Psychological Methods
2003, Vol. 8, No. 3, 384–393

1082-989X/03/$12.00 DOI: 10.1037/1082-989X.8.3.384

Overextraction of Latent Trajectory Classes: Much Ado About Nothing? Reply to Rindskopf (2003), Muthén (2003), and Cudeck and Henly (2003)

Daniel J. Bauer
North Carolina State University

Patrick J. Curran
University of North Carolina at Chapel Hill

The comments on D. J. Bauer and P. J. Curran (2003) share 2 common themes. The 1st theme is that model-checking procedures may be capable of distinguishing between mixtures of normal and homogeneous nonnormal distributions. Although useful for assessing model quality, it is argued here that currently available procedures may not always help discern between these 2 possibilities. The 2nd theme is that even if these 2 possibilities cannot be distinguished, a growth mixture model may still provide useful insights into the data. It is argued here that whereas this may be true for the scientific goals of description and prediction, the acceptance of a model that fundamentally misrepresents the underlying data structure may be less useful in pursuit of the goal of explanation.

We begin by thanking Robert Cudeck, Susan Henly, Bengt Muthén, and David Rindskopf for providing comments on our work (Bauer & Curran, 2003). We could not have asked for a more talented and esteemed group of quantitative methodologists to comment on our article, and we greatly appreciate the time and effort they have invested in this endeavor. Their comments have raised several interesting and important topics that relate to specific aspects of our article, to growth mixture modeling in particular, and to the goals of science and applied research more generally.

We have organized our reply around what we view as two primary themes shared by the comments. The first theme, most prominent in the comments of B. Muthén (2003) and Rindskopf (2003) is that it may not be as difficult as we suggest to distinguish between a single-class model with nonnormally distributed observed variables and a true mixture. The second theme reflected to varying degrees in each of the three commentaries is that, when all is said and done, we should perhaps not care so much that these two possibilities might be difficult to distinguish in practice. This theme suggests that different models should be regarded as alternative representations of reality, none necessarily true in any real sense, but some perhaps more useful than others. From this perspective, whether the latent classes correspond to true groups or not is an irresolvable issue, outside of the artificial context of mathematical equations and Monte Carlo simulations. Subsequently, our claim that classes will likely be found when no such classes truly exist is really much ado about nothing. We find both of these points to be quite intriguing and explore each in turn.

Daniel J. Bauer, Department of Psychology, North Carolina State University; Patrick J. Curran, Department of Psychology, University of North Carolina at Chapel Hill.

This work was funded in part by National Institute on Drug Abuse (NIDA) Fellowship DA06062 awarded to Daniel J. Bauer and NIDA Grant DA13148 awarded to Patrick J. Curran. We thank Kenneth Bollen, Robert MacCallum, Andrea Hussong, and the members of the Carolina Structural Equations Modeling Group for their valuable input on this reply.

Correspondence concerning this article should be addressed to Daniel J. Bauer, Department of Psychology, College of Humanities and Social Sciences, North Carolina State University, Raleigh, North Carolina 27695-7801. E-mail: dan_bauer@ncsu.edu

Distinguishing Between Mixtures and Nonmixtures

Both Rindskopf (2003) and B. Muthén (2003) suggested that whereas in some instances it may be difficult to distinguish between data that are simply nonnormal and data generated by a mixture, in other cases it may not be. Rindskopf (2003) stated, "Had they [Bauer and Curran] simulated the distributions implied by the mixtures 'detected' by the software they used, they might have seen that the distribution did not look like that of the data" (p. 366). B. Muthén (2003) made a similar point, remarking that one can test whether the "skewness and kurtosis estimated by the model fit the corresponding sample quantities" (p. 371). Both of these authors suggested that such model-checking procedures might show that the estimated mixture really does not approximate the nonnormal distribution of the observed variables well, and this could be cause for rejecting the mixture hypothesis when the population is homogeneous but the data are nonnormal. This point was made most forcefully by B. Muthén (2003), who asserted that "using this new [skew and kurtosis] test, BC [Bauer and Curran] would not have made the faulty conclusions that were of concern" (p. 372).

