ARTICLE IN PRESS

Learning and Individual Differences xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Commentary on latent class, latent profile, and latent transition analysis for characterizing individual differences in learning☆

Bethany C. Bray[a,b,*], John J. Dziak[a]

[a] The Methodology Center, The Pennsylvania State University, United States
[b] College of Health and Human Development, The Pennsylvania State University, United States

ARTICLE INFO

Keywords:
Latent class analysis
Latent profile analysis
Latent transition analysis
Individual differences
Learning

ABSTRACT

The collection of articles in this special issue focus on latent variable mixture models including latent class analysis (LCA), latent profile analysis (LPA), and latent transition analysis (LTA). These are all methods for summarizing observed variables by postulating an underlying categorical latent variable representing a type or status; in the case of LTA, the status of an individual may change over time and the pathways of change are of interest. As the introductory article by Hickendorff, Edelsbrunner, McMullen, Schneider, and Trezise points out, these methods are useful when theory suggests that a learning or problem-solving process can occur in distinct modes or phases. They can also be useful when it is desirable to give qualitative descriptions of individuals' approaches to a task based on their responses across several variables rather than just simple numerical scores. The articles in this special issue use latent variable mixture models in creative and insightful ways, demonstrating their versatility and practicality. However, some challenges remain for researchers using these methods. A number of exciting future directions remain for quantitative methodologists and applied researchers to work together to address new questions in learning and individual differences research. Latent variable mixture modeling will continue to be a powerful tool learning researchers can use to address the critical, sophisticated, theoretically based research questions facing the field.

1. Introduction

This special issue's editors asked us to reflect on the articles in this special issue with respect to the merits, challenges, and future directions of latent variable mixture models for accelerating the pace of understanding individual differences in learning. As discussed in the introduction to this special issue, the fundamental idea underlying latent variable mixture models is that individuals can be classified into unobservable subgroups based on their responses to observed indicators. By identifying these subgroups, we can model heterogeneity among individuals' learning strategies and development. The articles in this special issue illustrate the flexibility of this family of models to accommodate multiple types of indicators, including categorical indicators with latent class analysis (LCA) and continuous indicators with latent profile analysis (LPA), as well as modeling changes in subgroup membership over time using latent transition analysis (LTA). A particular advantage of LTA for learning research is that it can model stage-sequential development without assuming a unidimensional skill variable or a single shared path of improvement. Collectively, these articles illustrate many possibilities for the fruitful use of latent variable mixture modeling in learning and individual differences research.

2. Merits of latent variable mixture modeling in learning research

Some classic work in cognitive and development psychology used the idea of holistic development in stages of learning. The work of Piaget, for example, focused on the idea that children attaining new development stages do not merely incrementally acquire new facts and smoothly improve on a single dimension of performance, but instead learn different perspectives and holistic strategies that make new solutions to problems possible (see, e.g., Sawada, 1972). In addition, these learning processes do not have to take place over years—in the work of Gestalt psychologists, such as Wolfgang Köhler, learners obtain holistic insights into a task rather than simply gradually increasing their performance by trial and error (Köhler, 1925). A natural statistical model for many aspects of learning, then, may be a latent categorical variable representing stage of understanding or mastery—a learner makes an unseen phase transition rather than simply steadily improving.

☆ This commentary was supported by an award from the National Institute on Drug Abuse (P50-DA039838 to L. M. Collins). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse or the National Institutes of Health. The authors declare no conflicts of interest.

* Corresponding author at: The Methodology Center, The Pennsylvania State University, 404 Health and Human Development Building, University Park, PA 16802, United States.
E-mail address: bcbray@psu.edu (B.C. Bray).

https://doi.org/10.1016/j.lindif.2018.06.001
Received 29 May 2018; Received in revised form 31 May 2018; Accepted 1 June 2018
1041-6080/

Please cite this article as: Bray, B.C., Learning and Individual Differences (2018), https://doi.org/10.1016/j.lindif.2018.06.001

Techniques such as LCA, LPA, and LTA, therefore, have long been attractive statistical tools for understanding learning. The potential for latent variable mixture models has long been recognized (e.g., Clogg & Goodman, 1984), but modern software packages have made them much more accessible and they are now being used in many informative, creative ways in educational, cognitive, and developmental psychology. For example, Lee and Bierman (2016) used LPA to summarize characteristics of educational environments that might act as risk or protective factors for children. As another example, Rhoades, Greenberg, Lanza, and Blair (2011) used LCA to categorize children's home environments into subgroups with unique profiles of risk and examined how they were linked to executive function development. Further, latent variable mixture models may be helpful for studying and classifying types of learning disabilities and other learning challenges (e.g., Geary et al., 2009). The articles in this special issue further showcase the continuing possibilities of these methods for the in-depth study of learning and problem solving.

Hickendorff, Edelsbrunner, McMullen, Schneider, and Trezise (2017) provide a comprehensive introduction to underlying principles of the use of LCA, LPA, and LTA, as well as guidance on practical modeling choices such as selecting the optimal number of latent classes and handling missing data. They point out that these methods are especially useful when there is interest in finding qualitatively distinct subgroups of participants who differ in their learning approaches. McMullen, Van Hoof, Degrande, Verschaffel, and Van Dooren (2018) use LCA to classify children on their ability to use and understand fractional and decimal numbers and they compare prevalence rates of these classes between Finish and Flemish children. They argue that the need to consider both a numerator and a denominator to understand fractional numbers, and the need to imagine an infinite continuum of numbers between any two points to understand decimal numbers, involves new ways of thinking that are not just extensions of existing understanding of integers. This is the kind of qualitative distinction that Hickendorff et al., and we, find especially well-suited to latent variable mixture modeling.

