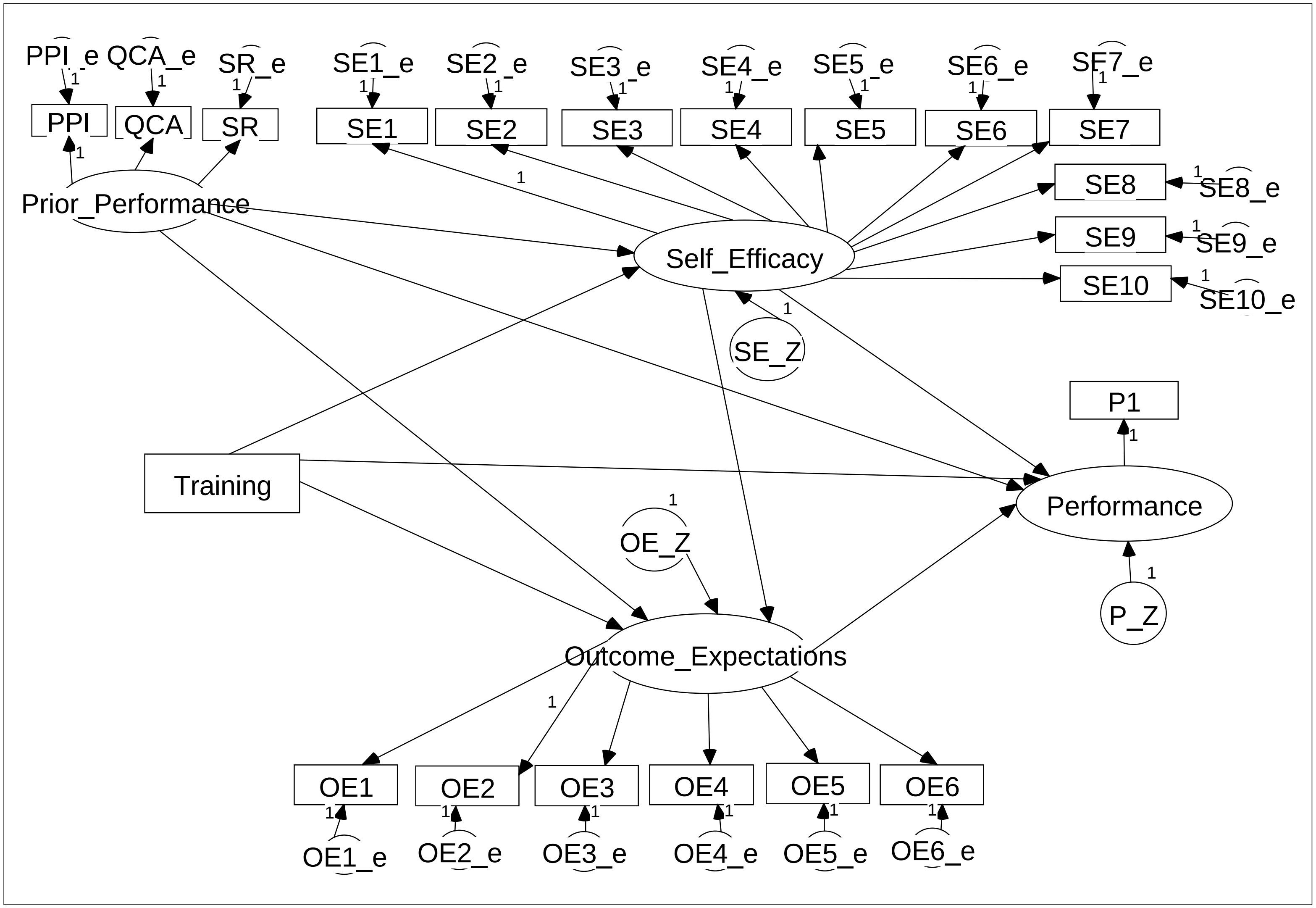

Figure 6.1 Path Diagram

6.8.2 Task Complexity Hypotheses (10 through 12)

Hypotheses 10 through 12 predicted interactions between the moderating variable, task complexity, and the relationships between training, performance, and self-efficacy. As in the analyses of the first three hypotheses that predicted relationships between training types and the endogenous variables in the model, Hypotheses 10 and 11, which address the training/performance and training/self-efficacy relationships, were analyzed using separate analyses of variance. It may be recalled that this type of analysis was necessary because the exogenous variable, training type, was bivariate and resulted in a matrix that was not positive definite.

As shown in Table 6.15, behavior modeling resulted in significantly higher performance scores than lecture based training ($p = .048$). The low task complexity level resulted in significantly higher performance scores than the high task level ($p = .003$). The interaction between task complexity and training type, however, was not significant ($p = .567$). Thus, although main effects were present, the interaction predicted by hypothesis 10 was not supported.

Table 6.15 Test of Hypothesis #10

Hypothesis 10: Behavior modeling will have a greater positive effect on performance than traditional lecture based training when task complexity is high.

Analysis of Variance

General Linear Model

Factor	Levels	Values	
Training	2	1	2
Task	2	1	2

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Training	1	0.36654	0.38264	0.38264	3.96	0.048 *
Task	1	0.90179	0.88829	0.88829	9.2	0.003 *
Training*Task	1	0.03182	0.03182	0.03182	0.33	0.567
Error	241	23.27885	23.27885	0.09659		
Total	244	24.579				

* p < .05

In the table above, the interaction between task complexity and training type indicates no significant differences in performance.