To provide a bit of perspective, at the time our manuscript was written and finalized (October 2001 to August 2002), best practice for determining the number of components in a growth mixture model, illustrated in the work of Muthén and colleagues (e.g., B. Muthén, 2001a, 2001b; B. Muthén & Muthén, 2000; B. Muthén & Shedden, 1999) consisted of determining the minimum Bayesian information criterion (BIC) score over growth models with different numbers of components (typically without predictors). Rather than restrict our comparison to the BIC and similar criterion, we also examined several other fit measures not routinely considered in the estimation of growth mixture models but which are sometimes used in the more general literature on finite mixture models. Every measure we considered indicated that the two-class model was superior to the one-class model, just as we had predicted on the basis of the underlying statistical model.

B. Muthén (2003) argued that the likelihood ratio test of Lo, Mendell, and Rubin (2001) is preferable to these indices, suggesting that it may be robust to minor departures from normality within components. However, a challenge arises when applying this test in growth mixture models. Specifically, growth mixture models are often based on mixtures of normals characterized by different covariance matrices (i.e., where either the growth factor covariance matrix or time-specific residual variances differ over classes, not just the growth factor means). Closer examination of the analytics underlying the Lo et al. (2001) test shows that this test can indeed be used for testing the number of components in a normal mixture involving heterogeneous covariance matrices, provided that proper constraints are imposed on the covariance matrices to assure that the likelihood function is bounded (Y. Lo, personal communication, May 27, 2003). However, such constraints are not routinely used in growth mixture model analyses. Further, no published analytical or Monte Carlo studies currently exist to establish the robustness or appropriate use of this test with growth mixture models with or without violations of distributional assumptions. Given this, we believe that although the application and possible extension of the Lo et al. test is promising, it is premature to broadly implement this test with growth mixture models, particularly when the classes are characterized by heterogeneous covariance matrices. To briefly examine this issue empirically, we applied the Lo et al. test to a random subsample of replications from our simulation study. Consistent with the other fit criteria we considered, the new test supported rejection of the one-class model in favor of the two-class model for each nonnormal condition. Although these results are intriguing, clearly much additional research on this new test is needed.

B. Muthén (2003) also asserted that the clear superiority in fit of the two-class model in our simulation study was an outcome of the "high degree of nonnormality" (p. 371) and "strongly nonnormal data" (p. 369) that we considered. We view our choice of distributions in a somewhat different light. We selected these specific distributions to represent what we considered to be small and modest departures from normality. In our minor nonnormal condition, the data were generated to be characterized by a univariate skew and kurtosis of 1.0 and 1.0, respectively. We chose these values given that they are well within the range commonly encountered in psychological research (Micceri, 1989) and would likely not be cause for concern in many empirical studies in the behavioral sciences. However, in this condition the two-class model was deemed superior to the true one-class model almost 100% of the time. The fact that multiple classes may be fit to trivially nonnormal data is further supported by B. Muthén's own results, which demonstrate that a multiclass solution may appear optimal even for data with univariate skew and kurtosis values all well below 1.0. This supports our very point: Nonnormality is requisite to fit multiclass models (under proper specification), and even minor departures from normality may be sufficient to extract more than one latent class. We thus do not believe that our results can be attributed to an artifact stemming from the study of severely nonnormal and unrealistic distributions; instead, we believe these results generalize to many empirical conditions encountered in applied behavioral research.

As both Rindskopf (2003) and B. Muthén (2003) nicely pointed out, we did concern ourselves only with *comparative* fit assessments and, whereas these all favored two classes, the two-class model might still have been rejected by a test of *absolute* fit. At issue is whether a fitted normal mixture model can really approximate nonnormal data well even if it does not arise from a true mixture. Both Rindskopf and B. Muthén expressed cautious skepticism on this point. However, in the broader literature, the use of finite normal mixtures to provide semiparametric approximations to nonnormal or irregular distributions is well accepted (whether or not the nonnormality reflects a "true" mixture). In a seminal article, Ferguson (1983) noted that "using such mixtures, any distribution on the real line can be approximated to within any preassigned accuracy" (p. 287). Further, it is this use of finite mixture models that motivated, at least in part, the growth mixture modeling approaches of Nagin and colleagues (Nagin, 1999; Nagin & Land, 1993; Nagin & Tremblay, 2001) and Verbeke and LeSaffre (1996). Specifically, these authors have noted that a key advantage of growth mixture models is precisely that they may provide semi-parametric approximations to irregular but possibly homogeneous distributions of repeated measures, rather than requiring parametric assumptions that are typically made for convenience and are likely incorrect.