Latent variable mixture modeling is particularly well-suited to studying phenomena that are inherently multidimensional in nature. Latent variable mixture models are sometimes criticized because they often lead to classes that seem ordered, such as "low", "medium", and "high" levels of the construct being studied. In these cases, it is possible that the phenomenon being studied is better represented by a continuum, and discretizing it into subgroups by these methods may be artificial. However, the articles in this special issue show that these methods can be used to identify qualitatively different subgroups with multidimensional indicators that do not simply fall into a continuum from low to high. For example, Trezise and Reeve (2018) categorized both the performance and the anxiety of students solving math problems. Their findings on performance illustrate the ability of LCA to find nonlinear patterns. There were one high, two middle, and one low performance classes, in terms of accuracy, but response time was found to be slower in the intermediate performance class compared to the high and low classes. This is reasonable, as an overwhelmed student might guess at an answer, whereas a well-prepared student might find the answer without much thought. The usual approach of looking at correlations between linear dimensions would not have detected these differences. As another example, Koppenol-Gonzalez, Bouwmeester, and Vermunt (in press) examine children's changes in their use of processing strategies (e.g., verbal vs. visual) in performing short-term memory tasks over the course of a year. They find that although there was much variability, there did not seem to be an orderly, stage-wise progression from a lower to a higher form of processing as the children matured. Instead, children followed a variety of different paths from one strategy to another on repeated testing occasions, with little evidence of systematic change. The differences between children in their use of strategies are due partly to differing cognitive maturational levels, but also due partly to either differences in personal preference or perhaps to haphazard changes from day to day. The authors point out that this supports the importance of a flexible approach to teaching that combines visual and verbal information, so that different learners can use strategies in ways that works best for them.

Arguments about the identification of qualitatively different subgroups are related to the classic and continuing debate (McLachlan & Peel, 2000; Sher, Jackson, & Steinley, 2011) alluded to by Hickendorff et al. (2017) regarding reification of latent classes. That is, whether latent classes and trajectories are really best viewed as qualitatively distinct entities or are just a convenient way to divide an underlying continuum into a manageable number of zones. Hickendorff et al. find present latent variable mixture models to be useful in either case, but especially interesting in cases where a real, qualitative difference is of theoretical interest. Some researchers have pointed out that categorical views of behavior can obscure underlying commonalities or continua (Rutter, 2011) and could potentially lead to harmful labeling of individuals (Walters, 2011). It has also been argued that having too many latent classes can hurt statistical power and precision by requiring too many comparisons of small groups (Piasecki, Jorenby, Smith, Fiore, & Baker, 2003). These are valid concerns. However, other authors say, and we believe, that typologies can be helpful for simplifying and interpreting data if they are used conscientiously, regardless of whether classes are considered to be truly distinct entities or only abstractions (Nagin & Tremblay, 2005). Approaches that consider only one or two variables at a time can fail to identify complex ways in which variables interact—one of the reasons why latent variable mixture models are referred to as "person-centered" approaches (Lanza, Rhoades, Greenberg, Cox, and The Family Life Project Key Investigators, 2011).

In addition, one of the most interesting possibilities Hickendorff et al. (2017) raise for latent variable mixture models is the potential opportunity to reconcile contradictory theories and findings, because it may be that different theoretical models apply to different subgroups of individuals. Identifying subgroups of individuals for whom different theories apply avoids the need to specify a single, homogeneous theory for all individuals. Methods such as LTA are particularly well-suited to testing single theories about stage-sequential development, but also to comparing competing theories because individuals can take different paths through development. Although the articles in this special issue did not explicitly test competing theories about learning development, they did illustrate the strengths of LTA to examine learning development. For example, Edelsbrunner, Schalk, Schumacher, and Stern (in press) and Flaig et al. (in press) examined students' progress through latent classes representing less scientifically accurate to more scientifically accurate interpretations of phenomena. In contrast to Koppenol-Gonzalez et al. (in press), in which multiple strategies could each be seen as useful, the interpretational strategies of these studies could be ranked ordinally, with some reflecting more accurate understanding than others. However, the authors still found the idea of these being qualitatively distinct classes to be fruitful, because scientific learning does not only involve accumulating facts but also rejecting common misconceptions (such as weight determining whether an object sinks) and forming new and more insightful mental models (such as density determining whether an object sinks). In general, both studies found that different students followed somewhat different paths, sometimes moving through periods of fragmented or inconsistent thinking, but that successful students generally moved in a direction of interpretations that gave them the ability to make more accurate empirical predictions. The result is a model of learning that is somewhat like a classic model of stage-sequential development, but does not require that everyone follow the same pathway.

In addition, Stevenson and Hickendorff (2018) examined the progression of students' performance on an analogical reasoning task over the course of study and instruction. They described a series of distinct phases through which children passed as they learned the concept of analogy and showed that practice and coaching can help this process. In the simplest phases, children either duplicated the provided stimuli or