Quantitative methods for capturing processes and contexts in educational research

Métodos cuantitativos para el registro de procesos y contextos en la investigación educativa

Lars-Erik MALMBERG, PhD. Associate Professor. University of Oxford, UK (*lars-erik.malmberg@education.ox.ac.uk*).

Abstract:

Technological and methodological advances enable new substantive research questions to be posed, and new study designs to be implemented, in educational research. In this paper I review emerging methods relevant for capturing learning and teaching processes over time —the sequences of learning events— which take place in multiple contexts.

To do so, the concepts of *nomothetic* and *ideographic* research are traced through the use of Cattell's (1952) cube, posing persons, variables and time as the three key dimensions for determining study-designs. For educational research, a fourth dimension —context— is important to consider given the nested structures (e.g. student-teacher dyads, peer-relations, student-groups, classrooms, teachers, and schools) learning and teaching occurs in. Several developments of quantitative methods enable researchers to a) establish quality of measurement (e.g. factor analysis, item response models), b) across sequences of time-points (e.g. autoregressive models), c) in complex multilevel structures (e.g. multilevel models, random effects models), also using estimators which are robust for small-n studies (e.g. Bayesian models). Educational researchers are encouraged to design studies fitting multilevel models for hierarchically and cross-classified data, and to think in terms of intraindividual learning processes.

Keywords: educational research, quantitative methods, statistical models, multilevel model, intensive longitudinal data.

Resumen:

Los avances tecnológicos y metodológicos permiten formular nuevas preguntas de investigación fundamentales y aplicar nuevos diseños de estudios en la investigación educativa. Este artículo revisa los métodos emergentes empleados para el registro de los procesos de aprendizaje y enseñanza en el tiempo —las secuencias de eventos de aprendizaje— que tienen lugar en contextos múltiples.

Revision accepted: 2018-15-04.

Cite this article as: Malmberg, L. (2018). Métodos cuantitativos para el registro de procesos y contextos en la investigación educativa | *Quantitative Methods for Capturing Processes and Contexts in Educational Research. Revista Española de Pedagogía, 76* (271), 449-462. doi: https://doi.org/10.22550/REP76-3-2018-03

https://revistadepedagogia.org/ ISSN: 0034-9461 (Print), 2174-0909 (Online)

Para este fin, se emplean los conceptos de investigación *nomotética* e *ideográfica* utilizando el cubo de Cattell (1952), que identifica a las personas, las variables y el tiempo como las tres dimensiones clave para describir diseños de estudios. En la investigación educativa es importante tener en cuenta una cuarta dimensión —el contexto— dadas las estructuras anidadas (p. ej. díadas alumno-profesor, relaciones entre pares, grupos de alumnos, aulas, profesores y colegios) en las que se produce el aprendizaje y la enseñanza. Existen varios métodos cuantitativos que permiten a los investigadores: a) determinar la calidad de la medición (p. ej. el análisis de factores, los modelos de respuesta a ítems), b) en secuencias de puntos temporales (p. ej. modelos autorregresivos), c) en estructuras multinivel complejas (p. ej. modelos multinivel, modelos de efectos aleatorios), empleando también estimadores sólidos en estudios de n pequeña (p. ej. modelos bayesianos). Se invita a los investigadores en educación a diseñar estudios apropiados para modelos multinivel con datos clasificados jerárquicamente o con clasificación cruzada, y a pensar en términos de procesos de aprendizaje intraindividuales.

Descriptores: investigación educativa, métodos cuantitativos, modelos estadísticos, modelo multinivel, datos longitudinales intensivos.

1. Introduction

One key aim of educational research is to investigate how students learn, and how their learning can be supported. Learning occurs in *processes*, that is in sequences of learning situations. In these learning situations individuals experience different levels of challenge, expectations, engagement, understanding, and positive and negative affect. Appropriately collecting and analysing process data (cf. micro-longitudinal, intensive longitudinal, intraindividual data) that captures such sequences of situations, gives insights into the shape and form of students' learning experiences. After establishing the shape, form and variation of the processes, researchers can then investigate how instruction contributes to the learning process (Schmitz, 2006). Instruction is complex, constituted by multiple forms of interaction with learning contents and instructional formats when studying solo, and through student-teacher and student-peers interactions. The interactions are also shaped depending on the organisation of student-groups, classrooms, teachers, and schools the learning and teaching occurs in. The aim of this paper is to provide an overview of emerging quantitative methods for the analysis of processes in contexts.