Hypothesis 11, which predicted that behavior modeling would have a greater positive effect on computer self-efficacy than lecture based training at the high level of complexity, also was not supported in this experiment (p = .348). Table 6.16 presents the results of this analysis. As shown in this table, the main effect of training method (p = .035) did demonstrate a significant difference on measures of self-efficacy; however, task complexity (p = .639) exhibited no significant effects on self-efficacy scores.

Hypothesis 12 predicted that self-efficacy would have a greater positive effect on performance for both behavior modeling and lecture based training when task complexity was low than when it was high. To analyze this interaction, a *multi-group* strategy was employed in AMOS© by constraining the path coefficient of interest (self-efficacy --> performance) across groups and comparing its results to the same multi-group model without constraints (Jaccard & Wan, 1996).

Table 6.16 Test of Hypothesis #11

Hypothesis 11: Behavior modeling will have a greater positive effect on computer self-efficacy than traditional lecture based training when task complexity is high.

Analysis of Variance
General Linear Model

Factor	Levels	Values	
Training	2	1	2
Task	2	1	2

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Training	1	1137.6	1143.3	1143.3	4.48	0.035 *
Task	1	48.5	56.3	56.3	0.22	0.639
Training*Task	1	225.8	225.8	225.8	0.88	0.348
Error	241	61550.4	61550.4	255.4		
Total	244	62962.2				

*** p < .05**

In the table above, the interaction between task complexity and training type indicates no significant differences in self-efficacy scores.

The multi-group strategy in structural equation modeling requires two steps. In the first step, goodness of fit is determined for the combined set of groups. A group is defined as a single level of the moderating variable. In this study, the two levels (groups) of the task complexity moderator proposed in hypotheses 12 were low complexity and high complexity. Thus, goodness of fit was determined for all data at the low level of complexity and again for all data at the high level of complexity. No constraints were imposed on any path coefficients for either of these data sets. That is, the software freely estimated all of the path coefficients separately for each of the two levels of complexity. It then determined a goodness of fit index for the two groups combined. Results of this model are presented in Table 6.17.

After these initial tests of model fit were performed, the second step, a test of the interaction, was executed. This second step imposed a constraint on the relationship between self-efficacy and performance. In step 1, the software freely estimated this path coefficient for both levels of complexity and produced an overall goodness of fit index. In step 2, it constrained the path coefficient of the low complexity level to be equal to that of the high complexity level. It then produced a goodness of fit index for the combined levels of complexity in step 2. Results of this step are presented in Table 6.18.

The difference between fit indices that was found between steps 1 and 2 was -0.041 (714.252-714.293). The final part of the analysis involved determining whether this difference was statistically significant. Representing the differential effects of the complexity levels, this result is chi-square distributed with degrees of freedom equal to the absolute value of the degrees of freedom of the constrained model, subtracted from those in the unconstrained model, ($df_2 - df_1 = |364-365| = 1$).

Table 6.17 Test of Hypothesis #12 (part 1) Model Without Constraints

Multi-group Model without constraints

Computation of Degrees of Freedom

Number of distinct sample moments:	462
Number of distinct parameters to be estimated:	98
Degrees of freedom:	364

Chi-square = 714.252
Degrees of freedom = 364
Probability level = 0.000

Results for group: High Task
Maximum Likelihood Estimates

Regression Weights:	Estimate	S.E.	C.R.
Self_Efficacy <--- Prior_Performance	0.680	0.227	2.999 *
Self_Efficacy <------------ Training	0.536	0.277	1.938 **
Outcome_Expectations <- Self_Efficacy	0.086	0.060	1.433
Outcome_Expectations <- Prior_Performance	0.101	0.086	1.177
Outcome_Expectations <----- Training	0.121	0.114	1.065
Performance <--------- Self_Efficacy	-0.008	0.032	-0.241
Performance <----- Prior_Performance	0.048	0.047	1.020
Performance <-- Outcome_Expectations	0.067	0.064	1.037
Performance <-------------- Training	-0.113	0.063	-1.811 **

Results for group: Low Task
Maximum Likelihood Estimates

Regression Weights:	Estimate	S.E.	C.R.
Self_Efficacy <--- Prior_Performance	0.404	0.197	2.054 *
Self_Efficacy <------------ Training	0.102	0.317	0.322
Outcome_Expectations <- Self_Efficacy	0.005	0.064	0.080
Outcome_Expectations <- Prior_Performance	0.135	0.097	1.385
Outcome_Expectations <----- Training	-0.006	0.180	-0.032
Performance <--------- Self_Efficacy	0.000	0.019	-0.009
Performance <----- Prior_Performance	0.059	0.034	1.740 **
Performance <-- Outcome_Expectations	-0.015	0.029	-0.516
Performance <-------------- Training	-0.740	0.050	-1.503

* $p < .05$ Critical value = 1.96
** $p < .10$ Critical value = 1.65

Table 6.18 Test of Hypothesis #12 (part 2) Model With Constraint

Multi-group Model
Constraint on Self-efficacy --> Performance

Computation of Degrees of Freedom

Number of distinct sample moments:	462
Number of distinct parameters to be estimated:	97
Degrees of freedom:	365

