

Published in final edited form as:

*Measurement ( Mahwah N J)*. 2004 ; 2(4): 219–247.

## Interindividual Differences in Intraindividual Variation: Balancing Internal and External Validity

Patrick J. Curran and R. J. Wirth

Department of Psychology, University of North Carolina at Chapel Hill

We are honored to have the opportunity to offer a few words in response to Molenaar's manifesto (this issue) calling for the return of the individual to scientific psychology. His arguments are clearly articulated, strongly supported, and quite timely, particularly given the recent surge in popularity of longitudinal studies of individual developmental trajectories. Despite our enthusiasm for both his theoretical and statistical perspectives, we would like to briefly articulate a somewhat different view. We do so for two reasons.

First, if we simply agreed with his manifesto, this would make for a rather boring commentary and we would rather not bore you. Second, we do not perceive our current set of interindividual variation (IEV) methodologies to be as limited as Molenaar seems to imply. More specifically, Molenaar makes the argument that (a) scientific psychology is almost exclusively based on statistical methods designed for the study of IEV. (b) individual development is defined by processes that are inherently characterized by intraindividual variation (IAV). ergo (c) IEV statistical methods cannot be validly applied to the study of individual development. We agree with Molenaar that there can be a dangerous disjoint between theory and statistics and we have made related arguments elsewhere (Curran & Willoughby. 2003). However, we also believe that existing IEV methodologies can be validly used to gain important insights into the behavior of both individuals and groups.

Indeed, we believe that the application of IEV methodologies to data comprised of samples of individuals with the goal of making generalizations to broader populations is a cornerstone of empirical science. To better articulate our perspective, we begin by considering the distinction between internal and external validity.

### INTERNAL AND EXTERNAL VALIDITY

A core concept of any empirical science is the distinction between internal and external validity. Cook and Campbell (1979) defined internal validity as the approximate validity of the assertion that an observed relation between two variables reflects a causal process or that the lack of an observed relation reflects the lack of a causal process. In contrast, they defined external validity as the approximate validity to which an inferred cause (or lack of cause) can be generalized to different individuals, contexts, and times. There are three issues we highlight here. First, by definition, internal validity precedes external validity; that is, we cannot generalize a causal relation prior to establishing the validity of a causal relation. Second, external validity is focused on the generalization of a causal process to people, places, or times that were not under study in a given sample. Finally, both definitions include the term *approximate*, emphasizing that all models are incorrect to some degree and

our substantive conclusions must always be tempered by this fact (see, e.g., MacCallum, 2003).

We believe the distinction between internal and external validity plays a key role in considering the IEV and IAV distinction articulated by Molenaar. More specifically, we believe that Molenaar focused almost exclusively on issues of internal validity without equal consideration to associated implications for external validity. Our perspective is that internal and external validity are often counterbalanced such that increasing the focus on one tends to lessen the ability to establish the other. For these two forms of validity to be more equitably balanced, we should consider the simultaneous study of IEV in the presence of IAV. To focus exclusively on the study of one or the other undermines our ability to draw valid generalizations about the complexities of the world around us.

Although he does not use this same terminology, Molenaar implicitly argues that the inappropriate application of IEV methodologies to IAV processes is a direct threat to internal validity (what Cook & Campbell, 1979, specifically termed *statistical conclusion validity*). Molenaar cites prior simulation studies to demonstrate this threat by fitting a one- or two-factor model to a sample that was generated by individually varying factor models. We agree with Molenaar's conclusion that the inappropriate application of a single fixed-factor model to data derived from individually varying factor models directly threatens the internal validity of the hypothetical study. That is, concluding that a two-factor model holds in the population when individuals are truly characterized by between one- and five-factor models is incorrect.

However, in our opinion, what is not also considered are the corresponding implications for external validity. At the extreme, Molenaar's argument implies that each individual is unique to the point that no valid generalizations can be drawn to any other individual based on the empirical study of a given individual. In other words, the intense study of one individual does not provide us any degree of understanding of developmental processes beyond that particular person. To understand the developmental processes of another individual, we must then intensively study that particular person. We are left with an infinite number of life stories, the characteristics of any one wholly unrelated to any other. This does not allow for the systematic building of an empirically based knowledge structure about human development because no knowledge can be generalized beyond the specifics of any single individual. We see this perspective as undermining one of the key goals of an empirical science, that being external validity.

We believe that an empirical psychological science should strive to validly draw on prior observations of a set of individuals to make generalizations to a broader set of unobserved individuals (or similar individuals embedded within different contexts or different developmental periods). This is a simple restatement of external validity. Having said that, we do not see our goal as empirical scientists to derive valid explanations of the behavior of *all* individuals within *all* contexts at *all* developmental periods (as is reflected in fitting a single factor loading matrix to the entire population). Instead, we see our goal to be the derivation of valid explanations for the behavior of *types* of individuals within *particular* contexts at *specific* developmental periods. To accomplish this, we turn to what we term *conditional IEV methodologies*. These general approaches are characterized by the fitting of models to groups of individuals under the assumption that, whereas models may vary in quantity or form across subgroups, these approximately hold for all individuals within subgroup.

