DOCUMENT RESUME

ED 476 862 TM 034 956

AUTHOR	Hamilton, Jennifer; Gagne, Phillip E.; Hancock, Gregory R.
TITLE	The Effect of Sample Size on Latent Growth Models.
PUB DATE	2003-04-00
NOTE	23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 21-25, 2003).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Mathematical Models; Monte Carlo Methods; Research Methodology; *Sample Size
IDENTIFIERS	*Latent Growth

ABSTRACT

A Monte Carlo simulation approach was taken to investigate the effect of sample size on a variety of latent growth models. A fully balanced experimental design was implemented, with samples drawn from multivariate normal populations specified to represent 12 unique growth models. The models varied factorially by crossing number of time points, variance of intercept factor, and variance of slope factor. Simulation results show that sample size was found to influence the convergence rates of the models, with larger samples resulting in fewer improper estimates and failures. However, this effect is lessened if the variances of the slope and intercept factors are low. It is also lessened when more timepoints are added to the model. This research reinforces previous findings on the importance of sample size used in latent models. The paper also makes recommendations about sample size for researchers hoping to use latent growth models. (Contains 7 tables, 6 figures, and 15 references.) (SLD)

The Effect of Sample Size on Latent Growth Models

Jennifer Hamilton
Phillip E. Gagné
Gregory R. Hancock

1

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in Chicago, IL
April 2003

05/19/03

Jennifer Hamilton, Phillip E. Gagné, and Gregory R. Hancock
University of Maryland, College Park

The Effect of Sample Size on Latent Growth Models

BACKGROUND

Understanding how individuals change over time has long been of interest to researchers and practitioners from many disciplines. Degree of depression, socioeconomic status, level of reading ability, and quality of mother-child attachment are all examples of the wide applicability of longitudinal investigations in the social sciences. There are many statistical methods available to researchers wishing to examine such changes over time. They include univariate and multivariate analysis of variance and covariance, as well as auto-regressive and cross-lagged multiple regression techniques.

One increasingly popular statistical approach, latent growth-curve modeling (LGM), has emerged fairly recently from the area of structural equation modeling (SEM) as a relatively simple and flexible technique for modeling change over time at the individual as well as group level. However, while LGM has many advantages over more traditional methods, the issue of sample size must be addressed so researchers know under what conditions they may pursue, and trust, LGM. As in the more general SEM, asymptotic behavior of data-model fit (i.e., chi-square) and of LGM parameters is assumed whereas the actual behavior at smaller and more realistic sample sizes is largely unknown. Research to determine how large sample sizes should be for SEM applications has so far been inconclusive (e.g., MacCallum, Roznowski, & Necowitz, 1992; Tanaka, 1987). Instead, various sample size guidelines have been proposed -- 50 observations per variable, no less than 100 observations total, 5 to 10 observations per parameter, and so on. To further complicate matters, adequate sample size might depend on many conditions including the size, complexity and type of model, distribution and reliability of the variables, and strength of the relations among those variables (Marsh, Han, Balla & Grayson, 1998; Muthén & Muthén, 2002; Gagné & Hancock, 2002). As LGM focuses on models with characteristics unlike most typical structural models (e.g., fixed loadings), and as growth models typically have fewer free parameters requiring estimation, behavior of LGM under smaller samples sizes requires its own focused investigation.

The purpose of this simulation is to understand the behavior of a variety of latent growth models at different sample sizes, with enough different models tested to generalize to most situations found in the field. Because researchers using LGM rarely have access to large sample sizes, it is critical that simulation studies are conducted to investigate the effects of small and moderate sized samples.

Understanding the effects of sample size on model fit and parameter estimates will help researchers using latent growth models more fully explain longitudinal growth in their areas of study.

METHOD

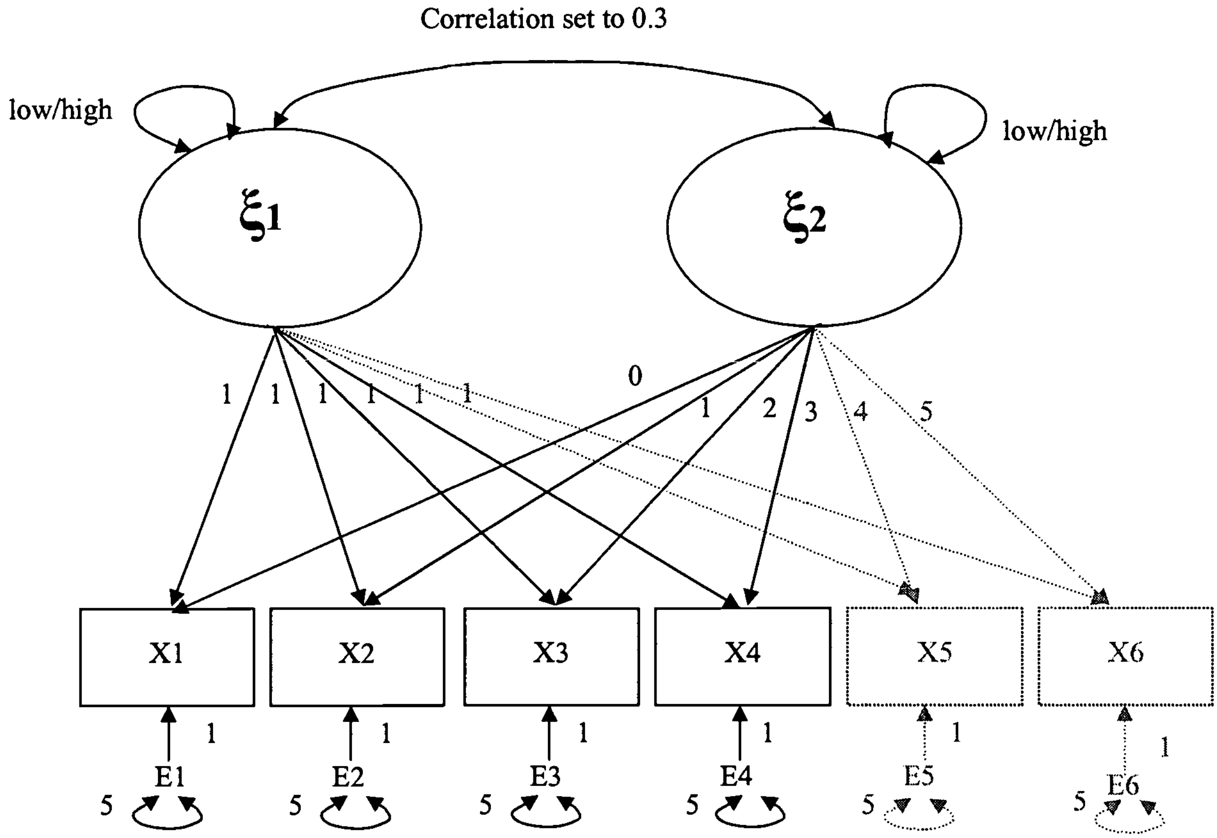
A Monte Carlo simulation approach was taken to investigate the effect of sample size on a variety of latent growth models. A fully balanced experimental design was implemented, with samples drawn from multivariate normal populations specified to represent 12 unique growth models. The 12 models varied factorially by crossing number of time points (three levels: 4, 5, 6), variance of intercept factor (two levels: high and low), and variance of slope factor (two levels: high and low). Table 1 provides a summary of the 12 models tested.

