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

Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis[☆]

Marian Hickendorff[a,*], Peter A. Edelsbrunner[b], Jake McMullen[c], Michael Schneider[d], Kelly Trezise[e]

[a] Leiden University, Leiden, The Netherlands
[b] ETH Zürich, Zürich, Switzerland
[c] University of Turku, Turku, Finland
[d] University of Trier, Trier, Germany
[e] University of Melbourne, Melbourne, Australia

A R T I C L E I N F O

Keywords:
Person-centered methods
Latent variable mixture models
Individual differences
Learning pathways
Developmental patterns

A B S T R A C T

This article gives an introduction to latent class, latent profile, and latent transition models for researchers interested in investigating individual differences in learning and development. The models allow analyzing how the observed heterogeneity in a group (e.g., individual differences in conceptual knowledge) can be traced back to underlying homogeneous subgroups (e.g., learners differing systematically in their developmental phases). The estimated parameters include a characteristic response pattern for each subgroup, and, in the case of longitudinal data, the probabilities of transitioning from one subgroup to another over time. This article describes the steps involved in using the models, gives practical examples, and discusses limitations and extensions. Overall, the models help to characterize heterogeneous learner populations, multidimensional learning outcomes, non-linear learning pathways, and changing relations between learning processes. The application of these models can therefore make a substantial contribution to our understanding of learning and individual differences.

1. Introduction

Learning research often seeks to characterize patterns and pathways of learning or development. Many learning theories emphasize both qualitative and quantitative differences in learners' knowledge, skills, and strategies at a specific point in time. Furthermore, learning pathways are often discontinuous or non-linear: Learning can take place in stages, learning pathways can vary substantially between learners, and learning can interact with learner abilities and characteristics (e.g., Carey, 2009; Meiser, Stern, & Langeheine, 1998; Van der Maas & Molenaar, 1992). For example, conceptual knowledge research shows qualitatively different mental models between children, and demonstrates that children differ in their transitions between concepts over time (e.g., Carey, 2009; Kleickmann, Hardy, Pollmeier, & Möller, 2011; Schneider & Hardy, 2013; Smith, Carey, & Wiser, 1985; Vosniadou & Brewer, 1992). To fully characterize learning processes research therefore needs to account for both quantitative and qualitative individual differences at a specific measurement point as well as in change over time. Unfortunately, traditional analytical approaches have limited capabilities to accomplish these goals.

In this article, we outline a set of analytical techniques that are highly useful for this purpose: Latent class and latent profile analysis, and their longitudinal extensions, latent transition analysis. Latent class analysis (for categorical variables) and latent profile analysis (for continuous variables) are used to trace back the heterogeneity in a group to a number of underlying homogeneous subgroups, at a specific measurement point. These techniques have been applied in various domains of learning, for instance in adolescents' literacy (Mellard, Woods, & Lee, 2016), homework behavior (Flunger et al., 2017), and undergraduate science education (Romine, Todd, & Clark, 2016). In the longitudinal extensions of latent class and latent profile analysis, a transitioning component is added to reflect changes in learners' subgroup membership over time, representing potentially non-linear learning pathways. These models have been applied for instance to first-year university students' learning pathways (Fryer, 2017), and to the identification of English language learners at risk for reading disabilities (Swanson,

[☆] The first author took the lead in drafting this article and in outlining the structure. All other authors contributed equally and are listed in alphabetical order.

[*] Corresponding author at: Education and Child Studies, Leiden University, PO Box 9555, 2300 RB Leiden, The Netherlands.
E-mail addresses: hickendorff@fsw.leidenuniv.nl (M. Hickendorff), peter.edelsbrunner@ifv.gess.ethz.ch (P.A. Edelsbrunner), jamcmu@utu.fi (J. McMullen), m.schneider@uni-trier.de (M. Schneider), kelly.trezise@gmail.com (K. Trezise).

https://doi.org/10.1016/j.lindif.2017.11.001
Received 1 December 2016; Received in revised form 23 October 2017; Accepted 2 November 2017
1041-6080/

2017).

The current paper aims to familiarize researchers in the domain of learning and individual differences with this family of techniques and to illustrate how they can make a substantial contribution to our understanding of learning and individual differences. Note that there are other introductions available, tailored to clinical research (Collins & Lanza, 2010), pediatrics (Berlin, Williams, & Parra, 2014), and developmental research (Kaplan, 2008; Lanza & Cooper, 2016). In the following, we will first elaborate on the usefulness of these models for learning research in comparison to the limits of more common analytic techniques. Then we will give more details about the four types of models that are central to the current paper: latent class analysis, latent profile analysis, latent class transition analysis and latent profile transition analysis. Next, we will discuss the current best practices in application of these techniques to empirical data, addressing several practical and statistical issues that researchers frequently encounter. The concluding remarks will summarize the usefulness of these approaches in learning research. While some basic knowledge of latent variable models may be helpful to understand the present article, the main goal here is to introduce the relevance of these models without expanding too far into the details of the statistical makeup.