Nevertheless, Rindskopf (2003) and B. Muthén (2003) raised the important possibility that model-checking procedures, such as comparing the implied distribution of the mixture to the actual distribution of the observed data, may reveal that the model is only providing an approximation function in the absence of a true mixture. We strongly agree that such comparisons should be made to determine the quality of the model, but the question remains, would they lead one to reject the mixture hypothesis in favor of the hypothesis that the population distribution is homogeneous but nonnormal? We suspect that they often would not. Our reasoning is highlighted in Figure 1. Here we have plotted nonparametric kernel density estimates of the data distributions for the first and last time points of a single replication (chosen at random) from the Skew 1.5 Kurtosis 6 condition with N = 600. The density estimates are clearly not normal, and

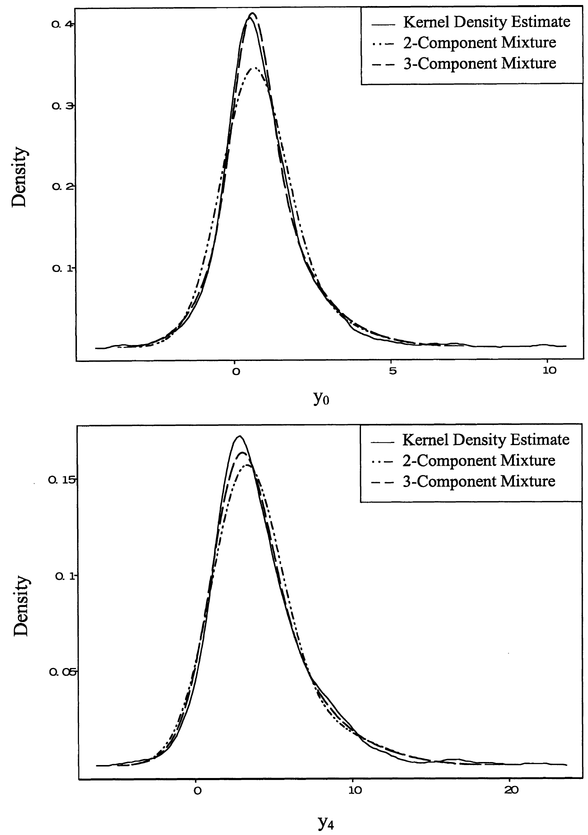

Figure 1. Estimated densities for the population distributions of the first and last time points (y_0 and y_4) of the growth model considered in Bauer and Curran (2003) for a single replicate from the Skew 1.5 Kurtosis 6 condition. Kernel density estimates were constructed using Gaussian kernels with bandwidths chosen to minimize the approximated mean integrated squared error (.30 for y_0 and .73 for y_4). Two- and three-class model densities are the densities implied by fitting growth mixture models of the form described by Bauer and Curran.

their general shape resembles the bottom panel of Figure 4 from Bauer and Curran (2003), as it should. Juxtaposed with the nonparametric kernel density estimates are the implied densities of two- and three-class growth mixture models.

Figure 1 supports the point made by Rindskopf (2003) and B. Muthén (2003) that the two-class model might not fit in an absolute sense. Clearly, the two-class model does not fully reproduce the features of the observed data distributions, though it might be considered "close enough" by some. In response to early drafts of our original manuscript, B. Muthén proposed a new skew and kurtosis test that formally evaluates this lack of fit. Although intriguing, we cannot comment in greater detail about this new test because the associated analytics were not explicated in B. Muthén's comment and the supporting work cited by B. Muthén was still in preparation at the time of this writing. However, we can consider the empirical performance of this new test as it is currently implemented in L. K. Muthén's commercial software package Mplus (L. K. Muthén & Muthén, 1998). We thus applied this test to the same randomly selected replication from our simulation considered above.

Just as B. Muthén (2003) reported in his comment, the new skew and kurtosis test rejects the two-class model for this replication at both the univariate and the multivariate levels.[1] If we include a third class, the fit of the model improves considerably. This is reflected in Figure 1, which shows that the implied density of the three-class model follows the nonparametric kernel density estimates quite closely. However, the skew and kurtosis test of B. Muthén rejects this model, suggesting that the test may be so powerful as to detect even relatively trivial departures of the data from the model. This naturally leads to questions about the practical usefulness of the test when the data do in fact arise from a true mixture. Specifically, because this test assumes strict multivariate normality within components, it must be determined whether true mixtures with minor deviations from this distributional assumption would also be rejected. Although we suspect that this may be the case, this is of course only conjecture on our part, and much future analytical and empirical work is needed to more fully evaluate this new test.