We live in times where statistical developments are rampant, freeware are mushrooming, datasets are growing larger, computers are getting faster, and artificial intelligence outperforms humans on narrowly defined tasks. Methodological advances spur us to pose new research questions, come up with new study designs, or revisit existing research questions with revitalized energy. Research questions which are difficult to answer with existing methods spur further methodological development.

This codependence between findings appropriate methods for unanswered questions, and revisiting old findings with new methodological tools provides opportunities for methodological-substantive synergies (Marsh & Hau, 2007). I provide an illustration of state-of-the art quantitative methods which enable investigations of process in contexts. But why process? Why context?

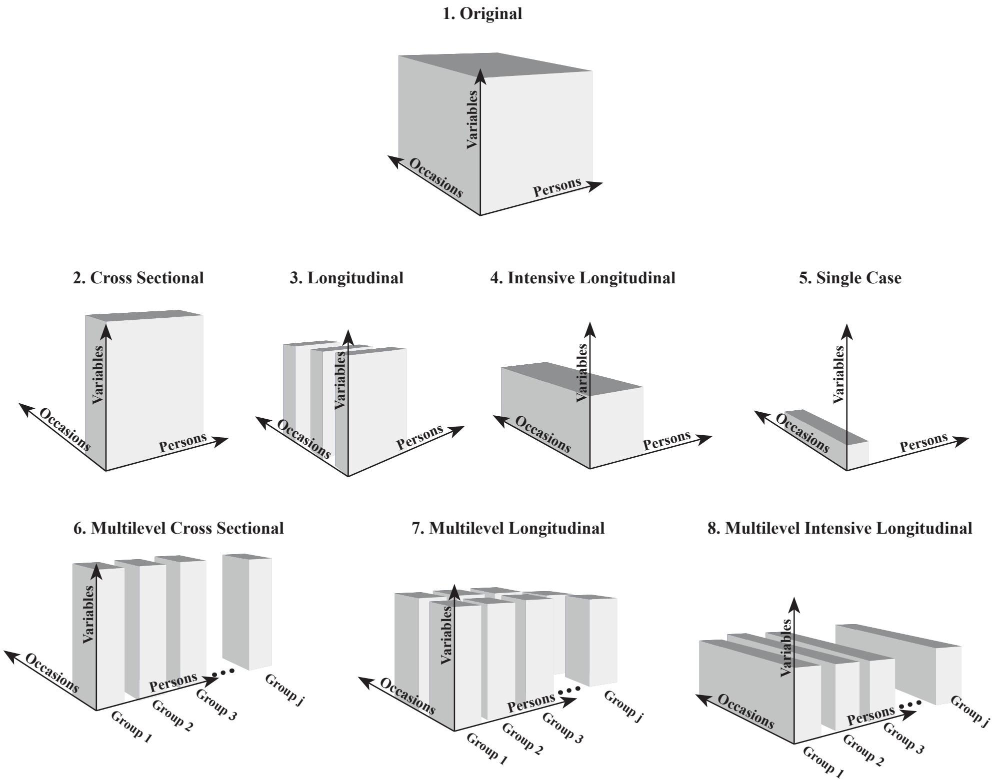
2. Persons, variables, occasions, and contexts

Going back to the turn of the 20th century, Windelband in 1894 defined *nomothetic* research as the search for «what always is», that is generalizable knowledge (Lamiell, 1998, p. 27). This knowledge included general laws, as in the natural sciences, and its applications in medicine and psychiatry. Research in natural sciences is often designed at the researchers desk (e.g. derivations of equations), and are later put to the empirical test. Case-studies, particularly in medicine and psychiatry, have long informed the search of generalizable knowledge. The term *ideographic* was used in the search of «what once was» as in the humanities (Lamiell, 1998, p. 27). These were the study of singular events that would not necessarily be repeated. Following the work of Galton, Pearson, Fisher, Allport and other large-sample survey researchers, the pursuit of the average (aggregate or the «general») was perceived as the norm (see Lamiell, 1998, p. 32). The critique against nomothetic studies was that the aggregate is not always true of what applies in general. *Ideographic* research came to be equivalent with small sample intraindividual research. The schism between the two has been fueled also by discussions of distinctions between *qualitative* and *quantitative*. It is time to move beyond such false dichotomies.