Table 1: Summary of model variations

Model No.	Time points	Variance Intercept	Variance Slope
1	4	High	High
2	4	High	Low
3	4	Low	High
4	4	Low	Low
5	5	High	High
6	5	High	Low
7	5	Low	High
8	5	Low	Low
9	6	High	High
10	6	High	Low
11	6	Low	High
12	6	Low	Low

The high and low variances for the intercept factor were set to 80 and 5, respectively, while the high and low variances for the slope factor were set to 16 and 1, respectively. This variance ratio of 5 to 1 is commonly found in the literature (Muthén & Muthén, 2002). Error variances were set to 5, and the correlation between the intercept and slope factors was constrained to be .3 for all models. This model is illustrated in Figure 1, following.

Figure 1: Latent growth model tested



The population covariance matrices for each of the 4 variances are provided in Table 2.

Table 2: Model-implied covariance matrices for each of the 4 model variance conditions

Variances: Intercept = 5 Slope = 1 (LL)

	X1	X2	X3	X4	X5	X6
X1	10.00					
X2	5.67	12.34				
X3	6.34	9.01	16.68			
X4	7.01	10.68	14.35	23.02		
X5	7.68	12.35	17.02	21.69	31.36	
X6	8.35	14.02	19.69	25.36	31.03	41.70

Variances: Intercept = 5 Slope = 16 (LH)

	X1	X2	X3	X4	X5	X6
X1	10.00					
X2	7.68	31.36				
X3	10.36	45.04	84.72			
X4	13.04	63.72	114.40	170.08		
X5	15.72	82.40	149.08	215.76	287.44	
X6	18.40	101.08	183.76	266.44	349.12	436.80

Variances: Intercept = 80 Slope = 1 (HL)

	X1	X2	X3	X4	X5	X6
X1	85.00					
X2	82.68	91.36				
X3	85.36	90.04	99.72			
X4	88.04	93.72	99.40	110.08		
X5	90.72	97.40	104.08	110.76	122.44	
X6	93.40	101.08	108.76	116.44	124.12	136.80

Variances: Intercept = 80 Slope = 16 (HH)

	X1	X2	X3	X4	X5	X6
X1	85.00					
X2	90.73	122.46				
X3	101.46	144.19	191.92			
X4	112.19	170.92	229.65	293.38		
X5	122.92	197.65	272.38	347.11	426.84	
X6	133.65	224.38	315.11	405.84	496.57	592.25

Each of the 12 model variations were tested at each of 6 sample sizes (n=25, 50, 100, 200, 500, 1,000), resulting in 72 (12x6) different cells. Within each condition, 1,000 replications were attempted. Then, additional replications were added until 1,000 properly converged replications per condition were obtained. Data generation and parameter estimation were conducted within EQS (Bentler, 1998) using maximum likelihood estimation. SAS (v8.01) was used to obtain the averages and variances of the parameter estimates and summary information about convergence.

Five general criteria were examined to assess performance of the 12 models at the 6 sample sizes. First, the convergence rate was examined to see if it appeared to be a function of sample size, or the other manipulated conditions. The number of condition and convergence codes were also examined to further describe the findings. Second, means for the model chi-square, comparative fit index (CFI), standardized root mean residual (SRMR), and root mean square error of approximation (RMSEA) were obtained, along with the variance of the chi-square values and the average lower and upper bounds for the 90 percent confidence interval of the RMSEA, as recommended by Hu and Bentler (1999). Third, the proportional bias and variability of each parameter, and the proportional bias of the standard deviations, were estimated.

RESULTS

Each of the 12 growth models were replicated at each of the sample sizes until 1,000 successful replications were obtained. The results from these replications are provided below. The results from our examination of the rate of successful completions are provided first, followed by the indices of model fit. Lastly, the results from the analyses of the empirical parameter estimates and standard deviations are provided.

Convergence Rates

All three manipulated conditions (sample size, number of time points, and relative size of the factor variances) influenced the rate of model convergence. Sample size had an effect on convergence rates, with the rates increasing as sample size increased. Generally, model convergence was quite good, even at low sample sizes. When n = 100 the average convergence across conditions was 99 percent, and it was still 96 percent, on average when sample size was 50. At the smallest sample size (n = 25) the convergence rate dropped down to 84 percent, on average (see Table 3).