For example, it is clear in Molenaar's simulation that, given how these participants were simulated, the two-factor model does not describe all of the individuals in the sample.

However, this two-factor model does conditionally describe the subset of individuals who are truly characterized by a two-factor model in the population. A three-factor model conditionally describes the subset of individuals who are truly characterized by a three-factor model in the population. And so on. The key here is to not consider a single fixed model that is expected to hold for all possible individuals, but instead to consider a smaller set of conditional models that hold for subsets of individuals. We see the consideration of conditional IEV models moving us into the interior of the space demarcated on one side by the strict study of the individual and on the other side the sole focus on the characteristics of the overall group. We strongly believe that many IEV methodologies can be validly applied in the estimation of these conditional models, and it is to these we now turn.

## INDIVIDUAL TRAJECTORY ANALYSIS

There has been a recent surge in popularity in the application of various forms of “growth curve” analysis in the study of individual developmental trajectories over time. These can be estimated using individual regressions (e.g., Rogosa & Saner, 1995), the general multilevel model (e.g., Bryk & Raudenbush, 1987), the structural equation model (e.g., Meredith & Tisak, 1990), or a variety of other approaches. We ourselves have spent much time exploring various aspects of latent curve analysis and have argued that these models provide a potential intersection between individual and group characteristics of development (e.g., Curran, 2003; Curran & Hussong, 2003; Curran & Willoughby, 2003). Molenaar voices a strong critique of these same latent curve models as being particularly insensitive to the presence of IAV. However, there is an interesting relation between the latent curve model as fitted to pooled sample statistics and the estimation of case-by-case trajectories for each individual considered in isolation that might in part address this critique.

For example, say that we were to fit a quadratic trajectory to each individual in isolation of all other individuals using the standard ordinary least squares (OLS) regression model. That is, for the first individual we fit the quadratic model, save out the three trajectory parameters, and repeat this process for all of the individuals in the sample. This appears to truly consider each individual in absence of all other individuals, and is thus an IAV technique. However, as was demonstrated by Rogosa and Saner (1995), if we gather together all of these individual trajectory estimates, the means of the OLS estimates are precisely equal to the fixed effects from a standard latent curve model and the variances of the OLS estimates (corrected for imprecision in estimation) are precisely equal to the random effects from a standard latent curve model. Given this isomorphism, which of these approaches is truly sensitive to IEV and which to IAV? Our perspective is that under many conditions, the careful application of growth curve models allows for valid inferences about both IAV and IEV processes.

## MULTIPLE GROUP AND FINITE MIXTURE MODELS

It appears that Molenaar’s primary concern in the misapplication of IEV techniques centers on the use of single-sample methods. This is reflected in his arguments regarding the inappropriateness of fitting a single factor loading matrix to an entire sample of individuals whose data were each generated by their own unique factor loading matrices. However, we believe that a promising option for simultaneously considering IAV and IEV in statistical models is through the consideration of a broad class of multiple group models. These types of models are not new, and multiple groups structural equation modeling (SEM) has been used for many years (e.g., Jöreskog, 1971). Instead of assuming that a single model holds for all individuals in the sample (as is a key concern of Molenaar), the multiple group approach allows for parameter values or even model structures to vary across discrete groups of participants. Thus, to partially address Molenaar’s concern regarding the imposition of just a

single factor loading matrix across all possible participants, we can simultaneously test whether unique matrices differing in magnitude or dimension hold within discrete subgroups of individuals. This of course still assumes that the model parameters hold within each group, but there is the opportunity to explicitly model a degree of heterogeneity across groups.

The classic multiple groups structural equation model assumes that the grouping variable is both discrete (either the individual is a group member or is not) and observed (group membership is known). However, a class of models exists that allows for the simultaneous estimation of model parameters and of group membership. These are generally referred to as *finite mixture models*. Finite mixture models have a long history (e.g., Pearson, 1894) and have recently been extended to confirmatory factor analysis (Yung, 1997), structural equation models (Armiger, Stein, & Wittenberg, 1999), and growth models (Muthén & Shedden, 1999), making them sufficiently general to be useful in a variety of research applications. A finite mixture assumes that a complex nonnormal aggregate distribution can be approximated by a sum of two or more simpler normal distributions. The aggregate density is then calculated as a weighted sum of the component densities where the weights represent the proportion of the full sample belonging to each component. Each individual has a specific probability of membership in all possible groups. When these probabilities are known and are equal to either 0 or 1, then this simplifies to the traditional multiple groups model in SEM.