For research designs Cattell (1952) devised his three-dimensional cube, consisting of three axes to represent three key dimensions of research designs: persons, variables and time-points (see top of Graph 1). The balance between these three dimensions depend on the type of study that is designed. The (1) original cube (on top) has been transformed to illustrate (2) a cross-sectional design (i.e. multiple variables of multiple persons at one time point). The face of the resulting rectangular block on the occasion-axis for this design is relatively slim, as there is only one time-point, but wide on the person-axis. (3) A longer term longitudinal study here with three time-points is depicted as three relatively slim occasions, and relatively fewer persons than in the cross-sectional design. (4) The intensive longitudinal design has relatively more time-points per person than a longitudinal study, but data are collected within a relatively shorter time-window and of fewer persons. (5) The single case case-study is the *narrowest* type of design we can pose, one participant providing relatively little information at each time, but for a relatively large number of time-points.

spanish journal of pedagogy
year LXXVI, n. 271, September-December 2018, 449-462

GRAPH 1. Modern designs in educational research following Cattell's (1952) original cube. The three dimensions in the original (top) were x = persons, y = variables, and z = occasions. Multilevel structures are included on the person-axis for modern designs, in cubes 6-8.



Source: Own elaboration.

Context can be incorporated in the cube in two ways. The first option is to include variables that capture the context (or experience of the context) at each occasion; for example, a student's perception of the teacher, the level of support he or she receives, or task difficulty. The second option is to consider the nested nature of institutionalized education as a 4th dimension in the cube, namely «context». Today's trends in research designs are a mixture of number of individuals, variables, time-points and contexts. This is illustrated by (6) a multilevel cross-sectional study, in which persons are nested within contexts, here students in schools. This is illustrated with students grouped in sub-sections of the original cube, i.e. «Group 1» to «Group j». Likewise (7) longitudinal and (8) intensive longitudinal studies are depicted to be nested within groups. The single-case case-study (cube 5) can be expanded to a multiple-case design by adding more persons (not depicted). As will be discussed shortly, this *simple* multilevel structures can be expanded to include higher levels (e.g. students in schools in countries), or more complex structures. Examples of more complex structures are when persons belong to multiple contexts simultaneously (e.g. students belonging to both schools and neighbourhoods) or cross-classified (e.g. students belonging to different student groups for different school subjects). What about experimental designs then?

In educational research we distinguish between non-experimental and experimental designs. The experimental design is essentially repeated measures designs with an experimental intervention between at least two of the time-points. This design can be thought of as a *classical* longitudinal design (cube 3 in Graph 1) with two groups, but can be designed as an intensive longitudinal study (cube 4) or single-case interventions (cube 5). Real-world experimental and intervention designs can incorporate all features of multilevel longitudinal or intensive longitudinal designs. The multilevel nature of both experimental and non-experimental designs has important implications for our statistical toolbox, for power calculations and estimation of effects at higher levels in the hierarchical structure. Advances in quantitative methods are being made for all designs depicted in cubes 5-8. For the purpose of investigating processes in context, we can learn much from methodologies which have been developed in large-scale multilevel cross-sectional designs (cube 6), particularly on making inference to the underlying populations data stem from, and for quality of measurement. Longitudinal (cube 3) and multilevel longitudinal (cube 7) designs can inform us on methods which take repeated measures into account, as well as quality of measurement. Intensive longitudinal (cube 4) and multilevel intensive longitudinal designs (cube 8) inform us about models for time-points nested in persons, and the role of conceptualization of time. Single and multiple-case case-study designs (cube 5) inform us about methods which are particularly well suited for small-n designs with rich occasion data.

3. Contextual and random effects and hierarchical structures

Multilevel strategies have evolved rapidly within the traditions spurred by Robinson's studies of contextual effects (Robinson, 1950), and Henderson's random effects between parents and offspring (Henderson, 1982) (see section on intraindividual research). A key issue in the 1950s was how contexts had an effect on the individual, and how to account for this in statistical models, so that inferences could be made at the appropriate level to avoid individual fallacies (i.e. attributing a contextual effect to the individual) and ecological fallacies (i.e. attributing an individual effect to the context). The multilevel model was rapidly developed based on the general linear model (GLM), which made school effectiveness research blossom. The baseline school effectiveness study includes a sufficient number of schools for drawing inference about the effects of schools on individuals. The models then estimate the gain in students' achievements by including concurrent achievement as outcome controlling for prior achievement (for an overview see Goldstein, 1997). Other modern studies apply the logic of hierarchically nested structures to both data-collection and analyses.

More and more countries join international comparisons. In 2015, for example, more than half a million students in 72 countries took part in the science-test of the Program for International Student Assessment (PISA). The international comparison programmes have developed multi-stage sampling procedure and advanced procedures for weighting the data for drawing inference for results at the country level (i.e. comparisons of country-means and differential effects of various covariates). A debate about the methodological advances on how to best establish structural validity of both test-scores and self-reported constructs (e.g. motivation) is in motion, particularly on how to establish cross-national compatibility of the measures accounting for the multilevel structure.