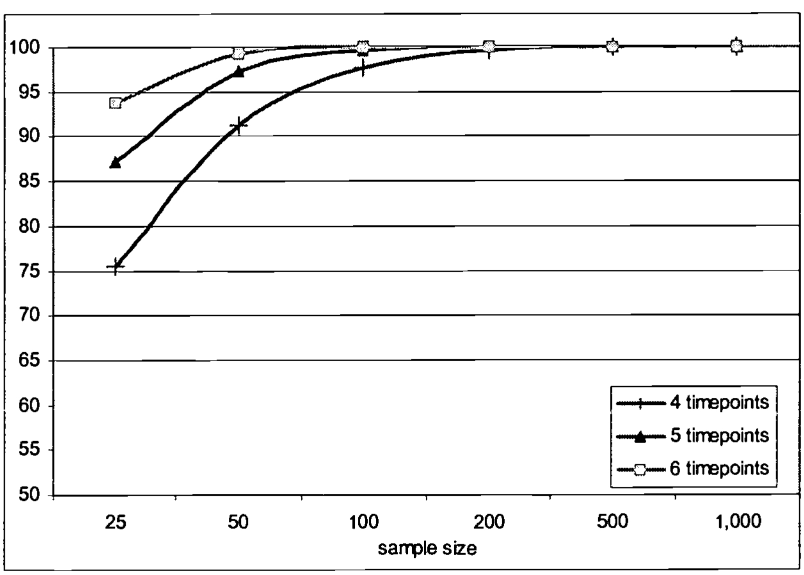
Table 3: Convergence rates

		Sample size					
		25	50	100	200	500	1,000
4 time	**LL***	84.8	97.1	100.0	100.0	100.0	100.0
	LH	74.2	89.6	96.9	99.5	100.0	100.0
	HL	81.7	96.8	99.9	100.0	100.0	100.0
	HH	61.6	81.2	93.7	99.3	100.0	100.0
5 time	**LL**	93.7	99.9	100.0	100.0	100.0	100.0
	LH	86.8	96.7	99.4	100.0	100.0	100.0
	HL	93.3	99.6	100.0	100.0	100.0	100.0
	HH	74.8	92.7	98.9	99.9	100.0	100.0
6 time	**LL**	97.1	99.8	100.0	100.0	100.0	100.0
	LH	95.4	99.4	99.9	100.0	100.0	100.0
	HL	96.7	99.9	100.0	100.0	100.0	100.0
	HH	86.2	98.1	100.0	100.0	100.0	100.0

*H= high variance, L=low variance. The intercept factor is the first, followed by the slope factor. This convention is followed throughout the paper.

The number of time points in the model also influenced the convergence rate, with more time points resulting in improved convergence. By adding one additional time point at the smallest sample size tested, the number of iterations necessary to obtain 1,000 successful completions dropped by 14 percent. By adding two time points, the number dropped by 21 percent (see Figure 2).

Figure 2: Average convergence rates by number of time points



Lastly, the relative size of the factor variances in the population had an effect on the rates of model convergence. When both the slope and intercept variances are high, the overall rate of convergence was at its lowest across all sample sizes. This is a result of the very high number of condition codes, 100 percent of which can be traced to improper estimates of the error variances. Conversely, when both variances are low (LL), the overall convergence rate is at its best. Of the failures that did occur, many were due to improper estimates of the intercept and slope variances. In fact, in the 4 time point model at the smallest sample size, the offending intercept and slope variance estimates accounted for the majority (57 percent) of all improper estimates.

In the two models where the slope variance is high (HH, LH), there were 36 percent more improper estimates than in the models where the slope variance is low (LL, HL). When the intercept of slope factor is manipulated to be low, it is the slope factor that incurs erroneous estimates. For example, in the 4 time points model with a sample size of 25, the number of improper estimates for the low slope variance (LL and HL) is 95 for each. However, when the slope variance is high in the same model (HH and LH), the number of improper estimates drops to zero. The same pattern holds across sample sizes (where improper estimates occur in adequate numbers) and across time points.

Table 4: Number of improper estimates

		Sample size					
		25	**50**	**100**	**200**	**500**	**1,000**
4 time	**LL**	182	30	0	0	0	0
	LH	356	116	32	5	0	0
	HL	226	34	1	0	0	0
	HH	638	230	68	7	0	0
5 time	**LL**	63	1	0	0	0	0
	LH	149	34	6	0	0	0
	HL	63	4	0	0	0	0
	HH	310	74	11	1	0	0
6 time	**LL**	30	2	0	0	0	0
	LH	46	6	1	0	0	0
	HL	29	1	0	0	0	0
	HH	139	19	0	0	0	0

Error variances with the greatest number of improper estimates (E4 and E5 when the slope variance is high), also produced the largest standard deviations. For example, the standard deviation of the fourth error variance was 19.4, in the 4 time point, HH model at the smallest sample size. On the opposite end of the spectrum, the first error variance for the LL model had a standard deviation of 5.5, which is 72 percent smaller.

Indices of Model Fit

For each cell of the design, data were collected for four fit indices. For the 1,000 fully converged replications, means for chi-square, comparative fit index (CFI), standardized root mean residual (SRMR), and root mean square error of approximation (RMSEA) were obtained, along with the variance of the chi-square values and the average lower and upper bounds for the 90 percent confidence interval of the RMSEA. Proportional bias for the mean and variance of the chi-square values was computed with the equation:

$$\text{pbias} = (\text{observed} - \text{expected}) / \text{expected} ,$$

with

$$E(\chi^2) = df$$

for computing the proportional bias of the mean chi-square value and

$$E[Var(\chi^2)] = 2df$$

for computing the proportional bias of the variance of the chi-square values. The results for the other fit indices are explicated in terms of their average values. Because the true model was fit to the data in every replication, the fit of the models was expected to be excellent.

Proportional bias in chi-square values

The bias in the chi-square values was generally positive; among the cells with negative biases none had a bias greater in magnitude than 5 percent. At n = 25, only one cell had less than 5 percent bias, with all of the cells in the six time-point condition having bias in excess of 10 percent. Within each combination of intercept and slope variance, the bias at n = 25 increased as the number of time points in the model increased. Except for the combination of HH with four time points, bias decreased in all conditions and time points when sample size was increased from 25 to 50. Without exception, increasing *n* from 50 to 100 decreased bias in all combinations. With n = 200, the only bias greater than 5 percent was that for LL with four time points (.0527). Bias decreased in the LL combination as the number of measured time points increased. No other cell with $n \geq 200$ had a proportional bias above 5 percent.