Adding more classes further improves the performance of the model on the skew and kurtosis test. This is a predictable consequence of the enhanced ability of the model to approximate the nonnormal repeated measures distribution given more classes (and more model parameters). In fact, with four classes the new test of skew and kurtosis is no longer significant at either the univariate or the multivariate level, so we see that in this case the data are consistent with a mixture, but a mixture made up of even more components than we examined previously. Hence, we continue to make the same "faulty conclusions" even with this new test. The same pattern of results can be expected more generally, as the closeness of the implied and sample values of skew and kurtosis tend to increase monotonically with the number of components estimated (this is in contrast to measures such as the BIC that counterbalance model fit with parsimony). The nonparametric density estimates in Figure 1 may be viewed as the most extreme expression of this point. Because we used Gaussian kernels for the nonparametric density estimates, even these density functions may be viewed as finite normal mixtures,[2] with N components of equal variance and class proportions of N^{-1} (see Everitt & Hand, 1981, pp. 118–124; McLachlan & Peel, 2000, p. 8). Of course, we would naturally be skeptical of a model with N components, but it is often the case that far fewer components are needed to obtain almost identical fit to nonparametric density estimates (Scott & Szewczyk, 2001).

In sum, as demonstrated in the example above, we hypothesize that the new skew and kurtosis test will often suggest that more components are necessary to adequately reproduce the observed data distributions but that it will likely provide less help in differentiating true mixtures from nonmixtures. Counterexamples could, of course, be provided, such as distributions that are not properly continuous (such as those with strong floor or ceiling effects). In cases like these, use of a normal mixture might be questioned from the outset. Other model-checking procedures, such as posterior predictive checks (Gelman, Meng, & Stern, 1996; Meng, 1994) or residual diagnostics (Lindsay & Roeder, 1992), might fare better in this regard. In general, we are quite pleased that our article has spurred further consideration and development of such procedures, especially given that this was one of the original motivating goals of our work.

We now turn to the second theme of the comments,

---

[1] In many ways we find the visual depiction in Figure 1 preferable to simply contrasting skew and kurtosis values. For instance, B. Muthén (2003) emphasized the "modest" skew and kurtosis values of the examples given in his Equations 1–4. However, even the small values of negative kurtosis in his Equation 4 can be generated by strongly bimodal data. In general, we believe that plotting the actual density functions is a useful supplement to reports of summary statistics, as values of skew and kurtosis alone do not convey the same depth of information.

[2] We used Gaussian kernels in Figure 1 in part to make this conceptual point. The reader may wonder if the resemblance between the nonparametric density estimate and the two- and three-class density estimates in Figure 1 is an artifact of this choice. In actuality, the particular kernel function employed is often of little consequence for the ultimate shape of the density estimate (Silverman, 1986, pp. 42–43). Bearing this out, other choices of kernel for this data produced plots virtually identical to Figure 1.

which poses the question, Even if it is difficult to distinguish a normal mixture from a nonnormal nonmixture, should we be concerned? This theme prompts us to examine more closely the functions of models as representations of reality within an empirically based science.

Alternative Views of Reality

A key issue raised by all three comments is whether we should really be concerned by the fact that a mixture could be estimated where none existed. Even if the classes are spurious, the model nevertheless provides one depiction of the underlying data that may prove scientifically useful in certain circumstances. As Rindskopf (2003) nicely articulated, "In the end, researchers may not know what is right but only what model is most helpful in achieving other scientific goals" (p. 367). Cudeck and Henly (2003) highlighted this same issue in their provocative description of the game Guess the Model.[3] We interpret their general point to be that all simulation studies are likely irrelevant because they do not conform to the real practice of data analysis. Specifically, when selecting from models for a specific set of data, we have no "true" model to serve as a benchmark for indexing the degree of bias, badness of fit, or number of spurious effects. All of our models are wrong, and it is quite possible that there is no "right" model to discern whatsoever. The real task at hand is to decide which model is most useful.

We agree with all three commentators on this general point. Specifically, even if there exists some "truth as God sees it," ultimately this truth is not accessible to us. We can never know when we fit a growth mixture model whether the latent classes correspond to true subpopulations or whether they serve simply to approximate a nonnormal distribution of repeated measures. Interestingly, we view this as one of the primary points of our target article. It was a point we felt compelled to make because in the empirical applications of the model that we have seen to date, the approximation function of the latent classes has rarely been considered as an alternative possibility to the existence of discrete groups (with exceptions in the work of Nagin and colleagues [Nagin, 1999; Nagin & Land, 1993; Nagin & Tremblay, 2001] and Verbeke & LeSaffre, 1996, as noted previously).[4] Further, it is the very promise of the models to diagnose and treat population heterogeneity that seems to make them so appealing to applied researchers. Our goal has never been to discourage researchers from using these models for this purpose but to encourage researchers to think more carefully and skeptically about the alternative functions that the same model may serve.