Many methods for establishing the test-scores (e.g. literacy, numeracy, science) have been proposed; for example, «plausible values» (von Davier, González, & Mislevy, 2009), based on the Rasch-model, that is, the one parameter logistic model (1PL). Additional models are the 2PL (which includes an item discrimination parameter) and the 3PL model (which corrects for guessing). There are difficulties to establish cross-national equivalence of test-scores and self-report constructs. Models in which factor structures have the same factor loadings across nations (i.e. weak invariance) typically fit better than models in which the same mean-structure is imposed (i.e. strong invariance). Whether it is the models or our assumptions about invariance that are too strict is also argued. For example (Scherer, Nilsen, & Jansen, 2016), demonstrated that strict invariance (i.e. equal factor loadings, means and residuals) was possible to achieve among the Anglosaxon countries but not in all countries of the PISA studies. Marsh and colleagues proposed the extended alignment method for multi-group analysis (Marsh et al., 2017).

In addition to large-scale international comparisons and school effectiveness studies, other studies investigate the proximal structures students are embedded in their daily learning. Using models for dyad-analysis (e.g. parent-child, husband-wife, actor-partner) Mainhard and colleagues went beyond a model in which students are nested in teachers, to inspecting how teachers vary across student-groups they teach, at the same time as student-groups have different composition with their different teachers (Mainhard, Oudman, Hornstra, Bosker, & Goetz, 2018). They found considerable variance in student-teacher relationships and teacher-student-group relationships. So far such models have been carried out using manifest variables.

There is a growing number of software that can accommodate multilevel designs (i.e. students nested in schools nested in countries). MLWin (e.g. Rasbash, Steele, Goldstein, & Browne, 2017) has long been able to handle multilevel structures for three or more levels, and complex cross-classified structures. The more complex structures are enabled thanks to the Multiple Chain Monte Carlo (MCMC) algorithm. MLWin also enables interaction with the R, using the R2MLwiN package. For modelling of latent constructs in such complex structures, Mplus can handle three levels of multilevel analysis (or two levels and one cross-classification, Muthén, 1994 and xxM even more (Mehta, 2013). When modelling two-level data (or more) there are three common ways of treating the effects at multiple levels. First, when there are enough units at the higher level of data —more units than parameters we want to estimate— the higher level can be modelled efficiently (see e.g. Morin, Marsh, Nagengast, & Scalas, 2014). Second, the standard errors of the parameter estimates are adjusted for the nested structure of the data. Third, many researchers opt to go for a Bayesian estimation (e.g. Praetorius, Koch, Scheunpflug, Zeinz, & Dresel, 2017), as a sparser number of higher-level units can still be sufficient for robust estimates.

4. Longitudinal designs

Longitudinal educational research converges with the study of human development. Researchers painstakingly collect follow-up data in cohort studies (e.g. a birth-cohort study follows up a representative sample of newborns a number of times, usually spaced in years), prospective longitudinal designs (e.g., one group followed up over time), or cross-sequential or accelerated designs (e.g. two or more cohorts are followed up over time). Many such designs are timed to start prior to an educational transition and end after such a transition (e.g. transition from kindergarten to primary, or primary to secondary school). The number of time-points at which individuals are followed up varies once a year, twice a year (e.g. in autumn and spring; e.g. Skinner, Zimmer-Gembeck, & Connell, 1998) or more often.

5. Experimental longitudinal designs

In the classical test-retest paradigm in experimental and intervention research, researchers have applied repeated measures analysis of variance (ANOVA), analysis of covariance (ANCOVA), or some extension of a general linear model (GLM). These models can be extended to include random effects in either multilevel models, or mixed models to accommodate multilevel structures. Switching to a mixed model using long data (e.g. time-points nested in persons) instead of wide data (i.e. one row per participant), enables us to analyse experimental data much as we would analyse longitudinal data, but adding a key predictor, the experimental condition(s). In intervention studies where the randomization is at the school level, the nested structure can be modelled as a two level path-model (Rakoczy et al., 2018).

6. Autoregressive and reciprocal effects

Modelling of longitudinal data can be done with (1) different variables (constructs) at different time-points, or (2) with the same variable (construct) repeated at each time-point. The former type of design would be commonplace in studies of early childhood education as the instruments for capturing infant development capture qualitatively different phenomena at different time-points. Modelling would entail verifying the quality of measurement of each construct, and then proceed with SEM models of hypothesized paths. When observations are nested in child-care centres, neighbourhoods or other higher-order units, such structures can be modelled in multilevel models.