At n = 25, the bias of the variance of chi-square increased substantially within each combination of intercept and slope variance as the number of time points in the model increased, reaching as high as 34.85 percent in the LL combination with six time points. There otherwise appeared to be no consistent patterns in the proportional bias of the variance as a function of any of the model conditions. Tables 5 and 6 provide the proportional bias for the mean and variance of the chi-square values.

Table 5: Model chi-square values

		Sample size					
		25	50	100	200	500	1,000
4 time	LL	3.17	3.11	3.07	3.16	3.03	2.86
	LH	3.20	3.05	3.04	3.02	3.08	3.12
	HL	3.24	3.19	2.99	2.95	3.05	3.02
	HH	2.95	3.23	3.04	3.08	3.03	2.99
5 time	LL	7.57	7.44	7.11	7.22	6.84	6.90
	LH	7.83	7.60	7.28	7.11	6.87	7.02
	HL	7.65	7.52	7.20	6.79	6.97	6.84
	HH	7.47	7.44	7.13	7.06	7.23	6.89
6 time	LL	13.81	12.67	12.14	12.28	11.91	12.40
	LH	13.58	13.03	12.42	12.31	12.15	12.07
	HL	13.51	12.93	12.38	11.97	12.15	12.03
	HH	13.31	12.84	12.45	12.39	12.12	11.93

Table 6: Proportional bias in model chi-square values

		Sample size					
		25	50	100	200	500	1,000
4 time	LL	0.0575	0.0352	0.0226	0.0527	0.0084	-0.0462
	LH	0.0674	0.0172	0.0136	0.0069	0.0256	0.0384
	HL	0.0787	0.0643	-0.0040	-0.015	0.0181	0.0055
	HH	-0.0160	0.0769	0.0147	0.0253	0.0097	-0.0047
5 time	LL	0.0816	0.0631	0.0161	0.0311	-0.0229	-0.0147
	LH	0.1181	0.0853	0.0403	0.0161	-0.0182	0.0023
	HL	0.0925	0.0738	0.0287	-0.0300	-0.0048	-0.0229
	HH	0.0676	0.0625	0.0189	0.0090	0.0323	-0.0163
6 time	LL	0.1509	0.0556	0.0113	0.0235	-0.0074	0.0335
	LH	0.1317	0.0856	0.0350	0.0260	0.0123	0.0058
	HL	0.1256	0.0773	0.0313	-0.0020	0.0123	0.0025
	HH	0.1093	0.0698	0.0375	0.0324	0.0099	-0.0056

Comparative Fit Index (CFI)

At n = 500 and n = 1,000, the mean CFI to three decimal places was at least .999, so those cells will be ignored when discussing the behavior of CFI. The mean CFI increased as *n* increased. Across time points, mean CFI did not seem to behave in a consistent pattern. Within each time point and within each sample size, the average value of the CFI increased in the following order of intercept and slope variance combinations: LL; LH; HL; HH. Of note, however, is the fact that only three cells (LL with n = 25) had a mean CFI below .98 (range: .9722 – .9750). The next lowest mean CFI was .9881.

Standardized Root Mean Residual (SRMR)

In all models, the mean value of the SRMR decreased as *n* increased. Mean SRMR also decreased across the same sequence of intercept and slope variance that mean CFI increased: LL; LH; HL; HH. At n = 25, the intercept and slope variance combination of HH had an average SRMR below .05. Among the remaining eight cells at n = 25, only 4HL had a mean SRMR below .05. At n = 50, the mean SRMR was consistently above .05 in the three LL cells, but 5 LH was the only other cell above .05 (.0501). Only 5 LL and 6 LL had a mean SRMR above .05 when n = 100 (.0520 and .0540, respectively), and all cells with n ≥ 200 had mean SRMR values below .04.

The behavior of the mean SRMR as a function of time points was a bit more involved. With low slope variance, mean SRMR increased as the number of time points increased. In the models that had high slope variance, the mean SRMR consistently increased when going from four to four time points; from five time points to six, the mean SRMR remained steady or decreased slightly. This pattern in the high slope variance models, however, involved only slight differences in the mean SRMR, with the largest difference between consecutive time points being .0037.

<u>Root Mean Square Error of Approximation (RMSEA)</u>

The average value of the RMSEA decreased as sample size increased for all models tested. The average value generally decreased as the number of time points increased, somewhat less consistently at the two smallest sample sizes tested. At n = 25, no average was below .05. At n = 50, all of the averages were between .04 and .05. For the 12 cells with n = 100, the effect of number of time points was quite evident: Every mean RMSEA with four time points (range: .0313 – .0327) was greater than every mean RMSEA with five time points (range: .0280 – .0295), and all of those were greater than the mean RMSEA for the cells with six time points (range: .0250 – .0274). At n = 200, no mean RMSEA exceeded .025, at n = 500, the largest average RMSEA was .0146 (4 LH), and only one average RMSEA with n = 1,000 exceeded .01 (4 LH: .0104).

The average lower bound of the 90 percent confidence interval for the RMSEA decreased as *n* increased. At n = 25, two averages were above .007 and the lowest was .004996 (4 LH). At n = 200, the highest mean lower bound was all of .0021 (4 LL), and at n = 1,000, no average lower bound was above .0009.

The relationship between the average lower bound of the RMSEA and number of measured time points involved an interaction with sample size and combination of intercept and slope variance. With n = 25, the average lower bound increased as the number of time points increased. With n = 50, the average lower bound increased from four time points to five, and then decreased from five time points to six, except for the intercept and slope variance combination of HH, where the average lower bound first decreased then increased from four to five time points and five to six time points, respectively. At the other four sample sizes, the average lower bound of the RMSEA generally decreased as the number of time points increased, but for each sample size, there was always one combination of intercept and slope variance that did not follow this pattern.

The average upper bound of the 90 percent confidence interval of the RMSEA also decreased as *n* increased. There was also a very clear effect of number of time points. Within each sample size, the four cells with four measured time points had average upper bounds greater than those for the four cells with five measured time points, which were greater than those for the four cells with six time points. At the smallest sample size tested, the average upper bound was never less than .20. The average upper bound did not drop below .10 until n = 100 with six time points. At n = 500 and five time points, only the HH combination had an average upper bound greater than .05. All four cells with n = 500 and six time points had average upper bounds below .05, as did every cell with n = 1,000.