Chi-square = 714.293
Degrees of freedom = 365
Probability level = 0.000

Results for group: High Task
Maximum Likelihood Estimates

Regression Weights:	Estimate	S.E.	C.R.	
Self_Efficacy <--- Prior_Performance	0.678	0.224	3.022	*
Self_Efficacy <------------ Training	0.537	0.277	1.938	**
Outcome_Expectations <- Self_Efficacy	0.086	0.060	1.439	
Outcome_Expectations <- Prior_Performance	0.101	0.085	1.184	
Outcome_Expectations <----- Training	0.121	0.114	1.065	
Performance <--------- Self_Efficacy	-0.002	0.016	-0.129	
Performance <----- Prior_Performance	0.043	0.038	1.126	
Performance <-- Outcome_Expectations	0.065	0.064	1.026	
Performance <-------------- Training	-0.116	0.061	-1.901	**

Results for group: Low Task
Maximum Likelihood Estimates

Regression Weights:	Estimate	S.E.	C.R.	
Self_Efficacy <--- Prior_Performance	0.413	0.194	2.122	*
Self_Efficacy <------------ Training	0.101	0.317	0.318	
Outcome_Expectations <- Self_Efficacy	0.004	0.065	0.060	
Outcome_Expectations <- Prior_Performance	0.138	0.098	1.410	
Outcome_Expectations <----- Training	-0.006	0.180	-0.034	
Performance <--------- Self_Efficacy	-0.002	0.016	-0.129	
Performance <----- Prior_Performance	0.061	0.033	1.845	**
Performance <-- Outcome_Expectations	-0.016	0.029	-0.531	
Performance <-------------- Training	-0.074	0.050	-1.501	

* p< .05 Critical value = 1.96
** p< .10 Critical value = 1.65

Comparing this value (-0.041) to the two critical values of 3.841 (a = .05) and 2.706 (a = .10), which represent the critical values of the chi-square distribution for 1 degree of freedom, this difference was not deemed statistically significant. Thus, hypothesis 12 was not supported in this study.

6.8.3 Hypotheses 13 and 14

Structural equation modeling was not used to test Hypotheses 13 and 14, which addressed avoidance behavior. More precisely, hypothesis 13 predicted that subjects in the behavior modeling training would exhibit avoidance behavior less frequently (i.e. they would elect to use Solver more frequently) than subjects in the non-modeling groups. Because this hypothesis involved frequencies of occurrences, a non-parametric test (i.e. a chi-square test) was used. Data from both levels of complexity were combined for each training method, and for each group, the number of subjects who used and did not use Solver were summed.

As illustrated in Table 6.19, results of this test revealed a significant difference between training methods for the subjects who elected to use Solver and those who did not (p=.008). Approximately 13% of behavior modeling subjects failed to use Solver while nearly 30% of subjects receiving the lecture-based training failed to use Solver. Based on this result, hypothesis 13 was supported.

Hypothesis 14 predicted that subjects would be more likely to exhibit avoidance behavior at the high level of complexity than at the low level of complexity. Again, a chi-square test was used in this analysis. In this analysis, however, training methods were collapsed across levels of complexity. Table 6.20 presents the results of this analysis.

As shown in Table 6.20, there was no significant difference between levels of complexity for the number of subjects who elected to use Solver as compared to those who did not (p = .965). Thus, hypothesis 14 was not supported in this study.

Table 6.19 Test of Hypothesis #13

Hypothesis 13: Subjects who receive behavior modeling training will elect to use Solver more frequently than lecture based subjects.

Chi-Square Test

Expected counts are printed below observed counts

	Training Type		
	Behavior Modeling	**Lecture Based**	**Total**
Used Solver	118 109.35	127 135.65	245
Did Not Use Solver	15 23.65	38 29.35	53
Total	133	165	298

ChiSq = 0.685 + 0.552 + 3.166 + 2.552 = 6.956 df = 1

p = 0.008 *

The chi-square test above is used to test the frequencies of subjects who use Solver across the two training methods: behavior modeling and lecture based training. At alpha = .05, the null hypothesis (i.e. no difference between training methods) is rejected.

Table 6.20 Test of Hypothesis #14

Hypothesis 14: Subjects who receive the low complexity task will elect to use Solver more frequently than subjects who receive the high complexity task.

Chi-Square Test

Expected counts are printed below observed counts

	Complexity		
	High	**Low**	**Total**
Used Solver	121 120.86	124 124.14	245
Did Not Use Solver	26 26.14	27 26.86	53
Total	147	151	298

ChiSq = 0.000 + 0.000 + 0.001 + 0.001 = 0.002   df = 1

p = 0.965

The chi-square test above is used to test the frequencies of subjects who use Solver across the two levels of complexity: high and low. At alpha = .05, the null hypothesis (i.e. that no differences exist between complexity levels) cannot be rejected.

6.8.4 Empirical Analysis of CAI Training

Although not formally hypothesized, the relationships depicted in the original C&H study were tested for the CAI training method. As shown in Table 6.21, an assessment of multivariate normality was first performed to determine if the CAI data set suffered skewness and/or kurtosis. Using a ±2.0 benchmark, Table 6.21 illustrates that no items exhibited significant skewness or kurtosis.