Whereas our focus was on demonstrating that the same model can equally depict two quite different realities, the comments raise the important converse question, Can different models provide alternative depictions of the same reality? We believe that they can. Consider the nonnormal data conditions from our article. Both the standard single-class latent curve model and the two-class growth mixture model are "incorrect." The latent curve model fails to take account of the nonnormality of the data, whereas the growth mixture model assumes that there are two groups when the population is actually homogeneous. Yet both models capture different aspects of the data.

[3] Cudeck and Henly (2003) dedicated a separate portion of their comment to guessing the specifics of the model on which we focused our original article. We aimed this article at a diverse audience, and it is always difficult to strike the right balance between technical rigor and general accessibility. We believed then at the time of writing, and continue to believe now, that our presentation of the model was sufficiently complete (one reviewer was able to fully replicate all of our primary analyses solely on the basis of our original description) while also written in a manner that would be accessible to a general audience. Whereas Cudeck and Henly considered our presentation to be deficient, others might consider this very same presentation to be a strength of our work. For readers who desire additional technical details on the specific growth mixture model that we investigated, the references to B. Muthén and Shedden (1999) and L. K. Muthén and Muthén (1998, Appendix 8) should suffice. We note, however, that the conclusions we drew in our article are equally applicable to the model previously developed by Verbeke and LeSaffre (1996) that is featured in Cudeck and Henly's comment.

[4] For point of clarification, even recent work by B. Muthén and colleagues has not explicated that latent classes may serve to approximate nonnormal distributions. For example, B. Muthén and Shedden (1999) log transformed their repeated measures as part of an evaluation of the "sensitivity" of the growth mixture model to nonnormality, and B. Muthén (2001a, Figures 1.1 and 1.2) depicted the aggregate of two normal components as being itself normally distributed (a case in which the mixture would not be identified). We thus believe that our original article provides an important perspective on the assumptions of growth mixture models that may be unfamiliar to many applied researchers.

If our interest is in parameter estimates for the whole population, the latent curve model is the simplest representation of the data. If our interest is in the distribution of individual trajectories, the growth mixture model provides greater detail. Thus, the two models capture different aspects of the same reality. What remains is to decide which model is optimal. To some extent, this depends on our scientific goals, and the diverse opinions of the commentators on this matter reflect the fact that our goals may change from one application to another. Rindskopf (2003) placed a premium on parsimony, B. Muthén (2003) emphasized consistency with theory, and Cudeck and Henly (2003) focused on the utility of the model for description and prediction. We examine each of these potential goals in turn.

Parsimony

Rindskopf (2003) offered parsimony as a means for deciding between a mixture and a nonmixture as the optimal model for a given set of data. In debating between a piecewise linear approximation to a nonlinear relationship and a continuous quadratic function, Rindskopf noted that each model provides almost the same measure of prediction but that the quadratic model requires one curve and fewer parameters so and thus is preferred. We see this perspective as a time-honored application of Occam's razor. Within the context of growth mixture modeling, Occam's razor suggests that a single-class model should be used whenever possible because it requires fewer parameters. Following this logic, Rindskopf suggested that, in the presence of nonnormal outcomes, one might perform nonlinear transformations to normalize the data and retain a simpler model. To the extent that the transformation succeeded, the need for the latent classes would be reduced, if not obviated altogether. An interesting example of Rindskopf's point is provided in an analysis of tomato root growth by Gutierrez, Carroll, Wang, Lee, and Taylor (1995), who found that two normal components were necessary to capture the distribution of raw data but that only one component was necessary when the inverse transformation was applied (see McLachlan & Peel, 2000, pp. 177–178). Like Rindskopf, these researchers preferred the more parsimonious one-component model to the two-component mixture.