Parameter Estimates

The cornerstone of latent growth modeling is the estimation of group-level parameters that define the starting point and growth trajectory, as well as the estimation of individual variability around these group parameters. In order to determine the effects of various sample sizes on these estimates, several statistics were calculated. First, the proportional bias of each parameter was calculated using the population parameter values and the parameter estimate averaged over the first 1,000 successful replications of each model. To compute the bias, the population value was subtracted from the estimate and divided by the population value.

$$\text{Bias} = (\text{empirical - pop})/\text{pop}$$

Second, the variability of the parameter estimates around their true population value were computed by dividing the standard error of the empirical estimate from the first 1,000 successful replications by the true population value. Lastly, the standard error bias was computed in the same way as the parameter bias, using the standard errors instead of the parameter values.

Proportional bias of parameter estimates

While the size of the sample did not bias the average parameter estimates, it does play a role in the range of bias estimates. Overall, biases in the estimated parameters were very small, with an average of less than 1 percent. Of the three parameters estimated, the covariance was the most likely of the three to have biased estimates. The largest parameter bias in the study was an underestimation of the covariance parameter by 12.65 percent at n = 25, with both factor variances set to low.

While the average bias by sample size is close to zero, the range of bias values is also of interest (see Figures 3 - 5). These figures show that as sample size increases, fluctuations in biased parameters diminish. They also demonstrate that the covariance parameter produced the largest range in bias at all sample sizes, which is most pronounced at the smallest sample size.

Figure 3: Bias in estimates of intercept variance – range and average

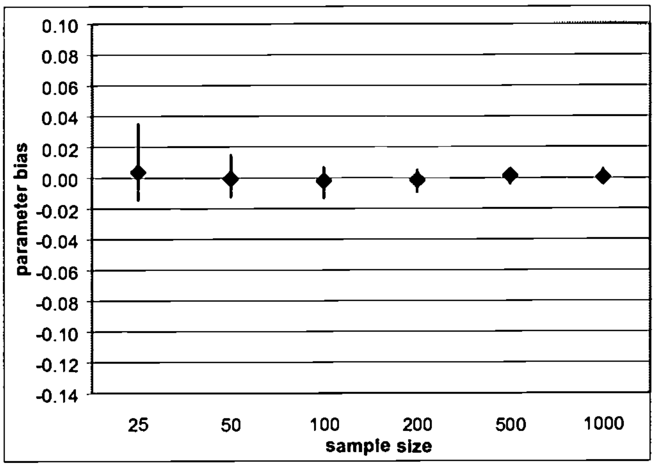


Figure 4: Bias in estimates of slope variance – range and average

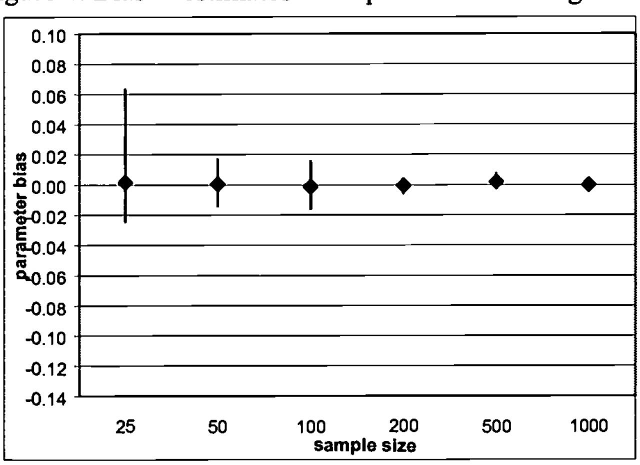
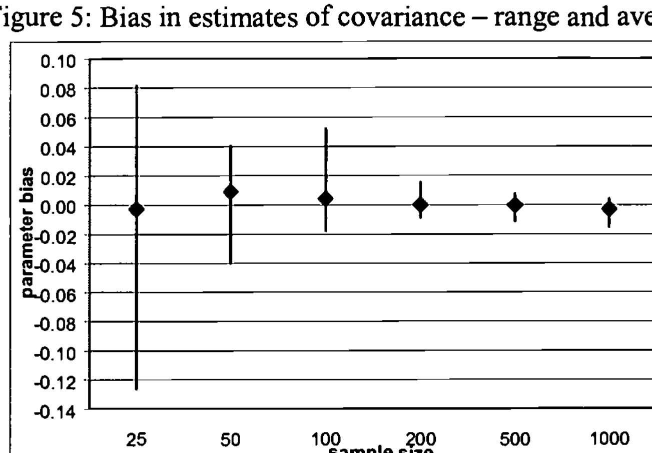


Figure 5: Bias in estimates of covariance – range and average



In terms of the four factor variance conditions, the models with low variances for both factors (LL) tended to produce parameter estimates with the largest proportional bias. Averaged across sample size and time points, the low variance model over-estimated both factor variance parameters and under-estimated the covariance parameter. All of the other conditions overestimated the covariance parameter.

Proportional variability

Proportional variability provides estimates of the degree to which the empirical parameter estimates fluctuate around the true population values. Values closer to zero are desirable because they indicate that the empirical estimates are more precise. Results indicate that the parameter estimates become less accurate at lower sample sizes, as can be seen in Figure 6, below. That is, larger variation around the true values occurred with smaller sample sizes -- each time the sample size doubled, this variability decreased by approximately 30 percent.