Table 6.21 Tests of Normality: Skewness and Kurtosis

Assessment of normality
Computer Assisted Instruction

Indicator	Skewness	Kurtosis	Indicator
Self-Reported Expertise	0.61	-1.628	Self-Reported Expertise
QCA	0.156	-1.021	QCA
Prior Performance Index	-0.383	-0.548	Prior Performance Index
Final Performance	-0.457	-1.422	Final Performance
Outcome Expectations #6	-0.279	-0.824	Outcome Expectations #6
Outcome Expectations #5	-0.524	-0.542	Outcome Expectations #5
Outcome Expectations #4	-1.267	1.861	Outcome Expectations #4
Outcome Expectations #3	-0.239	-1.22	Outcome Expectations #3
Outcome Expectations #1	-0.512	-0.438	Outcome Expectations #1
Outcome Expectations #2	-0.988	-0.367	Outcome Expectations #2
Self-Efficacy #10	-0.92	0.221	Self-Efficacy #10
Self-Efficacy #9	-0.76	-0.632	Self-Efficacy #9
Self-Efficacy #8	-0.011	-1.059	Self-Efficacy #8
Self-Efficacy #7	-0.64	-0.666	Self-Efficacy #7
Self-Efficacy #6	-0.306	-0.06	Self-Efficacy #6
Self-Efficacy #5	-0.253	-0.445	Self-Efficacy #5
Self-Efficacy #4	0.167	-0.731	Self-Efficacy #4
Self-Efficacy #3	-0.054	-0.639	Self-Efficacy #3
Self-Efficacy #2	0.654	-0.675	Self-Efficacy #2
Self-Efficacy #1	0.621	-0.658	Self-Efficacy #1

To assess any possible outliers in the model, the Mahalanobis distance statistic was produced and is illustrated in Table 6.22. As shown in this table, only one observation, (#4), had a value in the p2 column that was less than the .10 criterion established in the first analysis. After close inspection of this observation, it was determined that its unusually large distance from the centroid of the data was attributable to two "0" self-efficacy scores and a "0" final performance score. As in the previous analysis, these results were acceptable responses and thus, did not require elimination from the analysis.

Results of the path analysis are presented in Table 6.23. As shown in this table, only one path coefficient was deemed significant. This relationship was the prior performance --> self-efficacy path, and its significance was only at the .10 alpha level.

Table 6.22 Analysis of Outliers: CAI

Analysis of Outliers
Observations farthest from the centroid
(Mahalanobis distance)

Observation number	Mahalanobis d-squared	p1	p2
4	16.363	0.694	0.091
41	16.459	0.688	0.138
30	16.666	0.675	0.168
23	10.209	0.964	0.174
20	21.532	0.366	0.191
2	11.297	0.938	0.194
6	14.627	0.797	0.216
39	16.743	0.67	0.237
44	19.935	0.462	0.25
26	15.235	0.763	0.267

In the table above, the top ten observations that exhibit the greatest squared distance from the centroid of the CAI data set are presented. A threshold of .10 in the p2 column was used to identify possible outliers. Only observation #4 was less than .10, and analysis of its content revealed valid responses. No observations were considered outliers in the CAI data set.

Table 6.23 Path Analysis: CAI

Path Analysis Results
Computer Assisted Instruction

Computation of Degrees of Freedom

Number of distinct sample moments:	210
Number of distinct parameters to be estimated:	45
Degrees of freedom:	165

Chi-square = 259.689
Degrees of freedom = 165
Probability level = 0.000

Maximum Likelihood Estimates

Regression Weights:	Estimate	S.E.	C.R.	
Self_Efficacy <--- Prior_Performance	0.725	0.414	1.752	**
Outcome_Expectations <- Self_Efficacy	-0.022	0.063	-0.344	
Outcome_Expectations <- Prior_Performance	0.17	0.133	1.275	
Performance <--------- Self_Efficacy	0.007	0.054	0.13	
Performance <----- Prior_Performance	0.188	0.134	1.399	
Performance <-- Outcome_Expectations	-0.018	0.192	-0.092	

Results of the path analysis conducted for Computer Assisted Instruction are presented above. To determine significance of path coefficients, each path's critical ratio is compared to the critical value for the desired alpha level, where the critical value is 1.96 at alpha = .05 and 1.65 at alpha = .10. As shown above, only one path (Prior Performance --> Self-efficacy) is deemed significant at the .10 alpha level.

In addition to path analysis, an ANOVA was conducted to test the effectiveness of the CAI method on performance in relation to the behavior modeling and lecture based methods. Results of this analysis, as presented in Table 6.24, revealed a significant difference in performance between training methods (p = .075) at an alpha level of .10.

Table 6.24 Comparison of Training Methods: ANOVA for Final Performance

One-Way Analysis of Variance
Dependent Variance: Final Performance

Source	DF	SS	MS	F	p
Training	2	0.573	0.287	2.61	0.075**
Error	288	31.615	0.11		
Total	290	32.188			

```
                         Individual 95% CIs For Mean
                         Based on Pooled StDev
Level    N     Mean    StDev  ----+---------+---------+---------+--
  1    118   0.6651   0.2854                (------*-------)          Behavior Modeling
  2    127   0.5877   0.3414        (------*-------)                  Lecture Based
  3     46   0.5520   0.4056  (-----------*-----------)               CAI
                               ----+---------+---------+---------+--

Pooled StDev =  0.3313           0.480     0.560     0.640     0.720
```

** p < .10

The results above indicate a significant difference between training methods on the final performance measure at an alpha level of .10.

To determine which levels of training were responsible for this difference, a Tukey's pairwise comparison was conducted and presented in Table 6.25. As illustrated, the behavior modeling technique proved significantly better in terms of overall performance scores than both the lecture based and CAI methods. No differences between the two non-modeling methods, however, were exhibited.

An analysis of variance also was conducted to determine the effectiveness of all three training methods on levels of outcome expectations. Table 6.26 illustrates a significant difference between training methods ($p = .06$) at an alpha level of .10, and Table 6.27 depicts the pairwise comparisons of these methods. Results indicate that the CAI method significantly outperformed both the behavior modeling and lecture based training methods on outcome expectations scores at alpha = .05. No differences were found between the behavior modeling and lecture based methods, however.