Suppose, however, that one really believed that a mixture was "present." We set aside for the moment whether this belief stems from a theoretical model of the "true" process giving rise to the data or simply reflects an opinion that a mixture would provide a better model for description and prediction. In either case, performing a nonlinear transformation might actually suppress the very mixture that is hypothesized to exist. This result does not imply that there was no mixture in the first place, nor does it mean that the mixture would be less useful for prediction. It is always possible that follow-up studies will show quite different patterns of predictive relationships for the classes, a point that might be lost in using a single-class model with transformed data. Thus, whereas we fully agree with Rindskopf (2003) that nonlinear transformations may allow for the selection of a more parsimonious model, we also believe that there is no certainty that the more parsimonious model will ultimately prove to be the "better" representation of reality. As with all analysis decisions, caution must be exercised in deciding whether to perform a nonlinear transformation, as it may in some cases occlude the phenomena of most interest.

The Role of Theory

B. Muthén (2003; and also Rindskopf, 2003, to some degree) suggested that substantive theory should be used as a guide for selecting the optimal model among possible alternatives. The argument is that we should adopt a model if it generates results that are consistent with the motivating psychological theory. B. Muthén expressed concern that our article "missed the opportunity to contribute a thorough discussion of how psychological theory can guide [growth mixture modeling]" (p. 372). Interestingly, we believe that B. Muthén's perspective on this matter is quite similar to our own. Indeed, B. Muthén's remark that "once substantive theory has been formulated, it can be used to predict an interwoven set of events that can then be tested" (p. 373) closely parallels our own point that the construct validity of the latent trajectory classes is best assessed by constructing "a nomological network of results that are consistent with the idea of population heterogeneity" (Bauer & Curran, 2003, p. 359). We thus fully endorse B. Muthén's discussion on the value of using substantive theory to bring auxiliary information into a growth mixture model. This portion of his comment nicely expands on our own general views regarding the construct validity of the latent trajectory classes.

That said, we believe there remains an important difference between our statement and the one quoted from B. Muthén (2003); namely, we explicated that not only should the results be consistent with population heterogeneity, they should also "not necessarily be expected if the population was homogeneous" (Bauer & Curran, 2003, p. 359). This qualifier follows the classic recommendation of Cronbach and Meehl (1955), who noted that assessments of construct validity require

> a statement like the conclusions from a program of research, noting what is well substantiated and what alternative interpretations have been considered and rejected. The writer must note what portions of his proposed interpretation are speculations, extrapolations, or conclusions from insufficient data. The author has an ethical responsibility to prevent unsubstantiated interpretations from appearing as truth. (p. 297)

In the spirit of this important charge, a primary goal of our target manuscript was to simply highlight that strong empirical evidence suggesting the existence of multiple latent classes might result from the actual existence of such classes or might instead be due solely to the violation of distributional assumptions.

To highlight the possible implications of failing to substantiate the construct validity of the latent classes, we draw on an example from B. Muthén's (2003) comment, the case in which a growth mixture model indicates that an intervention or treatment has a significant effect only in one segment of the population. Logically, this would suggest the implementation of a targeted intervention program that would be administered specifically to this population segment. Efforts would be made to include at-risk individuals with high probability of belonging to the "responsive" segment and to exclude at-risk individuals belonging to the "resistant" segment. This has the promise of being an important and appealing application of the mixture model. As our simulation study demonstrated, however, when a growth mixture model is applied erroneously, it is entirely possible that a covariate having a significant effect for the whole population would show a much stronger effect in one class than another. Suppose the same was true of the treatment program model: Though all individuals are equally likely to benefit from the program, spurious responsive and resistant classes are estimated. Classifying individuals into these groups and applying the treatment only to the so-called responsive class would result in the exclusion of a large number of individuals who could have benefited from the program (in addition to the possible stigmatization associated with classifying individuals).

We certainly agree with B. Muthén's (2003) important point that a population may often consist of latent groups that would differentially benefit from an intervention or treatment. We also agree that there may be qualitatively distinct patterns of development in any broadly sampled population. Consistent with the original goals of our work, however, we emphasize that the hypothesis of population heterogeneity must be rigorously substantiated, particularly with respect to alternative competing hypotheses. Currently available model fit indices do not provide this substantiation, though our observation of current best practice is that they have often been interpreted as such. Instead, empirical results must be accumulated that are both consistent with the interpretation of population heterogeneity and inconsistent with alternative explanations. We are hopeful that the discussion of this issue in our original article, the comments, and this reply will jointly motivate applied researchers to conduct investigations of developmental heterogeneity that directly address the construct validity of the hypothesized latent trajectory classes.