Figure 6: Proportional variability by sample size

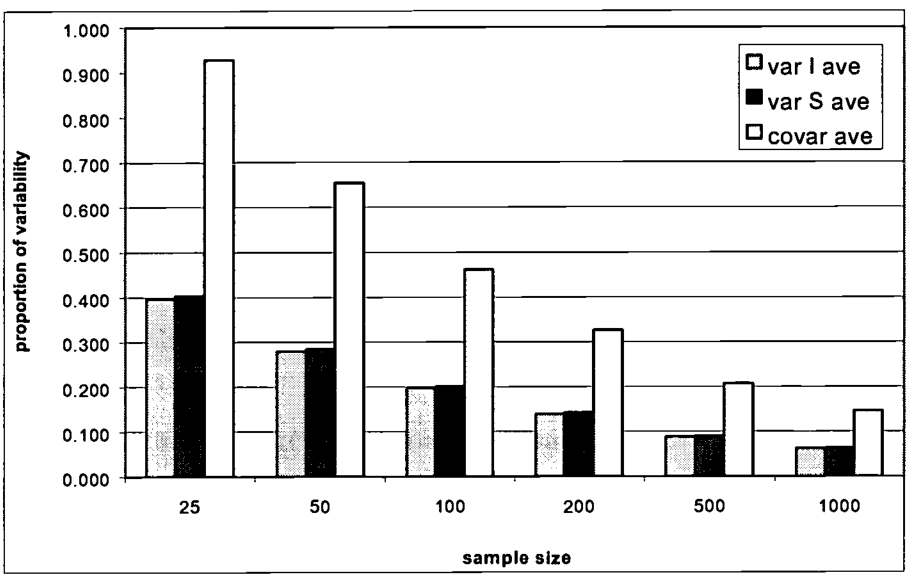


Figure 6 also illustrates the fact that the proportional variance of the covariance parameter, while decreasing with larger sample sizes, remains substantially larger than the other two parameters. Averaged across all sample sizes and number of time points, the fluctuations around the true covariance is greater when both intercept and slope variances are low (59.4 percent) than when both variances are high (35.3 percent).

While sample size affects the variability of the parameter estimates, it was found that the relative size of the factor variances also play a role. When the factor variances are low, the variability around the true value for that factor is high. Taking the slope variance (where n = 25 with 4 time points), as an example, when the variance of this parameter is low (the LL and HL conditions), the empirical estimate varies by about 70 percent. When the slope estimates are high (the HH and LH conditions), this variability drops to about 30 percent. The proportion of variability decreases as sample size increases, but the pattern holds across all sample sizes.

The accuracy of the covariance parameter is related to the variance size of both factors. When both factors have low variances (LL), the proportional variability of the covariance parameter is at its highest. Conversely, when both variances are high (HH), the estimates vary the least. The remaining two conditions (LH and HL) share similar degrees of variability and are within the two extremes.

Lastly, the number of time points used in the model did not appear to affect the variability of the empirical parameter estimates around the true population value. The improvement from 4 to 6 time points was negligible for the intercept and slope variance estimates (0.4 and 0.6 percent, respectively), while precision of the covariance estimate worsened by 0.8 percent.

<u>Proportional bias of standard errors</u>

The means of the asymptotic standard errors are very similar to empirical standard deviations for all models, resulting in low bias across all conditions. Very little bias in the standard deviations of the intercept variance parameter was found--the largest was 3.7 percent.

Sample size and the number of time points in the model did not appear to bias the standard errors, but the relative variance of the factors did influence the standard deviations of the slope parameter. There was some decrease in the proportion of bias as the sample size increased from 25 to 50, but after that, the bias estimates fluctuated around zero. However, the standard errors of the slope parameter showed substantial bias when both the slope and intercept variances were high. The bias remained steady at around 80 percent regardless of sample size and the number of time points. In comparison, the other models (LL, LH, HL) for this parameter produced biases from 0 to 4 percent (see Table 6). Similarly, proportion of bias in the covariance parameters standard deviation was much higher for the HH model than in the remaining models. The standard deviations were biased by approximately 67 percent, compared to the other models, whose biases did not exceed 3 percent (see Table 7).

Table 6: Proportional bias of standard deviations of slope variance parameters

		Sample size					
		25	**50**	**100**	**200**	**500**	**1,000**
ti	**LL**	-0.008	-0.010	-0.003	-0.002	0.000	0.002
	LH	-0.023	-0.007	-0.008	0.001	0.000	-0.001

	HL	-0.007	-0.016	-0.015	-0.002	-0.002	-0.005
	HH	-0.800	-0.799	-0.796	-0.797	-0.796	-0.796
5 time	**LL**	-0.031	-0.014	-0.009	-0.003	-0.004	0.001
	LH	-0.016	-0.016	-0.004	0.000	0.000	0.000
	HL	-0.031	-0.008	-0.002	-0.003	0.002	-0.007
	HH	-0.803	-0.802	-0.802	-0.802	-0.801	-0.801
6 time	**LL**	-0.013	-0.005	-0.016	-0.002	-0.003	-0.010
	LH	0.002	0.001	-0.009	0.004	0.003	0.002
	HL	-0.033	0.001	0.003	0.000	0.001	-0.006
	HH	-0.804	-0.805	-0.803	-0.803	-0.803	-0.803

Table 7: Proportional bias of standard deviations of intercept-slope covariance estimates

		Sample size					
		25	**50**	**100**	**200**	**500**	**1,000**
4 time	**LL**	-0.013	-0.010	-0.005	-0.003	-0.003	-0.003
	LH	-0.015	-0.004	-0.008	0.001	0.000	0.000
	HL	-0.005	-0.015	-0.008	-0.003	0.001	-0.001
	HH	-0.674	-0.675	-0.671	-0.672	-0.671	-0.670
5 time	**LL**	-0.029	-0.015	-0.009	-0.005	0.001	0.001
	LH	-0.022	-0.017	-0.010	-0.005	0.000	-0.001
	HL	-0.018	-0.004	-0.740	-0.004	0.001	-0.001
	HH	-0.678	-0.673	-0.674	-0.674	-0.673	-0.674
6 time	**LL**	-0.028	-0.008	-0.011	-0.005	-0.001	-0.001
	LH	-0.014	-0.009	-0.010	0.002	0.001	0.002
	HL	-0.027	-0.002	-0.003	-0.002	0.000	0.000
	HH	-0.678	-0.678	-0.674	-0.675	-0.674	-0.674

DISCUSSION

The goal of the present paper was to examine the effect of sample size on a variety of latent growth models. Specifically the effect of sample size on the rate of model convergence, model fit, and the estimation of model parameters and their standard deviations was explored. A discussion of the impact of sample size on each of these three areas is provided below.

Convergence Rates

In this section we examine the contribution of sample size, number of time points, and relative size of the slope and intercept variances to the incidence of failure due to either non-convergence or improper solutions. However, improper solutions will be the main focus, because they account for the majority (85 to 100 percent) of all failures.