Table 6.25 Comparison of Training Methods: Pairwise Comparisons for Final Performance

Tukey's pairwise comparisons

Dependent Variable: Final Performance

	Level	**N**	**Mean**
1	Behavior Modeling	118	0.6651
2	Lecture Based	127	0.5877
3	CAI	46	0.552

MSerror = .11 (from Table 6.24)

Critical Value: $[q(r_{max}, df_{error})][sqrt(MS_{error}/n)]$ = [q(3,288)][sqrt(.11/291)]
= (3.31)(0.019) = **0.0644**

	1	2	3
1	---	0.0774	0.1131
2		---	0.0357
3			---

Using Tukey's Pairwise comparisons, it can be seen that the behavior modeling training type is significantly better than both lecture based training and CAI in terms of its effects on performance. No differences, however, exist between the lecture based training and CAI on measures of performance.

Table 6.26 Comparison of Training Methods: ANOVA for Outcome Expectations

One-Way Analysis of Variance
Dependent Variance: Outcome Expectations

Source	**DF**	**SS**	**MS**	**F**	**p**
Training	2	109.4	54.7	2.84	0.06 **
Error	288	5544.4	19.3		
Total	290	5653.8			

```
                           Individual 95% CIs For Mean
                           Based on Pooled StDev
Level    N     Mean     StDev  --------+---------+---------+--------
  1    118   22.678    4.485  (------*------)
  2    127   23.433    4.678       (-----*------)
  3     46   24.457    3.103         (----------*---------)

                           --------+---------+---------+--------
Pooled StDev =   4.388              22.8      24.0      25.2
```

** p < .10

The ANOVA above indicates a significant difference between the levels of training on scores of outcome expectations at the .10 alpha level.

Table 6.27 Comparison of Training Methods: Pairwise Comparisons for Outcome Expectations

Tukey's pairwise comparisons

Dependent Variable: Outcome Expectations

	Level	**N**	**Mean**
1	Behavior Modeling	118	22.678
2	Lecture Based	127	23.433
3	CAI	46	24.457

MSerror = 19.3 (from Table 6.26)

Critical Value: $[q(r_{max}, df_{error})][sqrt(MS_{error}/n)] = [q(3,288)][sqrt(19.3/291)]$
$= (3.31)(0.258) =$ **0.8524**

	1	2	3
1	---	-0.755	-1.779
2		---	-1.024
3			---

Using Tukey's Pairwise comparisons, it can be seen that the CAI method resulted in significantly higher outcome expectations scores than both behavior modeling and lecture based training. Further, no differences were found between behavior modeling and lecture based training on outcome expectations.

To determine the effects of training on self-efficacy, an ANOVA was again conducted. Presented in Table 6.28, results of this analysis revealed a significant difference ($p = .094$) between training methods on measures of self-efficacy at an alpha level of .10. As shown in the pairwise comparisons presented in Table 6.29, behavior modeling resulted in significantly lower self-efficacy scores than lecture-based training and CAI.

Table 6.28 Comparison of Training Methods: ANOVA for Self-Efficacy

One-Way Analysis of Variance
Dependent Variance: Self-Efficacy

Source	DF	SS	MS	F	p
Training	2	1272	636	2.38	0.094 **
Error	288	76856	267		
Total	290	78128			

```
                        Individual 95% CIs For Mean
                        Based on Pooled StDev
Level    N     Mean    StDev  ----------+---------+---------+------
   1   118   52.03   16.61  (--------*-------)
   2   127   56.35   15.31            (-------*-------)
   3    46   56.13   18.28       (------------*-------------)
                        ----------+---------+---------+------

Pooled StDev =   16.34            52.5      56.0      59.5
```

** p < .10

The table above illustrates a significant difference between training methods on measures of self-efficacy at an alpha level of .10.

Table 6.29 Comparison of Training Methods: Pairwise Comparisons for Self-Efficacy

Tukey's pairwise comparisons
Dependent variable: Self-Efficacy

Level	N	Mean
1 Behavior Modeling	118	52.03
2 Lecture Based	127	56.35
3 CAI	46	56.13

MSerror = 267 (from Table 6.28)

Critical Value: $[q(r_{max}, df_{error})][sqrt(MS_{error}/n)] = [q(3,288)][sqrt(267/291)]$
$= (3.31)(0.958) =$ **3.1706**

	1	2	3
1	---	-4.32	-4.1
2		---	0.22
3			---

The table above illustrates a significant difference between behavior modeling and lecture-based training and between Behavior modeling and CAI on measures of self-efficacy. No differences are noted between lecture and CAI. In all instances, behavior modeling produced significantly lower self-efficacy scores than the other two training methods.

Task complexity also was tested for its potential moderating effect on performance, self-efficacy and outcome expectations. As shown in Table 6.30, the interaction between training methods and complexity on performance was statistically significant (p = .041).

An illustration of this interaction is provided in Figure 6.2. Examining this graph reveals that at low levels of complexity, behavior modeling and lecture based training outperformed CAI, but that this effect was not exhibited for the high level of complexity. In fact, Table 6.31 provides pairwise comparisons on this interaction and reveals that behavior modeling and CAI significantly outperformed lecture based training at the high level of complexity. At the low level of complexity, however, behavior modeling resulted in significantly higher performance scores than both lecture based training and CAI. Further, lecture based training significantly outperformed CAI at this low level of complexity.