The latter —repeated administration of the same measure— would be more commonplace for studies of changes over time through the primary school years. Researchers would first establish the equivalence of the factor structure of the psychometric constructs across the time-points (i.e. so we know that the same phenomenon is measured at all three time-points). In case the construct is a measure of a phenomenon that changes over time (e.g. cognition) by including easier items earlier on, more difficult items later on, and a sufficient number of overlapping items across the time-points (i.e. so called *anchor* items), then it is possible to test whether the continuum of cognitive growth can be established across the time-span of the study.

After establishing the quality of measurement one would then proceed to testing the autoregressive paths (i.e. the associations between the variable at Time T and at the preceding time-point T-1) (Little, 2013). These paths indicate the rank-order stability of individuals at T-1 and T, higher values indicating that the rank-order is maintained across the time-points. Lower values indicate more fluctuations (e.g. individuals can overtake each other) and developmental discontinuities. After establishing the autoregressive effects researchers then proceed to test whether one of the constructs affects change in the other construct over time. These can provide important insights into the directionality of the effect. Is for example, AT -1 more likely to predict change in BT controlling for BT -1, or the other way around?

From the change-score model of two time-points we can derive the growth model in which we can estimate changes of time at the group-level, and also individual differences in change over time (Singer & Willett, 2003). In the growth model the researcher hypotheses what the shape of change will be over time, linear (i.e. a straight line), quadratic (i.e. a line with a curvature), cubic (i.e. a roller coaster ride), or non-linear (Ram & Grimm, 2007). The growth model can include also autoregressive parameter combining analyses of both the mean-structure of the data and the structural relationships.

Given the complex nature of institutionalized education recent studies have included more than two time-points (i.e. growth of student achievement over three or more time-points) and multiple cohorts in accelerated growth models (Ortega, Malmberg, & Sammons, n.d.). As schools (being inanimate objects) cannot actually *do* anything to students, switching focus to teacher-effects is a reasonable thing to do. Teacher effects include complex cross-classifications or multiple membership statuses as teacher of a student can change over time (Ortega, Malmberg, & Sammons, 2018).

Other complex structures are reviewed in Duncan, Duncan, & Strycker, 2006, for example, the family-of-curves and curves-of-a-family models.

7. Micro-longitudinal designs

Micro-longitudinal designs have become increasingly popular given the relative ease at which data can be collected (Hamaker & Wichers, 2017). Data-collection points can be set randomly as in experience sampling, at fixed interval as in ecological momentary assessments, or event driven (e.g. the GPS recognises you are in the library and asks you to report on your learning experiences in the library). The resulting real-time data enhances closeness between an event and the report of the event, reducing retrospection bias (Walls & Schafer, 2006). It also enables the researcher to ask about experiences of the context in which the report takes place. Contextual variables can be organized in two ways. First, contexts can be thought of as levels of the hierarchical structure (e.g. time-points nested in students, nested in teachers, or classrooms). Second, contextual information can also be organized as intraindividual variables, such as individuals perceptions of the context: task difficulty, perception of classroom climate, or interaction with the teacher (Malmberg, Lim, Tolvanen, & Nurmi, 2016). Many studies model intra-individual data as «individuals as their own controls» using the multilevel model with time-points nested in persons (Murayama et al., 2017).

8. Experimental micro-longitudinal designs

Two types of models are applied to micro-longitudinal data. The first type is the random effects general linear models that grew out of the genetic modelling of Henderson (e.g. Henderson, 1982). Early models investigated genetic relatedness between offspring and parent generations (sires and dames) by applying random effects models. Also this methodology grew out of the GLM. Experimentalists have incorporated these models, particularly for the study of reaction time of various cognitive tasks. These models, common in experimental psychology, can incorporate complex random effects (Bates, Mächler, Bolker y Walker, 2014; Matuschek, Kliegl, Vasishtha, Baayen, & Bates, 2017).

Diary studies of micro-longitudinal data have been carried out (Perels, Gurtler, & Schmitz, 2005). While micro-longitudinal data collection is relatively rare in experimental designs, there are several ways in which aggregates of intraindividual variability could be assessed (Malmberg et al., 2016), or associations (couplings) between variables could change as a function of the intervention (Schmitz, 2015). As engaged students have been found to have low intraindividual variability (i.e. small magnitude of change from one moment to another) in competence beliefs and intrinsic motivation, a two-time-point study in which intraindidivual data would be collected at both time-points could unravel if intra-individual variability had diminished (i.e. less oscillations between low and high motivation across the time-points) as a result of the intervention. An intervention could then focus on decreasing the fluctuations in perceptions and beliefs. An alternative could be to increasing the synchronicity between beliefs, for example intentions (goal-setting) the previous lesson with self-regulated learning (goal-accomplishments) the subsequent lesson (Schmitz, 2015). An intervention could be designed as a within-person encouragement and the outcome could be a stronger magnitude between joy and working memory across time-points (Schmiedek, 2016). Interventions could then be carried out with single-case or multiple-case ABAB designs (Walls, Barta, Stawski, Collyer, & Hofer, 2013), A indicating control or «business as usual» and B sequences when the participant receives encouragement prompts. Single-case designs utilizing such burst designs are on the increase (Kratochwill & Levin, 2010). One methodology for single-case designs is the Bayesian unknown change-point model to investigate and quantify immediacy (Natesan & Hedges, 2017). Complex methodologies require complex preparation. For these types of intraindividual (and cross-classified) designs power calculations and effect size estimation, requires advanced computations or simulation studies (Moerbeek & Teerenstra, 2016).