The failure of a model to converge to a fully proper solution can be traced to one of two root causes; either the model did not converge within the maximum allowed number of iterations, or the model produced improper estimates that were then fixed to a boundary value (condition codes). Improper estimates take on values that would be impossible for the corresponding parameters, such as a correlation greater than one or a variance that is negative.

As Van Driel wrote in 1978, improper variance estimates result from one of three causes: (1) sampling fluctuation, (2) model misspecification, or (3) under-identification of the model. Because our models were correctly specified with known properties, options 2 and 3 may be ruled out, leaving only sampling fluctuation to account for the improper estimates. In our models, four conditions contribute to these sampling fluctuations; the size of the sample, population variance size, the standard deviations of the parameter estimates, and the number of time points in the model. Each of these will be discussed in turn.

As expected, all models demonstrated better successful completion rates as sample size increased. This is consistent with past studies that suggest an inverse relationship between sample size and frequency of failure (Boomsma, 1985; Chen et al., 2001). Because small samples have greater fluctuations in sampling, which leads to a greater likelihood that some of the fluctuations will result in impossible parameter estimates. Therefore, the larger the sample, the more stable the estimates of the standard errors.

The size of the factor variances in the population also impacted the convergence rates in our models. The closer the size of the variance in the population is to zero, the greater the probability of the sampling fluctuating below zero. Boomsma (1983, 1985) found that improper solutions occurred more frequently in models with smaller population values of error variances. Our results show that 86.6 percent of all improper estimates at $n = 25$ were due to Heywood cases. While the error variances in our population were set to 5, it is possible that if they had been larger, fewer Heywood cases would have occurred. This leads to the conundrum that, in models where the latent factors do an excellent job of accounting for the variance in the manifest variables, the error variances will be small, resulting in more Heywood cases than another model whose factors did a poor job of accounting for the variances in the observed variables.

In addition to the error variances, the sizes of the factor variances in the population also influence the convergence rate. Our evidence suggests that improper solutions occur more frequently to factors that have low population variances. Furthermore, it is the relative size of the slope variance that drives the overall number of improper solutions. This may be due to the relationship between the factor variances and the estimated variances of the measured variables. The slope variance in later time points exerts more of an influence on the later time points because the squared loadings are larger than the earlier loadings. Therefore, the error variances exert a lesser influence for these measured variables at later time points. Our findings support this theory, with the error variances for the later time points generally producing the largest number of improper estimates.

The standard deviations associated with the estimated variances also influenced the convergence rates, with higher standard deviations resulting in greater numbers of improper solutions. Greater standard deviations suggest a wider dispersion of estimates and hence a greater probability of a nonpositive estimate.

Lastly, our evidence suggests that the addition of more time points to growth models result in increased rates of successful completions. Our finding echoes previous Monte Carlo studies in Structural Equation Modeling that have found that by increasing the number of variables per factor, the success rate is improved (Boomsma, 1983, 1985; Anderson and Gerbing, 1984).

In sum, sample size was found to influence the convergence rates of the models, with larger samples resulting in fewer improper estimates and failures. However, this effect is lessened if the variances of the slope and intercept factors are low. It is also lessened when more time points are added to the model. This research reinforces previous findings on the importance of sample size used in latent models. For example, Marsh et. al. (1998) found that having more variables per latent factor compensates for smaller sample sizes.

Indices of Model Fit

The present study also investigated the effects of growth model conditions on the behavior of the following commonly used fit indices: χ^2; CFI; RMSEA, including 90 percent CI boundaries; and SRMR. The CFI, RMSEA, and SRMR all indicated improved model fit as *n* increased, the average CFI value was never less than .97, so improvements as a function of sample size were merely from great to fantastic.

Chi-square demonstrated improved model fit from smaller sample sizes to the moderate sample sizes, but from the moderate sample sizes into the larger sample sizes tested, the bias of the chi-square values either became negative or fluctuated in a manner that did not seem related to any of the model conditions varied in the present study. Additional relationships between chi-square bias and model facets were either complex beyond recognition or perhaps simply not present.

Values of CFI and SRMR indicated demonstrated improved model fit as the variance of the intercepts and slopes increased in the order: low/low (for intercept variance and slope variance, respectively); low/high; high/low; and high/high. The relationship between SRMR and number of measured time points interacted with the magnitude of the intercept and slope variances, as did the relationship between RMSEA and number of time points. The upper bound of the 90 percent CI for the RMSEA, however, had a very clear and potentially important relationship with time points, consistently decreasing as the number of measured time points increased.

Parameter Estimates

The degree to which sample size can bias parameter estimates is of interest to researchers using latent growth models. Our findings suggest that, on average, sample size does not bias the parameter estimates to a substantive degree. However, although the average bias is negligible, at low sample sizes, there was substantial fluctuations in the size of the bias. As sample size increased, these fluctuations in biased parameters diminished. This is consistent with other research findings (Burchinal, 1989).

It is also important to understand the degree to which small sample size may bias the standard deviations of parameter estimates, because an overestimation can mean that significant effects are missed. Similarly, if they are underestimated, significant effects may be overstated. However, in our study, sample size did not appear to bias the standard deviations. The number of time points also had no effect. However, the relative size of the factor variances were important in this regard, with the HH condition negatively biasing the slope standard deviation by 80 percent and the covariance standard deviations by 67 percent, across all sample sizes and time points.

The asymptotic theory underlying SEM implies that random fluctuations in the variance of parameter estimates will decrease as sample size increases. In fact, both sample size and relative size of the factor variances were found to influence the proportional variability of parameter estimates, while the number of time points in the model did not. By increasing the sample size, estimation errors are reduced because the accuracy of the parameter estimates are increased. That is, estimates of variances are more likely to reflect the population value when they are based on a large number of observations. This finding is consistent with Gerbing and Anderson (1985) who, in their confirmatory factor analysis Monte Carlo study, found that sample size had the largest effect on the variability in parameter estimates.