Table 6.30 Comparison of Training Methods: Test of Moderator on Performance

Analysis of Variance
General Linear Model
Test of Interaction: Training x Task
Dependent Variable: Final Performance

Factor	Levels	Values		
Training	3	1	2	3
Task	2	1	2	

Source	DF	Seq SS	Adj SS	Adj MS	F	P	
Training	2	0.5732	0.5674	0.2837	2.65	0.072	**
Task	1	0.4665	0.0669	0.0669	0.63	0.429	
Training*Task	2	0.6921	0.6921	0.346	3.24	0.041	*
Error	285	30.4564	30.4564	0.1069			
Total	290	32.1882					

* $p < .05$
** $p < .10$

The results presented above indicate a significant interaction effect between training levels and task complexity on measures of performance.

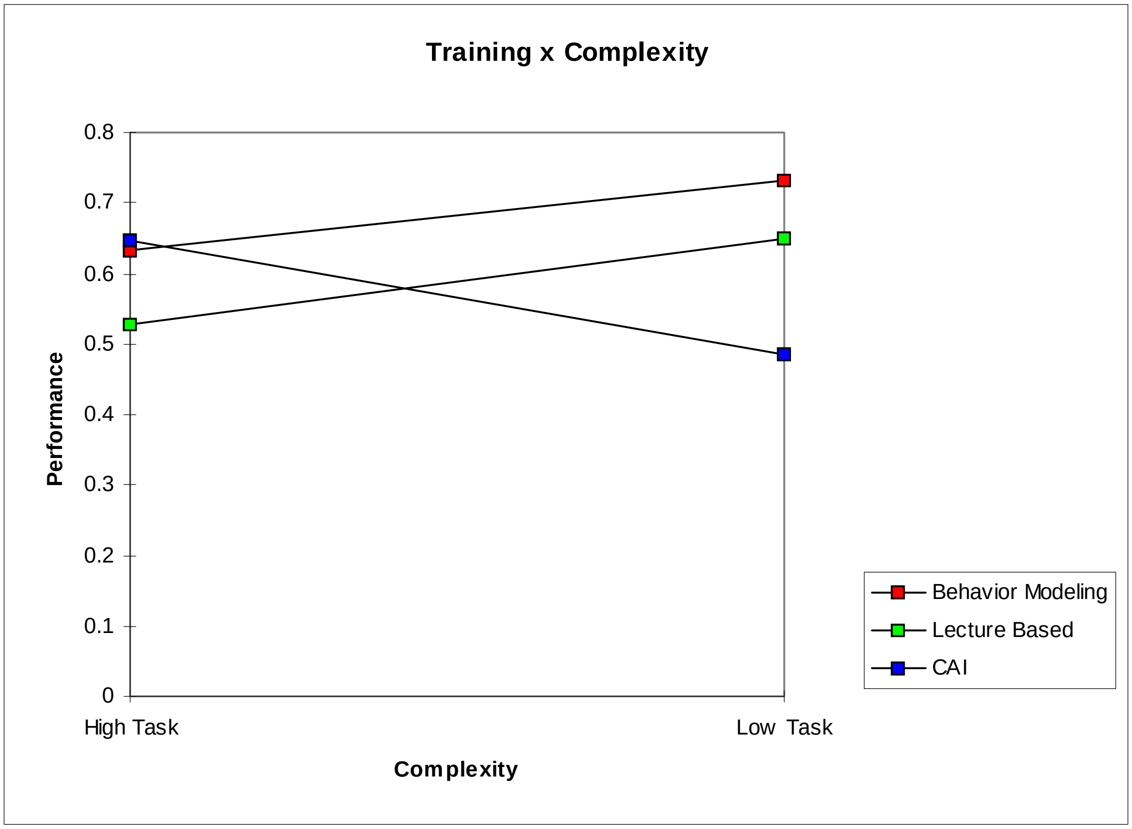

Figure 6.2 Chart of Training X Complexity Interaction

As shown in the chart above, Behavior Modeling and CAI outperformed Lecture-based training when task complexity was high. At the low level of complexity, however, CAI performed worse than both Behavior Modeling and Lecture.

Table 6.31 Comparison of Training Methods: Pairwise Comparisons for Task Complexity and Performane

Tukey's pairwise comparisons

Tests of interaction Training * Task

Dependent Variable: Performance

	Level	**N**	**Mean**
1	Behavior Modeling/High Task	59	0.631746
2	Behavior Modeling/Low Task	59	0.732143
3	Lecture Based/High Task	62	0.527385
4	Lecture Based/Low Task	65	0.65
5	CAI/High Task	22	0.645833
6	CAI/Low Task	24	0.485

MSerror = .1069 (from Table 6.30)

Critical Value: $[q(r_{max}, df_{error})][sqrt(MS_{error}/n)] = [q(6,303)][sqrt(0.1069/291)]$
$= (4.03)(0.019) =$ **0.077**

	1	2	3	4	5	6
1	---	0.100397	-0.10436	0.018254	0.01409	-0.14675
2		---	-0.20476	-0.08214	-0.08631	-0.24714
3			---	0.122615	0.11845	-0.04238
4				---	-0.00417	-0.165
5					---	-0.16083
6						---

Using Tukey's Pairwise comparisons, it can be seen that behavior modeling and lecture based training significantly outperformed CAI at the high level of complexity on measures of final performance. At the low level of complexity, however, behavior modeling resulted in significantly higher performance scores than both lecture based training and CAI. Further, a significant difference between lecture based and CAI was found, with lecture based outperforming CAI at the low level of complexity.

Table 6.32 illustrates the interaction between complexity and training on measures of self-efficacy, and Table 6.33 presents the results of this test on outcome expectations. As shown in these tables, no significant interactions were demonstrated for either of these dependent variables (p = .287 and p = .909, respectively).

Table 6.32 Comparison of Training Methods: Test of Moderator on Self-Efficacy

Analysis of Variance
General Linear Model
Test of Interaction: Training x Task
Dependent Variable: Self-Efficacy

Factor	Levels	Values		
Training	3	1	2	3
Task	2	1	2	

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Training	2	1271.7	1253.6	626.8	2.35	0.097 **
Task	1	278.7	601	601	2.26	0.134
Training*Task	2	668.2	668.2	334.1	1.25	0.287
Error	285	75908.9	75908.9	266.3		
Total	290	78127.6				

* p < .05
** p < .10

The results presented above indicate no significant interactive effects between training type and task complexity on measures of self-efficacy.

Table 6.33 Comparison of Training Methods: Test of Moderator on Outcome Expectations

Analysis of Variance
General Linear Model
Test of Interaction: Training x Task
Dependent Variable: Outcome Expectations

Factor	Levels	Values		
Training	3	1	2	3
Task	2	1	2	

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Training	2	124.9	125.35	62.68	3.29	0.038
Task	1	9.99	4.88	4.88	0.26	0.613
Training*Task	2	3.64	3.64	1.82	0.1	0.909
Error	285	5764.59	5764.59	19.03		
Total	290	5903.13				

The results presented above indicate no significant interactive effects between training type and task complexity on measures of outcome expectations.

A final analysis of the CAI data set was performed to determine the moderating effect of task complexity on the relationship between self-efficacy and performance. It may be recalled that this relationship was addressed in hypothesis 12 for the behavior modeling and lecture based training types. Although not formally hypothesized, the results of this analysis for the CAI method revealed no significant differences between levels of complexity on the self-efficacy --> performance path. As in the analysis for hypothesis 12, a multi-group strategy was employed whereby the CAI data were divided into two groups, low and high task complexity. This model was analyzed, and a chi-square statistic was computed to assess overall goodness of fit. Table 6.34 presents the results of this analysis. Next, the relationship between self-efficacy and final performance was constrained across the two groups. A chi-square statistic was then computed for this constrained model. Table 6.35 illustrates this second step. Finally. significance was determined by subtracting the chi-square goodness-of-fit statistic of the constrained model from that of the unconstrained model. This result (755.947 - 753.996 = 1.951) was then compared to the chi-square critical value (3.841) for alpha = .05 with 1 degree of freedom (331 -330 = 1). Because the computed difference did not exceed the critical value, the test of this interaction was not deemed significant.

6.8.5 Summary of Findings

Table 6.36 illustrates a summary of the hypotheses' results. Table 6.37 presents the results of the empirical investigation of the CAI method in addition to a summary of the analyses comparing the three training methods. Interpretations of all results are presented in the following chapter with a discussion of their implications, limitations, and directions for future research.

Table 6.34 Path Analysis for CAI: Model Without Constraints

Path Analysis for CAI

Unconstrained Model

Computation of Degrees of Freedom

Number of distinct sample moments:	420
Number of distinct parameters to be estimated:	90
Degrees of freedom:	330

Chi-square = 753.996
Degrees of freedom = 330
Probability level = 0.000

Results for group: CAI High Task
Maximum Likelihood Estimates

Regression Weights:	Estimate	S.E.	C.R.	
Self_Efficacy <--- Prior_Performance	0.56	0.533	1.051	
Outcome_Expectations <- Self_Efficacy	0.007	0.032	0.204	
Outcome_Expectations <- Prior_Performance	0.057	0.07	0.814	
Performance <--------- Self_Efficacy	-0.061	0.089	-0.69	
Performance <----- Prior_Performance	0.212	0.203	1.045	
Performance <-- Outcome_Expectations	0.195	0.509	0.383	

Results for group: CAI Low Task
Maximum Likelihood Estimates

Regression Weights:	Estimate	S.E.	C.R.	
Self_Efficacy <--- Prior_Performance	0.843	0.502	1.681	**
Outcome_Expectations <- Self_Efficacy	-0.032	0.114	-0.276	
Outcome_Expectations <- Prior_Performance	0.244	0.221	1.101	
Performance <--------- Self_Efficacy	0.106	0.062	1.71	**
Performance <----- Prior_Performance	0.096	0.119	0.811	
Performance <-- Outcome_Expectations	0.034	0.142	0.239	

Results of the path analysis using a multi-group strategy are presented above. Note that all paths were freely estimated. To determine significance levels of individual paths, each path's critical ratio is compared to the critical value 1.96 for alpha = .05 or 1.65 for alpha = .10. As shown above, no paths were deemed significant at the high level of complexity. For the low level of complexity, however, the relationships between prior performance --> self-efficacy and self-efficacy --> performance were significant at the .10 alpha level.