2. Learners and learning: a person-centered approach

Inter-individual differences represent an important but complex issue for educators and learning researchers (Snow, 1986). Learners differ in their abilities, motivations, and preferences, which often interact while affecting their learning. Oftentimes, an "average" learning pattern is not an adequate description for many learners because this ignores the unobserved heterogeneity between learners. Assessing and modeling the heterogeneity that may arise from the complex interplay between abilities, motivations, and preferences is important for understanding how, and under which, circumstances learning takes place. In these cases, researchers may find latent class, latent profile, or latent transition analyses to be useful to more appropriately model the unobserved heterogeneity between and within individuals. These techniques constitute a powerful and informative toolbox to examine different subgroups of learners in cross-sectional data and different pathways of learning in longitudinal data.

By contrast, traditional analytical approaches from the general linear model such as ANOVAs, correlation and regression-based techniques, and factor analysis have serious limitations in appropriately characterizing heterogeneity and complex, non-linear learning patterns. These common analytical approaches are variable-centered, emphasizing the relations between variables (Bergman, Magnusson, & Khouri, 2003; Collins & Lanza, 2010). They assume that the relation between variables can be applied to all learners in the same way: in other words, that there is homogeneity in the nature of the individual differences (Bergman & Magnusson, 1997; Collins & Lanza, 2010). These linear techniques are thus restricted to quantitative individual differences, assuming that learners differ quantitatively in the amount of something, but not qualitatively (Lanza & Cooper, 2016; Sterba & Bauer, 2010). For example, in learning research it is common to perform statistical analyses on sum scores from learning measures. The use of sum scores implies the assumption of homogeneity in response patterns, and any heterogeneity – between individuals and within individuals – is primarily considered statistical noise. As a consequence, our understanding of learning processes is a general model that describes the average behavior of a sample. If qualitatively different subgroups exist within a population, they are not accurately represented by the general model.

One means of getting around the use of continuous measures to examine differences in learning processes is with arbitrary cut-off points (e.g., median-splits). The resulting groups often represent ability levels such as "high" and "low", and differences between the groups are then explored to infer learning differences. While the comparison of ability groups can provide a useful method of understanding implications of different abilities on learning outcomes, the use of arbitrary cut-off points is considered poor statistical practice: it is a-theoretical and introduces error which can result in a distorted picture of relations between variables (Altman & Royston, 2006; Irwin & McClelland, 2003; Maxwell & Delaney, 1993). Consequently, arbitrary cut-off points are never appropriate and should be avoided.

The most important limitation of variable-centered methods is their inability to deal with heterogeneity within and between individuals. Another constraint of *linear* variable-centered methods is their inability to accurately characterize non-linear and interactive patterns (Bergman et al., 2003). Consequently, the use of linear variable-centered analyses impedes our ability to test theoretical claims of learning that do not meet these assumptions, such as when heterogeneous patterns, discontinuous change, or interacting and changing relations between two or more learning processes are present. Although some non-linear variable-centered methods exist that allow analyzing some non-linear learning patterns and pathways, these are still limited to general patterns for the entire population. By contrast, person-centered approaches are not restricted to linear patterns and can model heterogeneity as well. Person-centered analyses place the emphasis on the individual, in order to account for heterogeneous patterns of variable interactions; "operationally, this focus often involves studying the individuals on the basis of their patterns of individual characteristics that are relevant for the problem under considerations" (Bergman & Magnusson, 1997, p. 293). In the present case, the problem under consideration is knowledge and learning.

The aim in learning research is rarely to describe a single learner, but rather to describe general patterns of learners' behavior and learning pathways. Understanding these patterns and pathways can enable educators to better understand why some learners are more successful with learning and some experience particular difficulties, or this understanding can be used to inform targeted learning interventions. The strength of person-centered approaches is that they can capture these different patterns and pathways, by identifying homogeneous subgroups of learners that exhibit similar patterns of characteristics (Bergman & Magnusson, 1997). Traditional clustering methods like *K*-means clustering (e.g., Kaufman & Rousseeuw, 2009) provide one approach to examine such subgroups. The family of model-based clustering methods that latent class and profile models belong to, however, have specific advantages over traditional cluster techniques. These models are more flexible, account for measurement error, and are able to handle longitudinal data (e.g., Magidson & Vermunt, 2002, 2004; Oberski, 2016; Vermunt, Tran, & Magidson, 2008). Most of the work on developing these models and estimation procedures has, in fact, been completed by statisticians and methodologists in the social sciences starting in the 1960s (e.g., Goodman, 1974; Lazarsfeld & Henry, 1968; McCutcheon, 1987; Wiggins, 1973). More recently, the approach has been adopted by psychology and education researchers to examine learners and learning processes.

2.1. What is a latent class or latent profile model?

The aim of latent class and latent profile models is to trace back heterogeneity in a population to a number of existing but unobserved subgroups of individuals, which are referred to as latent classes. The analyses are based on a set of observed variables that can be categorical and/or continuous. The classes are formed such that there is as much similarity within a class while at the same time as much differences between the classes as possible (Lanza & Cooper, 2016). The identification of these latent classes can be useful for characterizing qualitative differences between learners, which may be missed with traditional analytic approaches. For example, Fig. 1 depicts outcomes of two analytic approaches to the same example data with two variables: accuracy and response time on a particular measure. Note that the advantages of latent class and latent profile models are more pronounced