Interestingly, the finite mixture model provides a formal analytic framework for thinking about the relations between idiographic and nomothetic approaches to empirical science. At one extreme, the existence of only a single mixture reflects that one model holds for all individuals in the sample (consistent with a nomothetic perspective). At the other extreme, the existence of as many mixtures as participants reflects the need to fit a different model to every single individual (consistent with an idiographic perspective).<sup>1</sup> The goal of finite mixture modeling is to fall somewhere between these two extremes, thus allowing for some pooling over individuals while at the same time allowing for different models to be fit to different subsets within the sample. Although we feel that the finite mixture model offers much in the way of thinking more formally about balancing IEV and IAV, we also have several reservations about the current use of these methods in applied research (Bauer & Curran, 2003a, 2003b, in press). However, this is a powerful and exciting set of methodologies that has much promise, particularly when considering new methodologies that are more sensitive to Molenaar's concerns.

## TESTING AND PROBING MULTIPLICATIVE INTERACTIONS

Closely related to a multiple groups perspective is that of conditional effects defined by multiplicative interactions. Whereas a multiple groups approach assumes that individuals are embedded within some discrete clustering, interactions examine the conditions under which a particular effect might hold. For example, suppose we estimated a standard regression model examining the effectiveness of a substance use treatment program for adolescents. As Molenaar nicely articulates, the fixed main effects from this model may not generalize to any single individual. However, from the outset it may be misguided to ask the question, "Is the treatment program effective?" Instead, we might be better informed by asking the question, "For what type of adolescent is the treatment program effective?" This introduces the concept of moderation (e.g., Baron & Kenny, 1986), which allows for the explicit

<sup>1</sup>Note that this latter condition is for heuristic value only in that there is insufficient observed information to estimate a unique mixture for each individual in the sample. Fitting such individual-based models requires the use of methodologies like those described by Molenaar (2004).

modeling of conditional effects as a function of other individual or group influences. Methods for testing and probing multiplicative interactions are well developed for numerous IEV models including standard regression models (e.g., Aiken & West, 1991), random effects models (e.g., Bauer & Curran, 2004; Curran, Bauer, & Willoughby, in press), and latent curve models (e.g., Curran, Bauer, & Willoughby, 2004), among many others. Examination of conditional effects increases internal validity with the simultaneous goal of maximizing generalizability.

## MULTILEVEL MODELING

Traditional quantitative methodologies such as factor analysis and SEM assume that all residuals are independent and identically distributed. An alternative perspective that may be used to better examine IEV and IAV is the multilevel or mixed model. Multilevel models are designed to allow for multiple sources of variability arising from nonindependent data structures (Raudenbush & Bryk, 2002). Classic examples include children nested within classrooms, siblings nested within family, or time nested within individuals. Explicitly modeling these multiple sources of variability allows us to better differentiate the characteristics of individual and group processes. Multilevel models have recently been expanded to the full structural equation model (e.g., Bentler & Liang, 2003; Du Toit & Du Toit, in press) and these methodologies are particularly promising when seriously taking heed of Molenaar's warnings. For example, random effects can be included for parameters that have traditionally been treated as fixed and hence remove the constraint that a single vector of parameters represents an entire population. In principle, multilevel factor analysis might provide a mechanism for more appropriately studying the simulated data described by Molenaar through the estimation of individual variability across the factor loading matrices. The relaxation of these assumptions facilitates greater flexibility in modeling conditional effects and thus increases both the internal and external validity of the model.

## DATA VISUALIZATION AND EVALUATION OF CASE-BASED RESIDUALS

Regardless of whether we are fitting a single model to a single sample, or multiple models to multiple samples, we still make the key assumption that the model approximately represents the observed data of all individuals within the given sample. Of course this condition may not hold for all individuals and well-developed methods exist for the careful examination of *person model fit*. Such approaches can take the form of inferential methods (e.g., Fox, 1991) or be based on visualization techniques (e.g., Tufté, 2001; Wainer & Thissen, 1981). These techniques may allow for the identification of one or a small subset of individuals who are distinctly distant from the rest of the group or for whom a given model poorly represents their observed data. Recent advances within the SEM framework include Bollen and Armiger's (1991) case-based residual analysis and Reise and Widaman's (1999) study of individual contributions to the likelihood within the SEM framework. Taken together, individually focused visualization and residual analysis may allow for insights into which specific participants a given model may not hold.