Table 6.35 Path Analysis for CAI: Contraint on Self-Efficacy --> Performance

Path Analysis for CAI

Self-Efficacy --> Performance Constrained

Computation of Degrees of Freedom

Number of distinct sample moments:	420
Number of distinct parameters to be estimated:	89
Degrees of freedom:	331

Chi-square = 755.947
Degrees of freedom = 331
Probability level = 0.000

Results for group: CAI High Task
Maximum Likelihood Estimates

Regression Weights:	Estimate	S.E.	C.R.	
Self_Efficacy <--- Prior_Performance	0.488	0.47	1.038	
Outcome_Expectations <- Self_Efficacy	0.011	0.029	0.371	
Outcome_Expectations <- Prior_Performance	0.048	0.059	0.826	
Performance <--------- Self_Efficacy	-0.039	0.07	-0.566	
Performance <----- Prior_Performance	0.174	0.156	1.113	
Performance <-- Outcome_Expectations	0.262	0.441	0.594	

Results for group: CAI Low Task
Maximum Likelihood Estimates

Regression Weights:	Estimate	S.E.	C.R.	
Self_Efficacy <--- Prior_Performance	1.403	0.745	1.884	**
Outcome_Expectations <- Self_Efficacy	-0.211	0.184	-1.151	
Outcome_Expectations <- Prior_Performance	0.675	0.476	1.42	
Performance <--------- Self_Efficacy	-0.039	0.07	-0.566	
Performance <----- Prior_Performance	0.484	0.309	1.564	
Performance <-- Outcome_Expectations	-0.229	0.295	-0.777	

Results of the path analysis using a multi-group strategy are presented above. In this analysis, the relationship between self-efficacy and final performance was constrained across groups (i.e. tasks). This path is highlighted in the table above. As illustrated, paths marked with a single asterisk (*) are significant at the .05 alpha level, while paths marked with two asterisks (**) are significant at the .10 alpha level.

Table 6.36 Summary of Hypotheses' Results

Summary of Hypotheses
Full Sample Compared to Student Sample

	p values or critical ratios	Training Methods Examined	Comments
H1: Training --> Self-Efficacy	0.035*	BM LB	Behavior modeling resulted in lower self-efficacy scores than lecture. Opposite of predicted direction. H1 not supported.
	.094**	BM LB CAI	Behavior modeling resulted in lower self-efficacy scores than both lecture and CAI. Opposite of predicted direction. This was significant at alpha = .10 .
H2: Training --> Outcome Expectations	0.199	BM LB	H2 not supported.
	.06**	BM LB CAI	CAI resulted in higher outcome expectations than both behavior modeling and lecture.
H3: Training --> Performance	.056**	BM LB	Behavior modeling resulted in higher performance scores than lecture.
	.075**	BM LB CAI	Behavior modeling outperformed both lecture and CAI. No differences between lecture and CAI.
H4: Self-Efficacy --> Outcome Expectations	CR: 0.704	BM LB	Critical ratios are compared to the critical values of 1.96 for alpha = .05 and 1.65 for alpha = .10. As the CR values do not exceed these numbers, H4 was not supported.
	CR: 0.456	BM LB CAI	
H5: Self-Efficacy --> Performance	CR: -0.121	BM LB	Using the 1.96 and 1.65 thresholds, H5 also was not supported.
	CR: 0.196	BM LB CAI	
H6:Outcome Expectations --> Performance	CR: -0.004	BM LB	Using the 1.96 and 1.65 thresholds, H6 also was not supported.
	CR: 0.118	BM LB CAI	
H7: Prior Performance --> Self-Efficacy	CR: 3.348*	BM LB	Because the critical ratios exceed the critical value of 1.96, H7 was supported at alpha = .05 in all analyses.
	CR: 3.914*	BM LB CAI	
H8: Prior Performance --> Outcome Expectations	CR: 1.964*	BM LB	Because the critical ratios exceed the critical value of 1.96, H8 was supported at alpha = .05.
	CR: 2.461*	BM LB CAI	
H9: Prior Performance --> Performance	CR: 2.311*	BM LB	Because the critical ratios exceed the critical value of 1.96, H9 was supported at alpha = .05.
	CR: 2.757*	BM LB CAI	
H10: Training x Task --> Performance	0.567	BM LB	No significant interaction. H10 not supported.
	.041*	BM LB CAI	Behavior modeling and CAI outperformed lecture at high complexity. At low complexity, behavior modeling was the best, lecture was the next best, and CAI was the worst.
H11: Training x Task --> Self-Efficacy	0.348	BM LB	No significant interaction. H11 not supported.
	0.287	BM LB CAI	No significant interaction.
H12: Training x Task --> (Self-Efficacy --> Performance)	CR: .041	BM LB	As the critical ratios did not exceed the critical values of 3.841 at an alpha level of .05 nor 2.706 for alpha = .10, H12 was not supported.

* p < .05
** p < .10

Table 6.37 Summary of CAI Empirical Investigation and Comparison of Training Methods

Summary of CAI Empirical Investigation

Path Analysis	Significance
Self_Efficacy <--- Prior_Performance	yes**
Outcome_Expectations <- Self_Efficacy	no
Outcome_Expectations <- Prior_Performance	no
Performance <--------- Self_Efficacy	no
Performance <----- Prior_Performance	no
Performance <-- Outcome_Expectations	no

Comparison of Training Methods

Dependent variable	Result
Performance	Behavior Modeling outperformed both CAI and Lecture. No differences between Lecture and CAI.
Self-efficacy	No differences between training methods
Outcome Expectations	CAI produced higher outcome expectations than behavior modeling and lecture based training.

Task Complexity Interactions in all Training Methods

Dependent variable	Result
Performance	At the high complexity level, behavior modeling and CAI outperformed lecture. At low complexity, differences were significant between all three with ranking as follows: behavior modeling (best), lecture (middle), CAI (last).
Self-efficacy	No complexity x training method interaction
Outcome Expectations	No complexity x training method interaction
Self-efficacy --> performance (path)	No moderating effect of complexity in any training method.

* p < .05
** p < .10