Description and Prediction

When considering the critically important issue of construct validity, there is an implicit assumption that the construct has been imbued with some theoretical meaning. In other words, the construct is more an agent in a theoretical model for the process generating the data and less linguistic shorthand for representing complex data. However, we interpret Cudeck and Henly's (2003) perspective to be consistent with the latter when they argue that the primary scientific function of a model is description and prediction. Whereas this is certainly a valid perspective, we believe it downplays what we regard to be a fundamental goal of model building, namely, developing an understanding of the process that gave rise to the data.

In our opinion, it is this goal of explanation that guides much social science research. Questions of etiology are framed by mediational models; hypotheses on risk and resilience are cast in terms of moderated relationships; and questions of population heterogeneity are addressed with mixture models. We thus part ways with Cudeck and Henly (2003) in their assessment that "practical research . . . begins with interesting data where in all likelihood no operating model exists" (p. 381). Having both been trained as substantively oriented researchers ourselves, we do not believe that this is a broadly held view of how empirical research is typically conducted. We believe that the majority of psychological research does not begin with interesting data but instead begins with an interesting question. Given a theoretical model of some underlying etiological process, the primary goal of the analysis is not limited to description and prediction but also includes explanation. The observed data is then primarily interesting to the extent that it helps to empirically evaluate some aspect of the hypothesized theoretical model.

If our concern is focused solely on description and prediction, then in principle any two models can be considered equally valid if they equivalently reproduce the observed data structure. As B. Muthén (2003) nicely emphasized in his comment, the two competing models are then simply alternative descriptions of the same data. However, it is well known that models that equivalently fit the data (or nearly so) may be based on radically different explanations of the process that generated the observed data (e.g., Lee & Hershberger, 1990). Should these alternative models then be regarded as simply providing different and complementary views of the same data structure? This may be reasonable in some situations, but less so in others. To draw on a classic astronomical example, does it matter whether our model of the planetary system is geocentric or heliocentric if both models equivalently predict the phases of the moon? If the goal of the model is prediction without explanation, then whether the earth or the sun lies at the center of the system is ultimately inconsequential. On the other hand, the distinction is critical if we are to come to a better understanding of the process whereby the phases of the moon occur. Similarly, if our goal is to design and implement an experimental intervention for antisocial behavior in children, it is critical that we accurately understand the etiological mechanisms that give rise to the behavior (e.g., Conduct Problems Prevention Research Group, 1992).

Given all of the above, what can we say to a substantive researcher whose empirical data are consistent with the theoretical prediction of distinct subgroups within the population? Cudeck and Henly (2003) convincingly articulated the point that we can never know the true population structure and that it may be counterproductive to attempt to determine which model is "correct." Although we clearly agree with them on the first point, we also believe that it is important and productive to attempt to distinguish between alternative models. Within the context of growth mixture modeling, this means considering whether the population is truly characterized by qualitatively different developmental trajectories or whether the individual variation is continuous and simply nonnormal. What the growth mixture model provides is a structured method for evaluating the consistency of the empirical data with the model of population heterogeneity. However, determining that the data are consistent with this model (i.e., multiple latent classes are found to be optimal via fit criteria) does not confirm the hypothesis, nor does it rule out other hypotheses. When there is an interest in augmenting the goal of prediction with that of explanation, these alternative models should be carefully considered and thoughtfully investigated prior to making strong inferences back to theory.

In sum, we wholeheartedly agree with the commentators that, outside of the artificial context of simulation studies, no model is ever "correct." We also agree that different models may provide different perspectives of the same underlying reality. Which model is most useful of course depends on the specific circumstances of the application at hand, and the criteria for model selection offered by the commentators (parsimony, theoretical consistency, description and prediction, to which we add explanation) should be weighed according to the scientific goals of the particular application. It remains true, however, that if a given statistic or a particular model is used for the conventional purpose of testing a hypothesis about the process at hand, alternative explanations for the same results should be explicated and, when possible, evaluated. If a growth mixture model is used to test the theory that there are distinct population subgroups characterized by different patterns of growth, then the alternative explanation that the data are homogeneous but nonnormal is viable until additional information can be brought to bear.

Conclusion

So do we really believe that our work represents *much ado about nothing?* We do not; nor do we believe that it represents the final word on an important new methodology. We are excited about the insights afforded by this methodology and look forward to ongoing developments in this area. The primary goal of our own research was simply to provide a word of caution to substantive researchers against the potential overinterpretation of growth mixture models, a point echoed by the commentators.