## CONCLUSION

We thank Molenaar for providing such an insightful and well-articulated call for the permanent return of the individual to scientific psychology. We enthusiastically endorse the majority of his key points and plan to take better heed of these issues in our own programs of research. However, we also believe that our formidable collection of existing IEV statistical methodologies cannot only be validly applied in many studies of development, but sometimes *must* be applied in the scientific quest for generalizability. We have tried to briefly describe a few of the more promising methods that we believe allow for maintaining

a balance between the inherently intertwined IEV and IAV processes that characterize human development. We recommend that these techniques be used in conjunction with Molenaar's manifesto in the quest for a better understanding of the individual both in isolation and as a natural member of groups.

## References

- Aiken, L.S.; West, S.G. Multiple regression: Testing and interpreting interactions. Thousand Oaks: Sage; 1991.
- Armiger G, Stein P, Wittenberg J. Mixtures of conditional mean- and covariance-structure models. *Psychometrika*. 1999; 64:475–494.
- Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*. 1986; 51:1173–1182. [PubMed: 3806354]
- Bauer DJ, Curran PJ. Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*. 2003a; 8:338–363. [PubMed: 14596495]
- Bauer DJ, Curran PJ. Over-extracting latent trajectory classes: Much ado about nothing? *Psychological Methods*. 2003b; 8:384–393.
- Bauer DJ, Curran PJ. Inferential and graphical techniques for testing and probing interactions in fixed and random effects regression. *Multivariate Behavioral Research*. (in press).
- Bauer DJ, Curran PJ. The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*. 2004; 9:3–29. [PubMed: 15053717]
- Bentler, P.M.; Liang, J. Two-level mean and covariance structures: Maximum likelihood via an EM algorithm. In: Reise, S.P.; Duan, N., editors. *Multilevel modeling: Methodological advances, issues, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc; 2003. p. 53–70.
- Bollen, K.A.; Armiger, G. Observational residuals in factor analysis and structural equation models. In: Marsden, P.M., editor. *Sociological methodology*. Oxford, England: Basil-Blackwell; 1991. p. 235–262.
- Bryk AS, Raudenbush SW. Application of hierarchical linear models to assessing change. *Psychological Bulletin*. 1987; 101:147–158.
- Cook, T.; Campbell, D. *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin; 1979.
- Curran PJ. Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*. 2003; 38:529–569.
- Curran PJ, Bauer DJ, Willoughby MT. Testing and probing main effects and interactions in latent curve analysis. *Psychological Methods*. 2004; 9:220–237. [PubMed: 15137890]
- Curran, P.J.; Bauer, D.J.; Willoughby, M.T. Testing and probing interactions in hierarchical linear growth models. In: Bergeman, C.S.; Boker, S.M., editors. *The Notre Dame series on quantitative methodology*. Vol. I: Methodological issues in aging research. Mahwah, NJ: Lawrence Erlbaum Associates, Inc; (in press)
- Curran PJ, Hussong AM. The use of latent trajectory models in psychopathology research. *Journal of Abnormal Psychology*. 2003; 112:526–544. [PubMed: 14674867]
- Curran PJ, Willoughby MJ. Reconciling theoretical and statistical models of developmental processes. *Development and Psychopathology*. 2003; 15:581–612. [PubMed: 14582933]
- Du Toit, S.H.C.; Du Toit, M. Multilevel structural equation modeling. In: de Leeuw, J.; Kreft, I.G.G., editors. *Handbook of quantitative multilevel analysis*. Boston, MA: Kluwer; (in press)
- Fox, J. *Sage University Paper series on Quantitative Applications in the Social Sciences*, Series No. 07–079. Beverly Hills and London: Sage; 1991. Regression diagnostics.
- Jöreskog KG. Simultaneous factor analysis in several populations. *Psychometrika*. 1971; 36:409–426.
- MacCallum RC. Working with imperfect models. *Multivariate Behavioral Research*. 2003; 38:113–139.
- Meredith W, Tisak J. Latent curve analysis. *Psychometrika*. 1990; 55:107–122.

- Muthén B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*. 1999; 55:463–469. [PubMed: 11318201]
- Pearson K. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. Series A*. 1894; 185:71–110.
- Raudenbush, SW.; Bryk, AS. *Hierarchical linear models: Applications and data analysis methods*. 2. Thousand Oaks, CA: Sage; 2002.
- Reise SP, Widaman KF. Assessing the fit of measurement models at the individual level: A comparison of item response theory and covariance structure approaches. *Psychological Methods*. 1999; 4:3–21.
- Rogosa DR, Saner HM. Longitudinal data analysis examples with random coefficient models. *Journal of Educational and Behavioral Statistics*. 1995; 20:149–170.
- Tufte, RE. *The visual display of quantitative information*. Cheshire, CT: Graphics; 2001. (Rev. ed.)
- Wainer H, Thissen D. Graphical data analysis. *Annual Review of Psychology*. 1981; 32:191–241.
- Yung YF. Finite mixtures in confirmatory factor-analysis models. *Psychometrika*. 1997; 62:297–330.