9. How to treat the time-variable?

The scaling of time in longitudinal research has received attention, as the assumption of auto-regressive models is equidistant time-points. Time can be coded in two ways, discrete time (e.g. time-point 1, time-point 2, time-point t) or continuous time (e.g. 09:15, 09:45, 10:30, 11:40, 13:20…). For data-collection in which participants are observed, or asked to complete a questionnaire, within a certain time-frame, or deliberately, continuous time analysis could be an important way to *correct* the irregular time-intervals of discrete time-coding. Such models can be set up as flexibly coded growth models in which time is coded on the scale of interest, or by using the ctsem package in R. Parameter estimates can thus be adjusted for the time-intervals used. This enables both modelling of longitudinal data and time-series data with parameter estimates correctly taking the time lags into consideration (Voelkle, Oud, von Oertzen, & Lindenberger, 2012).

Another way of thinking of time-dynamics is using differential equation models, in which we can model the rate of acceleration (Deboeck, 2013). Think of a situation in which a student gets off-task and we want to estimate how fast he or she returns to on-task behavior. Alternative methods for working with intensive longitudinal data is functional data analysis.

10. To Bayes or not to Bayes?

Bayesian techniques are strongly on the rise with applications in commercial and freeware using Bayesian estimation (Kaplan & Depaoli, 2012; Muthén & Asparouhov, 2012; Van de Schoot et al., 2014). The Bayesian technique estimates the probability of a parameter given the data p (θ | data), rather than the probability of the data given the model p (data | θ) (i.e. null-hypothesis significance testing, NHST). It is suitable for situations when Maximum Likelihood might be underpowered, as it does not rely on large-sample theory. It is suitable for complex models such as SEM as it does not converge at improper estimates (e.g. negative residuals, correlations above 1, Zitzmann, Lüdtke, Robitzsch, & Marsh, 2016).

11. Conclusions and reflections

The aim of this overview paper was to inspect how processes and contexts can be investigated using state-of-the-art methodologies. Process-models of sequences of learning events require models that can incorporate time-points within students and can take the flow of time into consideration in the model. Learning contexts prove complex places. These are possible to conceptualize in terms of hierarchies, e.g. time-points nested in students, nested in classrooms, nested in schools, but also according to relationships and interactions: student-teacher, teacher-student-group, and student-peers relationships and interactions, student-groups. Models for complex (e.g. cross-classified, multiple memberships) structures require suitable estimators (e.g. MCMC, Bayesian). There are several examples in the literature in which complex multilevel and cross-classified structures are modelled using manifest scales as dependent variables. We anticipate the next generation of models in which latent variables can be used as dependent variables.

References

Allport, G. W. (1937). *Personality. A psychological interpretation*. London: Constable and company.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67* (1). doi: https://doi.org/10.18637/jss.v067.i01

Cattell, R. B. (1952). *Personality and motivation. Structure and measurement*. Yonkers-on-Hudson, NY: World Book Company.

Deboeck, P. R. (2013). Dynamical Systems and Models of Continuous Time. In Todd D. Little (Ed.), *The oxford handbook of quantitative methods in psychology: Vol. 2: Statistical analysis*. Oxford: Oxford University Press. doi: https://doi.org/10.1093/oxfordhb/9780199934898.013.0019

Duncan, T., Duncan, S., & Strycker, L. (2006). *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Application*. Mahwah, NJ: Lawrence Erlbaum.

Goldstein, H. (1997). Methods in School Effectiveness Research. *School Effectiveness and School Improvement*, *8* (4), 369-395. doi: https://doi.org/10.1080/0924345970080401

Hamaker, E. L., & Wichers, M. (2017). No Time Like the Present: Discovering the Hidden Dynamics in Intensive Longitudinal Data. *Current Directions in Psychological Science*, *26* (1), 10-15. doi: https://doi.org/10.1177/0963721416666518

Henderson, C. R. (1982). Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics*, *31* (2), 423-447.

Kaplan, D., & Depaoli, S. (2012). Bayesian Structural Equation Modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650-673). New York, NY: Guilford.

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the Scientific Credibility of Single-Case Intervention Research: Randomization to the Rescue. *Psychological Methods*, *15* (2), 124-144. doi: https://doi.org/10.1037/a0017736

Lamiell, J. T. (1998). Nomothetic' and 'idiographic': contrasting Windelband's understanding with contemporary usage. *Theory & Psychology*, *8* (1), 23-38. doi: https://doi.org/10.1177/0959354398081002

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford.

Mainhard, T., Oudman, S., Hornstra, L., Bosker, R. J., & Goetz, T. (2018). Student emotions in class: The relative importance of teachers and their interpersonal relations with students.

Learning and Instruction, *53*, 109-119. doi: https://doi.org/10.1016/j.learninstruc.2017.07.011

Malmberg, L. E., Lim, W. H. T., Tolvanen, A., & Nurmi, J.-E. (2016). Within-students variability in learning experiences, and teachers' perceptions of students' task-focus. *Frontline Learning Research*, *4* (5), 62-82.

Marsh, H. W., Guo, J., Nagengast, B., Parker, P. D., Asparouhov, T., Muthén, B., & Dicke, T. (2017). What to do When Scalar Invariance Fails: The Extended Alignment Method for Multi-Group Factor Analysis Comparison of Latent Means Across Many Groups. *Psychological Methods*, Jan 12. doi: https://doi.org/10.1037/met0000113

Marsh, H. W., & Hau, K. T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology*, *32* (1), 151-170. doi: https://doi.org/10.1016/j.cedpsych.2006.10.008

Matuschek, H., Kliegl, R., Vasishtha, S., Baayen, H., & Bates, D. (2017). Balancing Type I Error and Power in Linear Mixed Models. *Journal of Memory and Language*, *94*, 305-315. doi: https://doi.org/10.1016/j.jml.2017.01.001

Mehta, P. (2013). N-level structural equation modeling. xxM user's guide version 1. In Y. M. Petscher, Ch. Schatschneider, & D. L. Compton (Eds.), *Applied Quantitative Analysis in Education and the Social Sciences* (p. 329). London: Routledge.

Moerbeek, M., & Teerenstra, S. (2016). *Power analysis of trials with multilevel data*. Boca Raton: Chapmans CRC Press.

Morin, A. J., Marsh, H. W., Nagengast, B., & Scalas, L. F. (2014). Doubly latent multilevel analyses of classroom climate: An illustration. *Journal of Experimental Education*, *82* (2), 143-167. doi: https://doi.org/10.1080/00220973.2013.769412

Murayama, K., Goetz, T., Malmberg, L. E., Pekrun, R., Tanaka, A., & Martin, A. J. (2017). Within-person analysis in educational psychology: Importance and illustrations. *British Journal of Educational Psychology*, Series II, 12.

Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22* (3), 376-398.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17* (3), 313-335. doi: https://doi.org/10.1037/a0026802

Natesan, P., & Hedges, L. V. (2017). Bayesian unknown change-point models to investigate immediacy in single case designs. *Psychological Methods*, *4* (22), 743-759. doi: https://doi.org/10.1037/met0000134

Ortega, L., Malmberg, L., & Sammons, P. (2018). School effects on Chilean children's achievement growth in language and mathematics: An accelerated growth curve model. *School Effectiveness and School Improvement*, *29* (2), 308-337. doi: https://doi.org/10.1080/09243453.2018.1443945

Ortega, L., Malmberg, L.-E., & Sammons, P. (2014). *Teacher effects on Chilean children's achievement growth: a cross-classified multiple membership accelerated growth curve model*. Paper presented at the conference Advances in Multilevel Modelling for Educational Research, at University of Maryland, USA.

Perels, F., Gurtler, T., & Schmitz, B. (2005). Training of self-regulatory and problem-solving competence. *Learning and Instruction*, *15* (2), 123-139. doi: https://doi.org/10.1016/j.learninstruc.2005.04.010

Praetorius, A. K., Koch, T., Scheunpflug, A., Zeinz, H., & Dresel, M. (2017). Identifying determinants of teachers' judgment (in)accuracy regarding students' school-related motivations using a Bayesian cross-classified multi-level model. *Learning and Instruction*, *52*, 148-160. doi: https://doi.org/10.1016/j.learninstruc.2017.06.003

Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2018). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction,* February 2017. doi: https://doi.org/10.1016/j.learninstruc.2018.01.004

Ram, N., & Grimm, K. (2007). Using simple and complex growth models to articulate developmental change: Matching theory to method. *International Journal of Behavioral Development*, *31* (4), 303-316. doi: https://doi.org/10.1177/0165025407077751

Rasbash, J., Steele, F., Goldstein, H., & Browne, W. (2017). *A User's Guide to MLwiN Version 3.01*. Bristol: Centre for Multilevel Modelling, Bristol.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. American Sociological Review, *15* (3), 351-357.

Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology*, *7*, 1-16. doi: https://doi.org/10.3389/fpsyg.2016.00110

Schmiedek, F. (2016). *Experimental manipulation «in the wild»: Proposing a within-person encouragement design*. Seminar presented in the series Network on Intrapersonal Research in Education (NIRE) at University of Oxford. Retrieved from http://www.education.ox.ac.uk/network-on-intrapersonal-research-in-education-nire/seminar-5-oxford-mayjune-2016/florian-schmiedek/ (Consulted on 12/06/2018).

Schmitz, B. (2006). Advantages of studying processes in educational research. *Learning and Instruction*, *16* (5), 433-449. doi: https://doi.org/10.1016/j.learninstruc.2006.09.004

Schmitz, B. (2015). *The study of learning processes using time-series analyses*. Seminar presented in the series Network on Intrapersonal Research in Education (NIRE) at University of Oxford. Retrieved from http://www.education.ox.ac.uk/network-on-intrapersonal-research-in-education-nire/seminar-1/bernhard-schmitz/ (Consulted on 12/06/2018).

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford: Oxford University Press.

Skinner, E. A., Zimmer-Gembeck, M. J., & Connell, J. P. (1998). Individual differences and the development of perceived control. *Monographs of the Society for Research in Child Development*, *63* (2/3). doi: https://doi.org/10.2307/1166220

Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: application to developmental research. *Child Development*, *85* (3), 842-860. doi: https://doi.org/10.1111/cdev.12169

Voelkle, M. C., Oud, J. H. L., von Oertzen, T., & Lindenberger, U. (2012, jul). Maximum Likelihood Dynamic Factor Modeling for Arbitrary N and T Using SEM. *Structural Equation Modeling*, *19* (3), 329-350. doi: https://doi.org/10.1080/10705511.2012.687656

Von Davier, M., González, E., & Mislevy, R. J. (2009). What are plausible values and why are they useful? In M. von Davier & D. Hastedt (Eds.), *IERI monograph series: Issues and methodologies in large scale assessments: Volume 2* (pp. 9-36). Hamburg, Germany: IERI Institute.

Walls, T. A., Barta, W. D., Stawski, R. S., Collyer, C. S., & Hofer, S. M. (2013). Time-scale dependent longitudinal designs. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of deve-*

lopmental research methods (pp. 45-64). New York: Guilford Press.

Walls, T. A., & Schafer, J. S. (2006). *Models for intensive longitudinal data*. New York: Oxford University Press.

Zitzmann, S., Lüdtke, O., Robitzsch, A., & Marsh, H. W. (2016). A Bayesian Approach for Estimating Multilevel Latent Contextual Models. *Structural Equation Modeling*, *23* (5), 661-679. doi: https://doi.org/10.1080/10705511.2016.1207179

Author's biography

Lars-Erik Malmberg is Associate Professor of Quantitative Methods in Education, at the Department of Education of University of Oxford. He has more than 70 publications (peer-reviewed articles, book chapters, and reports) and is the Editor-in-Chief of the *Journal of Learning and Instruction*. His current research interests focus on intraindividual perspectives on learning processes and modelling of intraindividual data, using multilevel structural equation models.

https://orcid.org/0000-0002-5309-7403

This paper was originally written in English. There is a Spanish translation of this paper in the printed version of issue 271, and on the website of the journal.

revista española de pedagogía
año LXXVI, nº 271, septiembre-diciembre 2018
Spanish Journal of Pedagogy
year LXXVI, n. 271, September-December 2018

Table of Contents

Sumario

Pedagogical research journals today
Las revistas de investigación pedagógica en la actualidad

Book reviews

This is the English translation of the studies and book reviews in the original Spanish printed version of issue 271 of **revista española de pedagogía**. The complete Spanish version of this issue can also be found on the journal's website (http://revistadepedagogia.org).

ISSN: 0034-9461 (Print), 2174-0909 (Online)
https://revistadepedagogia.org/
Depósito legal: M. 6.020 - 1958
INDUSTRIA GRÁFICA ANZOS, S.L. Fuenlabrada - Madrid