We conclude by noting that the following challenges remain for using these models in practice: (a)

The existence of multiple classes will necessarily induce nonnormally distributed data (assuming the classes are not identical); (b) other processes exist that will similarly induce nonnormally distributed data that are wholly independent of the existence of multiple classes; (c) it is highly likely that two or more classes will be extracted in the presence of even trivially nonnormal data, regardless of the process that gave rise to this data; and (d) it is currently difficult to conclude whether clear support for multiple classes in the sample reflects the structure of the population (if we believe such a structure to exist) or whether the classes are simply serving to approximate a complex distribution. Although universally thoughtful and at times provocative, the comments on our work do not lead us to question these core conclusions. Our intent in highlighting these issues is not to slight the model or dissuade its use. Instead, we hope to better inform applied researchers about the alternative functions the model may serve and to motivate innovative analytical and empirical research that might allow us to better address these intriguing challenges in the future.

References

Bauer, D. J., & Curran P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8,* 338–363.

Conduct Problems Prevention Research Group. (1992). A developmental and clinical model for the prevention of conduct disorder: The FAST Track Program. *Development and Psychopathology, 4,* 509–527.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Cudeck, R., & Henly, S. J. (2003). A realistic perspective on pattern representation in growth data: Comment on Bauer and Curran (2003). *Psychological Methods, 8,* 378–383.

Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions.* London: Chapman & Hall.

Ferguson, T. S. (1983). Bayesian density estimation via mixtures of normal distributions. In M. H. Rizvi, J. S. Rustagi, & D. Siegmund (Eds.), *Recent advances in statistics* (pp. 287–302). New York: Academic Press.

Gelman, A., Meng, X. L., & Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica, 6,* 733–807.

Gutierrez, R. G., Carroll, R. J., Wang, N., Lee, G.-H., & Taylor, B. H. (1995). Analysis of tomato root initiation using a normal mixture distribution. *Biometrics, 51,* 1461–1468.

Lee, S., & Hershberger, S. (1990). A simple rule for generating equivalent models in structural equation modeling. *Multivariate Behavioral Research, 25,* 313–334.

Lindsay, B. G., & Roeder, K. (1992). Residual diagnostics for mixture models. *Journal of the American Statistical Association, 87,* 785–794.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88,* 767–778.

McLachlan, G., & Peel, D. (2000). *Finite mixture models.* New York: Wiley.

Meng, X. L. (1994). Posterior predictive *p*-values. *The Annals of Statistics, 22,* 1142–1160.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105,* 156–166.

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods, 8,* 369–377.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55,* 463–469.

Muthén, B. O. (2001a). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Mahwah, NJ: Erlbaum.

Muthén, B. O. (2001b). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In A. Sayer & L. Collins (Eds.), *New methods for the analysis of change* (pp. 291–322). Washington, DC: American Psychological Association.

Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24,* 882–891.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide* (Version 2) [Computer software manual]. Los Angeles: Muthén & Muthén.

Nagin, D. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods, 4,* 139–157.

Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology, 31,* 327–362.

Nagin, D. S., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological Methods, 6,* 18–34.

Rindskopf, D. (2003). Mixture or homogeneous? Comment on Bauer and Curran (2003). *Psychological Methods, 8,* 364–368.

Scott, D. W., & Szewczyk, W. F. (2001). From kernels to mixtures. *Technometrics, 43,* 323–335.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* New York: Chapman & Hall.

Verbeke, G., & LeSaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association, 91,* 217–221.

Received April 17, 2003
Revision received May 20, 2003
Accepted May 27, 2003 ■

New Editor Appointed for *Contemporary Psychology: APA Review of Books,* 2005–2010

The Publications and Communications Board of the American Psychological Association announces the appointment of Danny Wedding (Missouri Institute of Mental Health) as editor of *Contemporary Psychology: APA Review of Books*, for a 6-year term beginning in 2005. The current editor, Robert J. Sternberg (Yale University), will continue as editor through 2004.

All reviews are written by invitation only, and neither the current editor nor the incoming editor receives books directly from publishers for consideration. Publishers should continue to send three copies of books for review consideration, along with any notices of publication, to PsycINFO Services Department, APA, Attn: *Contemporary Psychology: APA Review of Books* Processing, P.O. Box 91600, Washington, DC 20090-1600 or (for UPS shipments) 750 First Street, NE, Washington, DC 20002-4242.