The relative size of the each factor variance was found to be inversely proportional to the variability of its own variance parameter. For example, when the population slope factor variance was low, the proportional variability for the slope parameter was high, while the proportional variability for the intercept factor variance was not affected. The fact that the opposite factor does not much influence the proportional variability of the factor of interest is understandable, because the correlation between the two factors was set to be quite low (r = 0.3).

The fluctuations in the covariance parameter around the population variable was substantially greater than the variance parameters for the intercept and slope factors. This may be because the degree to which the intercept and slope factors covary is influenced by the variance sizes of both of these factors, while the factor variances are not influenced to a great extent by the other factor variance.

In conclusion, results from this study support the contention that latent growth models are meaningfully impacted by the size of the sample. Sample size was found to influence the convergence rates, measures of model fit, and the proportional variability of the parameter estimates. It did not, however, result in biased parameter estimates or biased standard deviations. A summary of the effects of each of the manipulated conditions is found below, in Table 8.

Table 8: Summary of key findings

	Convergence	**Model fit**	**Parameter bias**	**Standard error bias**	**Proportional variability**
Sample size	Improves with larger samples	Improves with larger samples	Not affected	Not affected	Decreases with larger samples
Time points	Improves with more time points	RMSEA upper bound decreases with more time points	Not affected	Not affected	Not affected
Factor variances	Improves when variances are low	Improves when variances are high	Low variances increase bias	High variances negatively bias slope & covariance	Low variances increase fluctuations for that parameter

Recommendations

Based on the results of this simulation, a number of recommendations for researchers hoping to utilize latent growth models are made. Sample sizes of at least 50 should be used, in order to obtain model convergence. Smaller samples may be used when additional time points are present and when the variances of the slope and intercept factors are expected to be low. However, it should be noted that the models tested in this study were all correctly specified and misspecified models may require larger samples. Model fit also improves with sample size, with n = 25 providing generally poor fit across all indices. However, by adding more time points to the model, the upper bound of the RMSEA decreases, improving the likelihood that both upper and lower bounds will fall below the .05 level. Biased parameters, on average, are not likely to result from the use of small samples, but the range of parameter estimates will fluctuate substantially more with small samples. In this regard, samples of at least 100 are therefore recommended.

Limitations and Future Directions

An inherent limitation in Monte Carlo studies is the representativeness of the models tested. It is difficult to know whether the results from one set of models will transfer to other models. A related issue is that we have examined data simulated from a multivariate normal distribution. Maximum likelihood estimation makes the assumption of multivariate normality. Violations of these assumptions have been shown to be problematic (Muthén and Kaplan, 1992). It is unclear how well the findings from this study would generalize to nonnormal distributions. Other future research in this area could include adding a mean structure to the growth models and manipulating the growth trajectory, so that some models are non-linear.

REFERENCES

Anderson, J., & Gerbing, D. (1984). The effects of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49*, 155-73.

Bentler, P. (1995). *EQS Structural Equations Program Manual.* Encino, CA: Multivariate Software, Inc.

Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality*. Amsterdam, the Netherlands: Sociometric Research Foundation.

Boomsma, A. (1985). Non-convergence, improper solutions and starting values in LISREL maximum likelihood estimation. *Psychometrika, 50*, 229-42.

Burchinal, M. (1989). Comparison of models for estimating individual growth curves. Paper presented at the biennial meeting of the Society for Research in Child Development. Kansas City, MO, April 27-30, 1989.

Chen, F., Bollen, K., Paxton, P., Curran, P., & Kirby, J. (2001). Improper solutions in structural equation models. *Sociological Methods & Research. 29*(4), 468-508.

Gagné, P., & Hancock, G. (2002). *Relation of sample size and solution propriety in latent variable models as a function of construct reliability.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. April, 2002.

Gerbing, D., & Anderson, J. (1985). The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. *Multivariate Behavioral Research, 20*, 255-271.

Hu, L., & Bentler, P.M. (1999. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling. 6,*(1) 1-55.

MacCallum, R. & Roznowski, M. & Necowitz, L. (1992) Model mispecifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin. 111*, 490-504.

Marsh, H., Hau, K., Balla, J., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*, 181-220.

Muthén, B., & Kaplan, D. (1992) A comparison of some methodologies for the factor analysis on non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology, 45*, 19-30.

Muthén, B., & Muthén, L. (2002). How to use a Monte Carlo study to decide on sample size and determine power. Preliminary version of a paper scheduled to appear in *Structural Equation Modeling.*

Tanaka, J. (1987). “How big is big?”: Sample size and goodness-of-fit in structural equation modeling with latent variables. *Child Development. 58*, 134-146.

Van Driel, O. (1978). On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika 43*, 225-43.

APPENDIX – EXAMPLE OF EQS CODE

```
/title
        Simulation for model: 6 time points, HH n = 500, data is generated from the model
/specifications
        cas = 500; var = 6; me = ml;
/equations
        v1 = f1        + e1;
        v2 = f1 +   f2 + e2;
        v3 = f1 + 2f2 + e3;
        v4 = f1 + 3f2 + e4;
        v5 = f1 + 4f2 + e5;
        v6 = f1 + 5f2 + e6;
/variance
        f1 = 80.0*;
        f2 = 16.0*;
        e1 = 5.0*;
        e2 = 5.0*;
        e3 = 5.0*;
        e4 = 5.0*;
        e5 = 5.0*;
        e6 = 5.0*;
/covariance
        f1,f2 = 10.73*;
/output
        data = '6HH500a.rst'; se; pa;
/simulation
        seed = 500541;
        replication = 1000; population = model;
/end
```

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC
Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

TM034956

I. DOCUMENT IDENTIFICATION:

Title: The Effect of Sample Size on Latent Growth Models	
Author(s): Jennifer Hamilton, Phillip Gagné, Gregory Hancock	
Corporate Source: University of Maryland - College Park	Publication Date: April 24, 2003

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 [X]	Level 2A [ ]	Level 2B [ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Jennifer Hamilton	Printed Name/Position/Title: JENNIFER HAMILTON	
Organization/Address: University of Maryland - College Park	Telephone: 301-738-3628	FAX:
	E-Mail Address: jenniferhamilton@westat.com	Date: 5/19/03

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse: **University of Maryland**
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions