Comparing Single-Level and Multilevel Item Response Theory and Confirmatory Factor Analysis for Assessing the Social Composition of School Classes

Paper to be presented at the RC33 Eight International Conference on Social Science Methodology, Sydney 2012

Dominik Becker, Kerstin Drossel, Jasmin Schwanenberg, Heike Wendt, Nadja Pfuhl

Version of April 10, 2012

Authors' Affiliation:

Technical University of Dortmund
Institute of School Developmental Research
Martin-Schmeißer-Weg 13
44227 Dortmund
Germany

Contact: Becker@ifs.tu-dortmund.de

1 Introduction

Although already Raju et al. (2002) demonstrated major similarities of item response theory (IRT) and confirmatory factor analysis (CFA), Skrondal and Rabe-Hesketh (2007, 723) recently noted that still "factor analysts and item-response theorists rarely cite each other, although their work is closely related and often published in the same journal, Psychometrika". In the paper at hand, we strive to overcome this shortcoming by cross-validating results from a polytomous IRT approach, the partial credit model (PCM; Masters, 1982, 1988) by means of ordinal confirmatory factor analysis (CFA; Bollen, 1989). Hitherto, in a couple of German large-scale educational surveys, the PCM was used in order to classify school classes along a set of parental socio-economic status variables that were deduced from Bourdieu's (1986) cultural capital theory in order to assign crucial resources to comparably disadvantaged schools (Bonsen et al., 2007). However, these analyses typically assume the one-dimensionality of the underlying latent social composition variable on the student level. In CFA terminology, this is equivalent to a group-factor model whose group-factor correlations are all set to one (Rindskopf and Rose, 1988, 55f.) – which might impose an untenable restriction on the data. Hence, in a first step, we fit a conventional PCM and test for the one-dimensionality of our latent social composition variable. In a second step, we only use those social composition items that achieved an acceptable fit in the PCM model to build up an ordinal CFA within subsequently both reflective and formative specifications (Bollen and Lennox, 1991; MacCallum and Browne, 1993; Diamantopoulos and Winklhofer, 2001). Tentative results based on a sample of highest-track secondary school students (N=3310) and their parents (N=2729) in the German federal state of North-Rhine Westphalia suggest that while in the PCM, a satisfactory-fitting one-dimensional social composition index could be obtained, this does not equally hold for the *reflective* CFA wherein a second-order (Rindskopf and Rose, 1988; Chen et al., 2005) three-factorial model fitted the data significantly better than imposing the restriction of one-dimensionality as well as the one of zero variance and unity of factor loadings (as would be the specification of a IRT model in the CFA context).

Also in a *formative* specification – viewing social background indicators as *causes* of the latent variable (which is reversely within a reflective measurement model) –, the model's χ^2 and corresponding p-values suggest that allowing for three-dimensionality can improve model fit.

Results from multilevel CFA and IRT admitting to fit a latent variable directly on the school-class level indicate that notwithstanding a few problems regarding both model convergence and negative values in the residual covariance matrix, the student-level measurement structure basically seems to hold also on class-level in that a three-factorial second-order measurement model suits the data better than both a simple one-factorial solution and an even more restricted model that was set up according to the specifications of IRT.

We conclude with an outlook on further cross-validation of the results obtained by both PCM and CFA by means of latent class analysis estimated on both student and school-class level (Vermunt, 2003), and with a critical discussion of whether rigorous factor structure tests such as PCM and CFA are suitable for assessing school-classes' social composition within the more 'lax' framework of cultural capital theory.

2 Theoretical and Methodological Background

2.1 Theory

Beginning with the seminal Coleman report (Coleman, 1966), and despite the intense debate the report fostered (Bowles and Levin, 1968; Coleman, 1968; Cain and Watts, 1968; Coleman, 1970; Cain and Watts, 1970), the importance of a school's social composition for student outcomes is well known. On the one hand, parental social backgrounds of course have a non-negligable impact on students' educational achievement. After reception of the influencial monographies by Coleman (1966) and Jencks (1972), and also of the *Wisconsin status attainment model* (Sewell et al., 1969, 1970), Boudon (1974) developed a frame accounting for *inequalities in educational opportunities* still influential today.

The *primary* effect of educational inequality states that the lower educational success of lower-SES children may be due to their lower capabilities – be they defined as educational interests, intellectual skills, effort or motivation (Müller-Benedict, 2007). While part of the primary effect may also be genetic, its presumably greater part is acquired during socialization (Erikson and Jonsson, 1996a, p. 10f.). The *secondary* effect, contrarily, operates via stratum-specific differences in educational decision making due to differential opportunity cost structures, and Boudon's crucial assumption is that secondary effects still take place once primary effects have been controlled for (Nash, 2005). The idea that utility considerations may shape students' (or their parents') educational decisions was further elaborated by Erikson and Jonsson (1996b); Goldthorpe (1996); Breen and Goldthorpe (1997) and Esser (1999).

Another influential theoretical account of educational inequalities in part counterbalancing the one of Boudon (1974) is Bourdieu's capital theory (Bourdieu, 1973, 1986; Bourdieu and Passeron, 1990) distinguishing between cultural capital, economic capital, and social capital. *Cultural capital* includes "all the goods, material and symbolic, without distinction, that present themselves as rare and worthy of being sought after in a particular social formation" Bourdieu (1977, p. 178). It may be either *institutionalized* in terms of educational qualifications, *objectified* in terms of physical objects such as work of arts or books, or *embodied* in terms of inherited propositions acquired over time and being reflected e.g. in individuals cultural practices. *Economic capital* is assessed by an individual's dispose of economic resources such as cash and assets, and *social capital* by an individual's social networks and group memberships.

Primarily the influence of family's cultural capital on students' educational attainment has been subject to a multiplicity of empirical studies (see Lareau and Weininger, 2003 for an overview and Jaeger, 2009 and Andersen and Hansen, 2011 for more recent applications) – but also the effect of was prominently emphasized also in other theoretical accounts such as Coleman's resource theory (Coleman, 1988). A hypothesized effect of economic capital on educational achievement would also be in line with the idea of *relative risk aversion* by Breen and Goldthorpe (1997) postulating that children from upper-class families need a relatively higher educational qualification to achieve at least the level of their parents than children from mid-class or below.

But on the other hand, many studies also suggest that school's social composition *on the aggregate level* is an important predictor of students' individual educational outcomes. Particularly studies of *school effectiveness* (Sammons, 1999; Scheerens, 2000; Rivkin et al., 2005) lay emphasize on the effect of school-level socioeconomic status on educational achievement. Moreover, Ditton (2010) extends the Scheerens (2000) model also on students' self-concept, and indeed found early sociological studies positive effects of school status on students' educational aspirations (Meyer, 1970; Alexander and Eckland, 1975; Alwin and Otto, 1977) – which was also corroborated by the study of Marsh et al. (2000).

In sum, there is sufficient evidence for the relevance of parental social backgrounds measured both on the individual level and on contextual level for students' educational outcomes. In a couple of German large-scale educational studies such as *KESS4* and *KESS7* (Bos and Pietsch, 2006; Bos et al., 2010), various information about all parental forms of capital was used to measure a latent variable of students' socioeconomic status in the framework of *I*tem *R*esponse *T*heory. In a second step, the resulting metric variable was aggregated on the school-class level and then categorized in order to arrive at different groups of school classes according to their SES. However, *IRT* is not the only approach to estimate latent variables – which is why in the next section, we will review a couple of available methods.

2.2 Models

2.2.1 Confirmatory Factor Analysis

The idea of *C*onfirmatory *F*actor *A*nalysis (CFA) is to empirically translate theoretical concepts into latent variables that are ideally mapped by a series of indicators in order to reduce measurement error. While conventional estimation strategies such as *O*rdinary *L*east *S*quares (OLS) regression typically measure each concept (e.g. education) with one indicator (e.g. certificate of highest qualification), CFA and its 'regression' counterpart, *S*tructural *E*quation *M*odelling (SEM) makes use of multiple indicators and explicitly models measurement error. The notation of CFA is as follows:

$$\mathbf{x} = \Lambda_x \xi + \delta \tag{1}$$

with $\mathbf{x}$ as a vector of manifest variables, $\mathbf{\Lambda_x}$ as the matrix of factor loadings, ξ as a vector of latent variables, and δ as a vector of *unique* or *specific* factors that reflects measurement error (Bollen, 1989, p. 233). Suppose that two latend variables ξ_1 and ξ_2 should be mapped by three items each. This factor structure can then be specified as follows (cf. Bollen, 1989, p. 234):

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{pmatrix} \cdot \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{pmatrix} \tag{2}$$

For ease of notation, we now reduce the above example to the single-factor case of ξ_1 mapped by three items x_{11}, x_{12}, x_{13}. Suppose that we observe an *empirical* covariance matrix $\mathbf{S}$ of our three items such as

$$\mathbf{S} = \begin{pmatrix} \sigma^2_{x_1} & & \\ \sigma_{x_2,x_1} & \sigma^2_{x_2} & \\ \sigma_{x_3,x_1} & \sigma_{x_3,x_2} & \sigma^2_{x_3} \end{pmatrix} \tag{3}$$

Based on the hypothesized factorial structure of the data, the *implied* covariance matrix $\mathbf{\Sigma}$ is estimated by means of the factor loadings $\mathbf{\Lambda_x}$ and the variances of δ:

$$\mathbf{\Sigma} = \begin{pmatrix} \hat{\sigma}^2_{x_1} & & \\ \hat{\sigma}_{x_2,x_1} & \hat{\sigma}^2_{x_2} & \\ \hat{\sigma}_{x_3,x_1} & \hat{\sigma}_{x_3,x_2} & \hat{\sigma}^2_{x_3} \end{pmatrix} = \begin{pmatrix} \lambda^2_{11} + \sigma^2_{\delta_1} & & \\ \lambda_{11}\lambda_{21} & \lambda^2_{21} + \sigma^2_{\delta_2} & \\ \lambda_{11}\lambda_{31} & \lambda_{21}\lambda_{31} & \lambda^2_{31} + \sigma^2_{\delta_3} \end{pmatrix} \tag{4}$$

The objective of the estimation process is to minimize the difference between the observed covariance matrix $\mathbf{S}$ and the implied covariance matrix $\mathbf{\Sigma}$ – which varies with the measurement level of the manifest indicators.

Continuous manifest variables In case of continuous manifest indicators, estimation is straightforward via maximum likelihood (ML). The fitting function to be minimized reads (cf. Jöreskog, 1969, p. 184).

$$F_{ML} = ln|\mathbf{S}| - ln|\mathbf{\Sigma}| + trace[(\mathbf{S})(\mathbf{\Sigma^{-1}})] - p, \tag{5}$$

where *trace* refers to the sum of the elements in the main diagonal, and p to the number of indicators.

Categorical manifest variables In case of categorical manifest indicators, conventional maximum likelihood estimation based on a usual variance-covariance matrix will be biased (Bollen, 1989, p. 433ff). Instead, it has been suggested to use a matrix of polychoric correlations as input covariance matrix and then either a *W*eighted *L*east *S*quares (WLS) estimator or a ML estimator with bootstrapped standard errors.

The basic idea of polychoric correlations of categorical variables is to compute the thresholds of an assumed underlying continuous variable (Olsson, 1979; Muthén, 1984; Aish and Jöreskog, 1990; Jöreskog, 1994) as input matrix. Concretely, each ordinal variable X is assumed to be a manifestation of an underlying continuous variable x^* which is normally distributed with mean μ_x and variance σ^2_x (Jöreskog, 1990):

$$x = i \Leftrightarrow \alpha_{i-1} > x^* \leq \alpha_i, i = 1, 2, ..., k, \tag{6}$$

where

$$\alpha_0 = -\infty, \alpha_1 < \alpha_2 < ... < \alpha_{k-1}, \alpha_k = +\infty \tag{7}$$

In case of WLS, the fitting function reads (Bollen, 1989, p. 443)

$$F_{WLS} = [\hat{\rho} - \sigma(\theta)]\prime\mathbf{W}-1[\hat{\rho} - \sigma(\theta] \tag{8}$$

where $\hat{\rho}$ is a vector of polychoric correlations, $\sigma(\theta)$ is the corresponding vector for the implied covariance matrix, and W is a consistent estimator of the asymptotic covariance matrix of $\hat{\rho}$.[1]

Maximum-Likelihood estimation of SEM models based on polychoric correlations as an observed matrix **S** in sense of equation (5) may lead to consistent estimates, but the standard errors, z-values and significance parameters will be biased (Bollen, 1989, p. 443) – which may be corrected by use of bootstrapping techniques (Zhang and Browne, 2006; Fox, 2006).[2]

Reflective vs. formative specification A crucial differentiation to be considered in the context of CFA is the distinction between reflective and formative indicators. While the above-specified model of a latent variable affecting the distribution of its indicators is a case of a *reflective* measurement model, a *formative* measurement model addresses the point that “in many cases, indicators could be viewed as causing rather than being caused by the latent variable measured by the indicators” (MacCallum and Browne, 1993, p. 553).

The formal notation of a formative measurement model reads

$$\nu = \gamma_1 X_1 + \gamma_2 X_2 + ... + \gamma_q X_q + \zeta \tag{9}$$

A well known example of case where a formative measurement model would have to be applied is the socio-economic status that is a composite of different items such as education, income, occupational prestige, etc. (Diamantopoulos and Winklhofer, 2001, p. 269f.). The difference between reflective and formative measurement is also illustrated graphically in figure 1.

Multilevel CFA Multilevel analysis or hierarchical linear modeling in general is a statistical tool to account for clustered data. Consider a conventional regression equation of the form

$$y_i = \beta_0 + \beta\mathbf{X_i} + \epsilon_i \tag{10}$$

[1] An estimator of W is provided by Muthén (1984) – which is beyond the scope of this overview.

[2] We do not consider this case here. See Babakus et al. (1987) and Rigdon and Ferguson Jr (1991) for issues of convergence rates and fit statistics of polychoric correlations depending on different types of categorization.

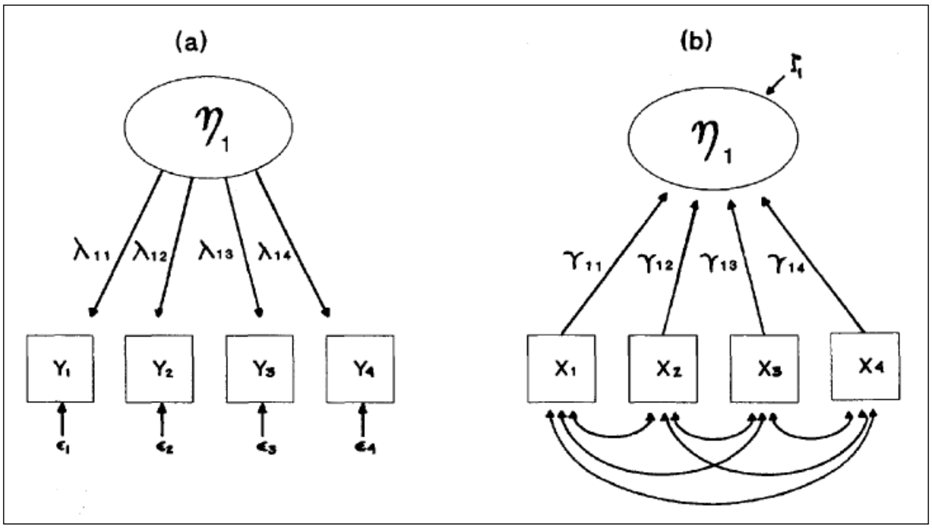


Figure 1: Reflective (a) and formative (b) measurement of latent variables (Bollen and Lennox, 1991, p. 306).

where y is the outcome of interest, $beta_{0i}$ the regression intercept, X a vector of predictors with slope β, and ϵ_i the error term for individual i, respectively. In this case, *O*rdinary *L*east *S*quares (OLS) regression provides the *B*est *L*inear *U*biased estimate of **X** and its corresponding standard errors (McElroy, 1967).

Consider now individuals i to be nested in cluster unit j (e.g. students nested in school classes). In this case, the outcome might be affected by predictors from both units of analyses (e.g. students' self-concept by both student-level and class-average achievement; see Marsh and Parker, 1984). Hence, we would start from

$$y_{ij} = \beta_0 + \beta_{\mathbf{i}}\mathbf{X_{ij}} + \beta_j\mathbf{Z_j} + \epsilon_{ij} \tag{11}$$

– where $\mathbf{X_{ij}}$ is a vector of lower-level predictors (such as student achievement), and $\mathbf{Z_j}$ is a vector of contextual-level predictors (such as class-average achievement; Snijders and Bosker, 1999).

While this would be denoted as a *fixed-effect* parametrization, one could explicitly allow for *random effects*. For instance, even controlling for both $\mathbf{X_{ij}}$ and $\mathbf{Z_j}$, some schools might show a higher average level of school-average student self-concept than other schools. This would be reflected by

$$\beta_{oj} = \gamma_{00} + \mu_{0j}, \tag{12}$$

where γ_{00} is the 'real' average intercept and the error term μ_{oj} a group-specific deviation from it. Since seperate error terms would have to be specified for each cluster unit, OLS is no longer BLUE. Hence, multilevel analysis partitions the variance of the outcome into a within-cluster-unit and a within-cluster-unit part in order to obtain consistent parameter estimates. Since contextual-level variables necessarily have less cases than lower-unit variables, not distinguishing these units would inflate the degress-of-freedom of the higher-level unit variables, decrease, the standard errors of their parameter estimates, thereby increase corresponding t- or z-values and thus lead to an overhastly acceptance of contextual-level hypotheses. This issue is also addressd by multilevel analysis in terms of calculating separate degrees-of-freedom of each unit of analysis.

Multilevel CFA grabs the idea of different units of analysis and allows to fit latent variables separately for each level. Thus, a factor η_{gi} is specified as

$$\eta_{gi} = \alpha + \eta_{Bg} + \eta_{Wgi}, \tag{13}$$

where α is the overall expectation for η_{gi}, η_{Bg} a random factor component capturing the between-group (e.g. school) effects, and η_{Wgi} a random factor component capturing the within-group effects (e.g. students within their schools; Muthen, 1994, p. 379). Hence, the total variance of η_{gi} may be decomposed into

$$\mathbf{V}(\eta_{gi}) = \Psi_T = \Psi_B + \Psi_W \tag{14}$$

Hence, the general structure of a two-level CFA becomes

$$y_{gi} = \nu + \Lambda_b \eta_{Bg} + \epsilon_{BG} + \Lambda_W \eta_{Wgi} + \epsilon_{Wgi} \tag{15}$$

with

$$V(y_{gi}) = \Sigma_B + \Sigma_W, \tag{16}$$

$$\Sigma_B = \Lambda_B \Psi_B \Lambda_B\prime + \Theta_B, \tag{17}$$

and

$$\Sigma_W = \Lambda_W \Psi_W \Lambda_W\prime + \Theta_W. \tag{18}$$

An exemplary two-level CFA is illustrated graphically in figure 2.

2.2.2 Item Response Theory

Unlike in classic test theory, and as in CFA, Item Response Theory (IRT) also accounts for measurement error. But unlike CFA, IRT more precisely distinguishes between person ability β_n and item difficulty δ_i.

Dichotomous items The simple *Rasch* model (Rasch, 1960) can be applied to items that have either been 'solved' or not. In this model, the probability of a positive response of person n on item i, π_{1ni}, is modeled via a logistic link function (cf. Masters, 1982, p. 152):

$$\pi_{1ni} = \frac{exp(\beta_n - \delta_i)}{1 + exp(\beta_n - \delta_i)} \tag{19}$$

The dichotomous Rasch model can be visualized via Item Characteristic Curves (ICC) as illustrated in figure 3. Each of the curve follows the logistic distribution. The more left a curve is located, the more easy the item and thus the higher π_{lni} given constant person ability.

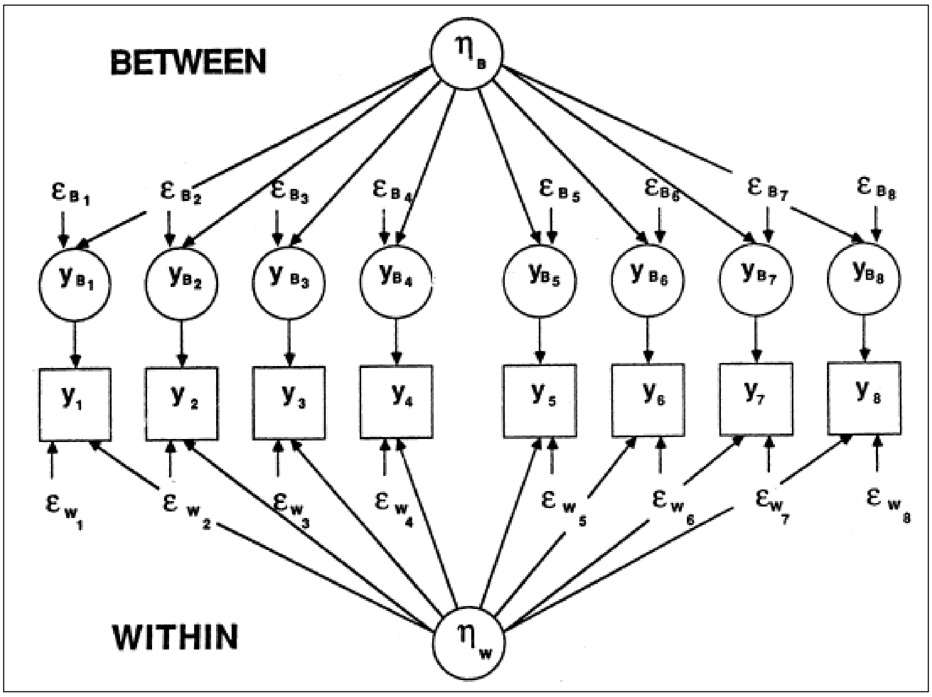


Figure 2: Multilevel Covariance Structure Path Diagram (Muthen, 1994, p. 386).

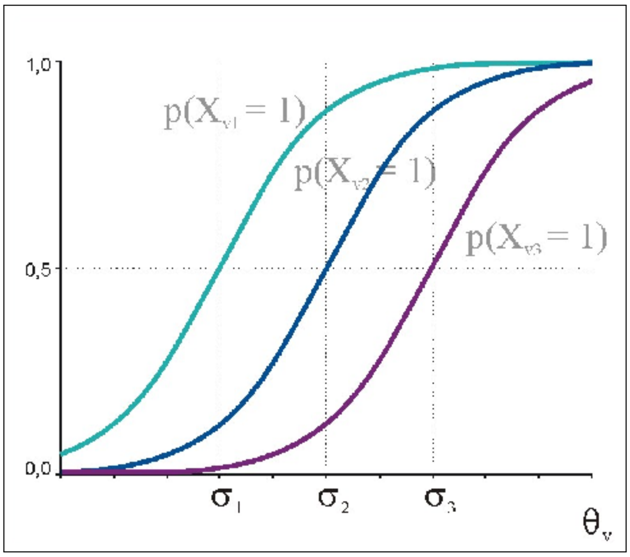


Figure 3: Exemplary Item Characteristic Curves.

IRT model parameter estimates are obtained via ML:

$$L = \frac{exp(\sum_{n=1}^{N} r_n\beta_n - \sum_{i=1}^{k} n_i\delta_i)}{\prod_{n=i}^{N}\prod_{i=1}^{k}(1+exp(\beta_n-\delta_i))} \quad (20)$$

where r_n is the number of correctly-answered items by person n, n_i is the number of correct item-answers of item i, β_n is the unknown person ability and δ_i the unknown item difficulty.

Polytomous items Masters (1982, 1988) proposed the *Partial Credit Model* in order to extend the Rasch model onto application with more than two ordered categorical indicators. The probability of person n with ability β_n to respond in category x ($x = 0$, 1, ..., m) of item i is as follows (Masters, 1988, p. 284):

$$\pi_{nix} = \frac{exp \sum_{j=0}^{k} n(\beta_n - \delta_{ij})}{\sum_{k=0}^{m} exp \sum_{j=0}^{k} n(\beta_n - \delta_{ij})} \text{ for } x = 0, m \tag{21}$$

Multilevel IRT Following Raudenbush et al. (2003), the Rasch model can be understood as a two-level logistic regressions with items nested within persons. But while in the classical Rasch model, person abilities and item difficulties would be fixed effects, this constrait can be relaxed in terms of letting these parameters vary by an additional cluster unit. In log-odds specification, the model proposed by Raudenbush et al. (2003) reads the following:

$$\eta_{ijk} = \sum_{p=1}^{P} D_{pijk}(\pi_{pjk} + \sum_{m=1}^{M_p - 1} \alpha_{pmjk}\alpha_{pmijk}), \tag{22}$$

where η_{ijk} is the log-odds that person j in cluster unit k (e.g. school classes) will positively answer item i, D_{pijk} accounts for items measured on different scales for dimension p, π_{jk} is the positive answer of person j on item i in cluster unit k within dimension p, $\alpha_{pmjk} = 1$ if item i is the m^{th} item within scale p (zero otherwise), and α_{pmijk} is the discrepancy between the log-odds of a positive response to the m^{th} item in scale p for person j in cluster unit k and the reference item within that scale (Raudenbush et al., 2003, p. 182).[3]

2.3 Research Questions

Having summarized a couple of latent variable models, the research questions to be answered in this paper are the following:

1. Do IRT and CFA arrive at approximately similar results in mapping the latent variable?
2. Once a student-level factor structure has been obtained, is this structure transferable also on the school class-level?

[3]While the authors also allow for multidimensionality, we here restrict analyses to the more simple case of only one dimension.

In order to answer the first research question, we first fit a partial credit model in `ConQuest` (Wu et al., 2007), and then use the remaining indicators to build a categorical CFA model in `MPlus`. We then test the CFA model against the restriction of unity of the factor loadings and unique variance of the latent variable – which can be considered to be an IRT specification in an CFA framework.

In order to answer the second research question, we fit a multilevel CFA based on a covariance structure both on the student and on the school-class level, and we also test this model against the corresponding IRT restriction.

3 Data

All indicators come from the student and parent survey of a comprehensive German longitudinal study started in autumn 2010. The data was surveyed in context of a German region-wide project called "Ganz In - All-Day Schools for a Brighter Future. The New All-Day Secondary School in North Rhine-Westphalia" (Berkemeyer et al., 2010). In order to reduce inequalities in educational opportunities, 31 upper-secondary schools joined the project and switched from half-day to all-day schooling. Amongst various means of school developmental advice, the evaluation process consists of regular quantitative and qualitative assessments. The quantitative data is collected in a longitudinal design comprising surveys of students, their parents, teachers, school's pedagogic staff apart from teachers, and school principals. In order to answer our research questions, we used the data from the parent questionnaire of the initial survey from 2010 which reached 2.742 parents (equal to a response rate of about 83%) of 5^{th} graders in 31 schools in a multitude of neighborhoods varying in social context. A particular limitation of this parent survey is the remarkable share of mothers in the sample (82.2%). Also, more than 60% of the parents are equipped with a gross income higher than 40.000 € a year (for an amount of about 27% also higher than 70.000 € a year), and almost 70% of the respondents dispose of an educational degree (Abitur) that qualifies for academic studies.

Indicators The parent survey comprises of various indicators of parental social backgrounds. We base our analysis on a set of variables that has already been used successfully in other studies that estimated a social composition index on the school class level (e.g. Bonsen et al., 2010).

Six dichotomous indicators refer to parental migration status, of which three are related to the country of birth (*ELTERN3a, ELTERN3b, SCHÜLER4*), and another three to mother tongue and colloquial language at home (*ELTERN4a1, ELTERN4a2, ELTERN4a3*). On each of these indicators, a value of one indicates to be born in Germany or to speak German language, respectively (and zero otherwise). Parental objectivized cultural capital is measured by the fact whether at least one of the child's parent disposes of a high school degree qualifying for academic studies (German *Abitur*) and by the number of books at home (1 'less than 100'; 2 'between 100 and 200'; 3 'more than 200'). Twelve items (*ELTERN5a* – *ELTERN5l*) measured by a four-point scale assess parents' incorporated cultural capital (1 'does not apply at all'; 2 'does not apply'; 3 'applies'; 4 'applies strongly'). Finally, economic capital is controlled via the household's yearly gross average income. Due to shortcomings in data return from schools, families' social capital could not be considered yet. Therefore, the results presented below are based on preliminary measurement models.

Table 1 summarises these indicators, their level of measurement, and corresponding means and standard deviations.

Table 1: Distribution of indicators

	count	mean	sd	min	max
ELTERN3a - born in Germany?	2547	0.72	0.45	0.00	1.00
ELTERN3b - father/male guardian born in Germany?	2059	0.70	0.46	0.00	1.00
ELTERN4a1 - mother tongue German?	2643	0.76	0.43	0.00	1.00
ELTERN4a2 - colloqial language with child German?	2646	0.85	0.35	0.00	1.00
ELTERN4a3 - partner: colloqial language with child German?	2648	0.65	0.48	0.00	1.00
ELTERN5a2 - visiting museums, exhibitions	2535	1.93	0.44	1.00	4.00
ELTERN5b2 - visiting blockparty, rummage, amusement park	2639	2.39	0.55	1.00	4.00
ELTERN5c2 - visiting philharmonic concerts, opera, theatre	2531	1.70	0.53	1.00	4.00
ELTERN5d2 - visiting cinema, pop concert, discotheque	2610	2.19	0.55	1.00	4.00
ELTERN5e2 - visiting sport events	2565	2.20	0.91	1.00	4.00
ELTERN5f2 - actively practicing sports	2567	3.12	1.04	1.00	4.00
ELTERN5g2 - practicing music or art	2553	2.61	1.10	1.00	4.00
ELTERN5h2 - meeting friends and relatives	2632	3.47	0.69	1.00	4.00
ELTERN5i2 - volunteering in associations	2486	1.67	0.95	1.00	4.00
ELTERN5j2 - engagement in citizens' action committee	2478	1.09	0.34	1.00	4.00
ELTERN5k2 - visiting religious events	2587	2.29	0.96	1.00	4.00
ELTERN5l2 - listening to classical music or jazz	2537	1.93	1.05	1.00	4.00
ELTERN7 - no. of books in household	2681	2.07	0.89	1.00	3.00
ABI - at least one parent with Abitur	2651	0.69	0.46	0.00	1.00
ELTERN12 - yearly gross average income of household	2329	2.48	0.72	1.00	3.00
SCHUELER4 - student born in Germany?	3200	0.96	0.19	0.00	1.00

4 Results

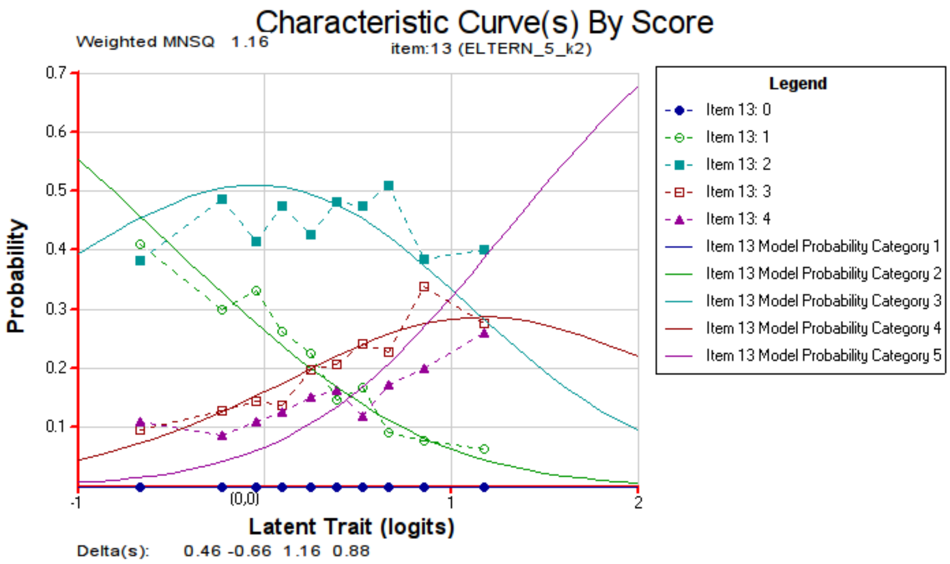
4.1 Single-level analyses

4.1.1 Item response theory

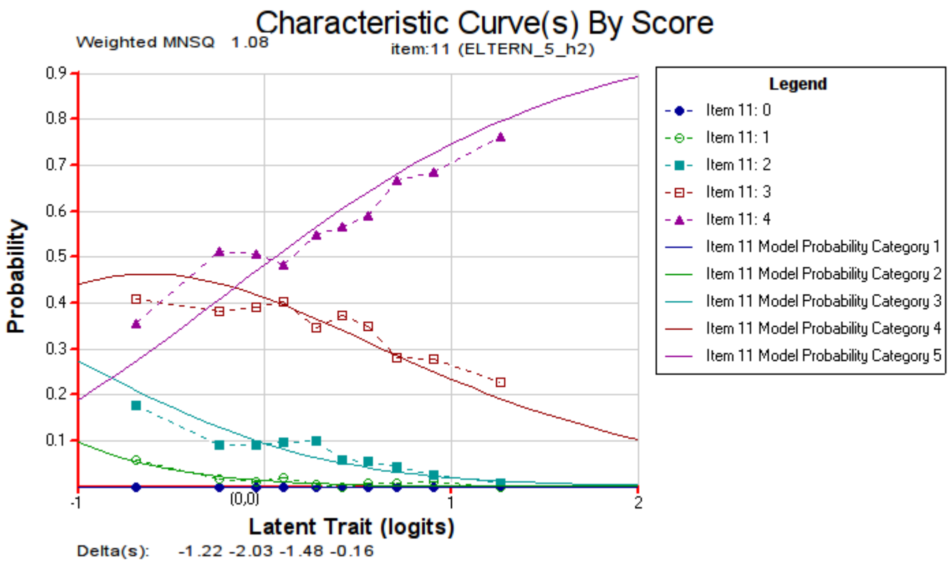
We first fitted a conventional partial credit model in ConQuest (Wu et al., 2007). In the first run, all items went into the analysis. Based on an initial run, IRT provides a couple of item fit statistics to decise whether an item appropriatly maps the latent dimension to meausre. These involve the weighted *M*ea*N SQ*uare error, its t-value, the *I*tem *C*haracteristic *C*urves, and the item discrimination parameter of classical test theory. The MNSQ is defined as the sum of the squared residuals divided by sample size and weighted by each residual's variance and has an expected value of 1. It should be neither smaller than .75 nor larger than 1.3 (Bond and Fox, 2001, cf. eg.). As known from other applications, corresponding t-values should not exceed the value of 1.96. The ICC plots help to judge how closely an *empirical* latent-trait-probability curve follows its *theoretical* expectation. The more an empirical ICC resembles the logistic distribution, the better the fit of the corresponding item. Finally, the threshold of the item discrimination statistic below an item would be interpreted as showing a bad fit is .2.

In the initial run of the partial credit model, all items showed MNSQ statistics within the acceptable range. However, a couple of items showed discrimination statistics lower than .20. Hence, up to run four, items with the lowest discriminiation were discarded subsequently. This affected *ELTERN5b* (visiting blockparty, rummage, amusement park) , *Eltern4a* (colloqial language with child German?), and *ELTERN5j* (engagement in citizens' action committee). After run four, all items showed satisfactory fit statistics, but the MNSQ of *ELTERN5k* (visiting religious events) had both a notable high t-value and an ICC curve conspiciously deviating from the expected curve. Hence, this item was dropped for another run 5 where *ELTERN5e* (visiting sport events) showed a high t-value but an ICC curve along expectations – which is why this and all other items were maintained. Figure 4 shows the ICC curves of the bad-fitting item *ELTERN5k* from run 4, and, as a contrast, of the well-fitting item *ELTERN5h* (meeting friends and relatives) from run 5.[4]

[4]Note: In the current specification, the empty zero category is estimated by ConQuest by default.



(a) attending religious events



(b) meeting friends and relatives

Figure 4: Exemplary item characteristic curves of a bad-fitting and a well-fitting item.

Having eleminated the misfitting items, run 5 is considered to provide a satisfactory partial credit model. Figure 5 shows the relative discrimination of the remaining items on the latent variable. The 'higher' an item is located on the vertical axis, the more 'difficult' it is for a respondent to answer positively (in case of dichotomous items) or to exceed the next threshold (in case of ordinal items).

We note that items *SCHÜLER4* (student born in Germany) and *ELTERN5h* (meeting relatives and friends) are the most 'easy' items. Even parental economic capital and institutionalized cultural capital in tems of Abitur are relatively ineffective in discriminating – which is of course due to the particular selectivity of our sample. A bit 'harder' is the second indicator of institutionalized cultural capital, the number of books at home. In contrast, the most discriminating items are the most 'highbrow' indicators of parental cultural practice, *ELTERN55a* (visiting museums and art exhibitions) and *ELTERN5c* (visiting philhharmonic concerts, opera, theatre). Hence, once a certain level of economic wealth and education is reached, 'highbrow' cultural practices still discriminate among parents.

4.1.2 Confirmatory factor analysis

Reflective specification Based on this set of indicators, we next performed a series of CFA in MPlus based on the WLSMV estimator. When trying to obtain a one-factorial model based on all indicators remaining from the IRT model, the model fit was very bad since many indicators either showed either insignificant or relatively low standardized factor loadings (see table 2, *model 1a*). Thus, in the next nine runs, these items were subsequently dropped from the measurement model – which could improved model fit (*model 1j*).

Table 2: One-factor CFA solutions

	Model 1a		*Model 1j*	
	λ	p	λ	p
SES	BY			
ELTERN3A	0.948	$< .001$	0.955	$< .001$
ELTERN3B	0.877	$< .001$	0.893	$< .001$
ELTERN4A1	0.928	$< .001$	0.948	$< .001$
ELTERN4A3	0.351	$< .001$	0.402	$< .001$
ELTERN5A2	0.397	$< .001$		
ELTERN5C2	0.413	$< .001$		
ELTERN5D2	0.04	0.077		
ELTERN5E2	0.112	$< .001$		
ELTERN5F2	0.224	$< .001$		
ELTERN5G2	0.331	$< .001$		
ELTERN5H2	0.091	$< .001$		
ELTERN5I2	0.083	0.001		
ELTERN5L2	0.229	$< .001$		
ELTERN7	0.64	$< .001$	0.611	$< .001$
ABI	0.554	$< .001$	0.549	$< .001$
ELTERN12	0.614	$< .001$	0.651	$< .001$
SCHUELER4	0.444	$< .001$	0.497	$< .001$
CFI	0.766		0.964	
TLI	0.787		0.966	
RMSEA	0.117		0.092	
WRMR	4.958		3.07	
χ^2	3458.981		460.69	
df	76		16	
p	$< .001$		$< .001$	

Note: All factor loadings are standardized.

However, according to theory, one could expect a measurement model with at least three distinct dimensions: Recall that we comprise of indicators assessing *migration background*, *institutionalized cultural capital*, *incorporated cultural capital*, and *economic capital* (while the latter is only measured by one single manifest variable). Hence, in a second step, we estimated a three-factorial measurement model with separate latent variables for migration background, institutionalized cultural capital, and economic capital (table 3). Again, the fit of the initial model with all indicators was not perfect due to a couple of indicators with factor loadings smaller than .4 (*model 2a*). Having subsequently dropped these items, model fit approaches a satisfactory level also better than that of the one-factorial structure (*model 2g*).

Table 3: Three-factor CFA solutions

	Model 2a		*Model 2g*	
	λ	p	λ	p
OBJKK	BY			
ELTERN7	0.883	0	0.883	0
ABI	0.657	0	0.657	0
MIG	BY			
ELTERN3A	0.968	0	0.971	0
ELTERN3B	0.897	0	0.892	0
ELTERN4A1	0.951	0	0.951	0
ELTERN4A3	0.376	0		
SCHUELER4	0.523	0	0.527	0
INKKK	BY			
ELTERN5A2	0.656	$< .001$	0.655	$< .001$
ELTERN5C2	0.708	$< .001$	0.757	$< .001$
ELTERN5D2	0.205	$< .001$		
ELTERN5E2	0.284	$< .001$		
ELTERN5F2	0.449	$< .001$		
ELTERN5G2	0.62	$< .001$	0.591	$< .001$
ELTERN5H2	0.265	$< .001$		
ELTERN5I2	0.299	$< .001$		
ELTERN5L2	0.558	$< .001$	0.598	$< .001$
MIG	WITH			
OBJKK	0.59	$< .001$	0.586	$< .001$
INKKK	WITH			
OBJKK	0.395	$< .001$	0.463	$< .001$
MIG	0.116	$< .001$	0.135	$< .001$
CFI	0.949		0.988	
TLI	0.948		0.987	
RMSEA	0.059		0.047	
WRMR	2.546		1.683	
χ^2	784.989		177.697	
df	64		22	
p	$< .001$		$< .001$	

Note: All factor loadings are standardized.

One might detect an interesting pattern in the remaining indicators of incorporated cultural capital and in those that were discarded: The remaining items without exception measure the 'highbrow' dimension of cultural capital (visiting museums, attending philharmonic concerts, etc.) – which does not hold for the indicators that were discarded. Therefore, we also tested for a four-factorial structure that separately modeling the 'highbrow' dimension of incorporated cultural capital (table 4).

The initial four-factor model already achieved a satisfactory model fit (see *model 3a*), but the latter could improved once more by dropping items *ELTERN5D* (attending cinema, pop concert, discotheque) and *ELTERN5H* (meeting friends and relatives). Having done so, the covariances between **SPANN** one the one hand and both **OBJKK** and **MIG** on the other hand that were insignificant in *model3a* turned out to be significant in *model3c*.

While tables 3 and 4 specified a correlated factorial structure, an alternative specification (and also a test of the relative importance of each latent variable) is a second-order measurement model (Rindskopf and Rose, 1988; Chen et al., 2005) wherein each latent variable is in turn an indicator of a higher-level latent variable. Table 5 displays the factor loadings of each first-order latent variable on the second-order latent variable **SES** separately for the three-factor- and the four-factor specification as estimated in tables 3 and 4.

Results show that similar to the first-order CFAs, the four-factorial second-order measurement model fits the data a bit worse than the three-factorial second-order measurement model. Moreover, the fourth latent variable **SPANN** shwns only a relatively weak factor loading on the second-order latent variable **SES**. Hence, for subsequent analyses, we prefer the three-factorial solution.

Table 4: Four-factor CFA solutions

	Model 3a		*Model 3c*	
	λ	p	λ	p
OBJKK	BY			
ELTERN7	0.883	$< .001$	0.883	$< .001$
ABI	0.657	$< .001$	0.657	$< .001$
MIG	BY			
ELTERN3A	0.971	$< .001$	0.971	$< .001$
ELTERN3B	0.893	$< .001$	0.893	$< .001$
ELTERN4A1	0.951	$< .001$	0.951	$< .001$
SCHUELER4	0.526	$< .001$	0.526	$< .001$
INKKK	BY			
ELTERN5A2	0.669	$< .001$	0.668	$< .001$
ELTERN5C2	0.733	$< .001$	0.737	$< .001$
ELTERN5G2	0.626	$< .001$	0.619	$< .001$
ELTERN5L2	0.573	$< .001$	0.577	$< .001$
SPANN	BY			
ELTERN5D2	0.264	$< .001$		
ELTERN5E2	0.497	$< .001$	0.515	$< .001$
ELTERN5F2	0.752	$< .001$	0.788	$< .001$
ELTERN5I2	0.424	$< .001$	0.447	$< .001$
ELTERN5H2	0.356	$< .001$		
MIG	WITH		WITH	
OBJKK	0.586	$< .001$	0.586	$< .001$
INKKK	WITH		WITH	
OBJKK	0.465	$< .001$	0.464	$< .001$
MIG	0.138	$< .001$	0.137	$< .001$
SPANN	WITH		WITH	
OBJKK	0.056	0.087	0.067	0.041
MIG	0.043	0.193	0.065	0.049
INKKK	0.457	0	0.4	0
CFI	0.975		0.981	
TLI	0.975		0.98	
RMSEA	0.044		0.045	
WRMR	1.845		1.748	
χ^2	397.573		298.458	
df	55		40	
p	$< .001$		$< .001$	

Note: All factor loadings are standardized.

Table 5: Second-order factor loadings for a three- and a four-factor solution

	Model 4a		*Model 4b*	
	λ	p	λ	p
SES	BY			
OBJKK	0.906	$< .001$	0.899	$< .001$
MIG	0.672	$< .001$	0.66	$< .001$
INKKK	0.343	$< .001$	0.38	$< .001$
SPANN			0.195	$< .001$
ELTERN12	0.758	$< .001$	0.753	$< .001$
CFI	0.975		0.949	
TLI	0.975		0.951	
RMSEA	0.06		0.067	
WRMR	2.286		2.75	
χ^2	358.881		747.832	
df	28		48	
p	$< .001$		$< .001$	

Note: All factor loadings are standardized.

As indicated by Rindskopf and Rose (1988), a one-factorial structure can be regarded as a special case of both a group-factor model and also of a second-order model. In second-order factor model terminology, a simple one-factorial measurement model is a second-order model whose first-order latent variable variances are all set to zero. Of course, we are able to impose this restriction on the data in order to directly test whether the second-order model without this restriction fits the data better. Since the two models are nested, it is possible to apply the Satorra-Bentler χ^2 difference test (Satorra and Bentler, 1999). The resulting Δ_{χ^2} of 1288.458 with 3 degrees of freedom indicates that imposing a one-factorial structure on the data at hand leads to a highly significant ($p < .001$) decrease in model fit and thus has to be considered as an empirically untenable assumption.

An even stronger restriction would arise if one would try to estimate *model 4b* in the IRT framework again. Statistically, this would equal to set all (first-order) factor loadings to 1 and also the (second order) factor variance fixed at 1. Hence, we can test whether this additional restriction compared to imposing the one-factorial structure on the data leads to another significant decrease in model fit – which is the case ($\Delta_{\chi^2} = 10586.117, df = 8, p < .001$).

Formative specification As indicated above, a perhaps more convenient specification for the latent **SES** variable is a *formative* measurement model (Bollen and Lennox, 1991; MacCallum and Browne, 1993; Diamantopoulos and Winklhofer, 2001) wherein each indicator is modeled as a *cause* of the latent variable rather than reversely. Table 6 lists the coefficients of two formative models: *Model 5a* is a formative specification of model 1a, and *model 5b* is a formative specification of the second-order model 4b. For model identification purposes, another variable as an outcome of the latent variable to be measured has to be specified. We opted for an item that had been discarded in the initial partial credit model, *ELTERN5b*.[5] As the one-factorial formative *model 5a* significantly differs from the data while the three-factorial formative *model 5b* does not, results are evidently in favour of the latter.

Table 6: Two formative measurement models

	Model 5a			*Model 5b*
	β			β
SES	ON	**SES**	ON	
ELTERN3A	0.036		**OBJKK**	0.473
ELTERN3B	0.022		**MIG**	0.722
ELTERN4A1	0.185		**INKKK**	-0.366
ELTERN4A3	-0.072			
ELTERN5A2	-0.019		**OBJKK**	ON
ELTERN5C2	0.162		ELTERN7	0.942
ELTERN5D2	-0.65		ABI	0.132
ELTERN5E2	-0.315			
ELTERN5F2	0.149		**MIG**	ON
ELTERN5G2	-0.008		ELTERN3A	0.309
ELTERN5H2	-0.534		ELTERN3B	-0.021
ELTERN5I2	-0.039		ELTERN4A1	0.629
ELTERN5L2	0.059		SCHUELER4	0.371
ELTERN7	0.122			
ABI	0.005		**INKKK**	ON
ELTERN12	-0.01		ELTERN5A2	0.537
SCHUELER4	0.111		ELTERN5C2	-0.943
			ELTERN5G2	0.481
ELTERN5B	ON		ELTERN5L2	-0.244
SES	-0.413			
			SES	ON
			ELTERN12	-0.197
			ELTERN5B	ON
			SES	-0.169
χ^2	144.11		χ^2	14.909
df	17		*df*	11
p	$< .001$		*p*	0.1867

Note: All regression weights are standardized.

[5]Since the models are just identified, no fit measures except χ^2 are provided.

4.2 Multilevel analyses

When individual-level measures should be used in order to classify contextual-level units, social scientists often simply aggregate indicators of a factor that have been considered well-fitting on the individual level. However, Marsh et al. (2009) argued that this may lead to considerable bias. Instead, it should first be tested if a factorial structure that was found to apply on the individual level also holds on the contextual level. Hence, we again start from the three-factorial second-order measurement model of model 3c and estimate this model separately on both student and school-class level (table 7, *model 6a*. We then impose the restriction of a one-factorial structure on the between-level by fixing the variances of the between-level first-order latent variables to zero. Since the models based on the $WLSMV$ estimator showed convergence problems, we switched to a robust ML estimator that also accounts for non-normal data. However, results should only be interpreted with caution – particularly since the output indicated a negative residual covariance matrix on the between-level. However, since the fit of the model is quite satisfactory, we see good reasons to use it for a first insight in the between-level structure of the data.

Apart from the overall acceptable model fit, results indicate in terms of within-level and between-level $SRMR$, that in both cases, the factor structure suits the individual level better than the contextual level. One might now ask whether a three-factorial second-order model is the better model for the school-class level given the data at hand. We applied the Satorra-Bentler χ^2 difference test (Satorra and Bentler, 1999) between that model *model 6a* and an alternative model with a three-factorial second-order model on the individual level but a simple one-factorial model on the school-class level (*model 6b*). The resulting Δ_{χ^2} of 17.354 with 3 degrees of freedom is highly significant ($p < .001$) which is a hint that the more simple one-factorial model also fits worse on the school-class level.

Finally, we also imposed the restriction of an IRT model by fixing all contextual-level factor loadings and the variance of a one-dimensional latent SES variable to unity. We then tested this model against the one-dimensional CFA model, and the resulting Δ_{χ^2} of 578.437 with 9 degrees of freedom again indicated a significantly worse fit of the more restricted model.

In sum, and cogizant of the nuisances regarding the estimation process described above, our tentative conclusion from these analyses would be that a three-factorial second-order measurement model suits the data best on both student and school-class level.

Table 7: Hierarchical confirmatory factor analysis

		Model 6a		*Model 6b*	
Within-Level		λ	p	λ	p
	OBJKK	BY			
	ELTERN7	0.649	$< .001$	0.658	$< .001$
	ABI	0.521	$< .001$	0.526	$< .001$
	MIG	BY			
	ELTERN3A	0.847	$< .001$	0.847	$< .001$
	ELTERN3B	0.685	$< .001$	0.686	$< .001$
	ELTERN4A1	0.835	$< .001$	0.835	$< .001$
	SCHUELER4	0.24	$< .001$	0.241	$< .001$
	INKKK	BY			
	ELTERN5A2	0.469	$< .001$	0.474	$< .001$
	ELTERN5C2	0.6	$< .001$	0.604	$< .001$
	ELTERN5G2	0.538	$< .001$	0.54	$< .001$
	ELTERN5L2	0.558	$< .001$	0.56	$< .001$
	SES	BY			
	ELTERN12	0.609	$< .001$	0.609	$< .001$
	SES	BY			
	OBJKK	0.759	$< .001$	0.746	$< .001$
	MIG	0.518	$< .001$	0.514	$< .001$
	INKKK	0.216	$< .001$	0.215	$< .001$
Between-Level					
	$\mathbf{M_O}$**BJKK**	BY			
	ELTERN7	0.997	$< .001$	0.934	$< .001$
	ABI	0.964	$< .001$	0.92	$< .001$
	$\mathbf{M_M}$**IG**	BY			
	ELTERN3A	1	$< .001$	1.001	$< .001$
	ELTERN3B	0.998	$< .001$	0.996	$< .001$
	ELTERN4A1	0.978	$< .001$	0.977	$< .001$
	SCHUELER4	0.596	0.001	0.586	0.001
	$\mathbf{M_I}$**NKKK**	BY			
	ELTERN5A2	0.754	$< .001$	0.59	$< .001$
	ELTERN5C2	0.781	$< .001$	0.64	$< .001$
	ELTERN5G2	0.648	$< .001$	0.62	$< .001$
	ELTERN5L2	0.422	0.003	0.3	0.016
	$\mathbf{M_S}$**ES**	BY			
	ELTERN12	0.945	$< .001$	0.931	$< .001$
	$\mathbf{M_S}$**ES**	BY			
	$\mathbf{M_O}$**BJKK**	0.9	$< .001$	1	999
	$\mathbf{M_M}$**IG**	0.986	$< .001$	1	999
	$\mathbf{M_I}$**NKKK**	0.763	$< .001$	1	999
	CFI	0.935		0.932	
	TLI	0.915		0.914	
	RMSEA	0.039		0.04	
	$SRMR_W$	0.05		0.05	
	$SRMR_B$	0.096		0.107	
	χ^2	441.330*		459.380*	
	df	84		87	
	p	$< .001$		$< .001$	

Note: All factor loadings are standardized.

5 Tentative conclusion and outlook

The objective of this paper was to test different measurement models of latent variables applied on a score of students' socio-economic status. In the theoretical section, we first briefly reviewed a couple of theories accounting for inequalities in educational opportunities. In the methodological section, an overview about Confirmatory Factor Analysis (CFA) and Item Response Theory (IRT) was provided.

Based on a German school sample, we first applied a partial credit model that eliminated a couple of misfitting items. We then used the remaining items to build up a CFA on both student and school-class level.

Preliminary findings suggest that on both levels of analysis, a three-factorial second-order factor model suits the data better than the more restrictive one-factorial solution and an even more restrictive constraint in terms of an IRT model.

In a revision of this paper, we aim to cross-validate the results obtained so far by means of latent class analysis (LCA) also on both student and school-class level.

References

Aish AM and Jöreskog KG (1990) A panel model for political efficacy and responsiveness: An application of LISREL 7 with weighted least squares. *Quality and Quantity* 24(4): 405–426.

Alexander K and Eckland BK (1975) Contextual effects in the high school attainment process. *American Sociological Review* 40(3): 402–416.

Alwin DF and Otto LB (1977) High school context effects on aspirations. *Sociology of Education* 50(4): 259–273.

Andersen PL and Hansen MN (2011) Class and Cultural Capital - The Case of Class Inequality in Educational Performance. *European Sociological Review* forthcoming.

Babakus E; Ferguson Jr CE and Jöreskog KG (1987) The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research* 24(2): 222–228.

Berkemeyer N; Bos W; Holtappels HG et al. (2010) „GanzIn“: das Ganztagsgymnasium in Nordrhein-Westfalen. In: Berkemeyer N; Bos W; Holtappels HG et al. (eds.) *Jahrbuch der Schulentwicklungsforschung - Band 16*, 131–152. Weinheim & München: Juventa.

Bollen KA (1989) *Structural Equations With Latent Variables.* New York: Wiley.

Bollen KA and Lennox R (1991) Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin* 110(2): 305–314.

Bond TG and Fox CM (2001) *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bonsen M; Bos W and Gröhlich C (2007) Die Relevanz von Kontextmerkmalen bei der Evaluation der Effektivität von Schulen. *Zeitschrift für Evaluation* 1: 165–174.

Bonsen M; Bos W; Gröhlich C et al. (2010) *Zur Konstruktion von Sozialindizes: Ein Beitrag zur Analyse sozialräumlicher Benachteiligung von Schulen als Voraussetzung für qualitative Schulentwicklung.* Berlin: Bundesministerium für Bildung und Forschung (BMBF), Referat Bildungsforschung.
URL: *www.bmbf.de/pub/bildungsforschung_band_einunddreissig.pdf*

Bos W; Bonsen M and Gröhlich C (2010) *KESS 7-Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen zu Beginn der Jahrgangsstufe 7*, vol. 5. Münster: Waxmann Verlag.

Bos W and Pietsch M (2006) *KESS 4-Kompetenzen und Einstellungen von Schülerinnen und Schülern am Ende der Jahrgangsstufe 4 in Hamburger Grundschulen.* Münster: Waxmann Verlag.

Boudon R (1974) *Education, Opportunity and Social Inequality: Changing Prospects in Western Society.* New York: John Wiley & Sons.

Bourdieu P (1973) Cultural Reproduction and Social Production. In: Brown R (ed.) *Knowledge, Education, and Cultural Change: Papers in the Sociology of Education*, 71–112. London: Tavistock Publications.

Bourdieu P (1977) *Outline of a Theory of Practice (Cambridge Studies in Social and Cultural Anthropology).* Cambridge: Cambridge University Press.

Bourdieu P (1986) *Distinction: A social critique of the judgement of taste.* London: Routledge.

Bourdieu P and Passeron JC (1990) *Reproduction in Education, Society and Culture.* London: Sage Publications Ltd.

Bowles S and Levin HM (1968) The determinants of scholastic achievement-An appraisal of some recent evidence. *Journal of Human Resources* 3(1): 3–24.

Breen R and Goldthorpe JH (1997) Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society* 9(3): 275–305.

Cain GG and Watts HW (1968) The controversy about the Coleman report: comment. *The Journal of Human Resources* 3(3): 389–392.

Cain GG and Watts HW (1970) Problems in making policy inferences from the Coleman Report. *American Sociological Review* 35(2): 228–242.

Chen FF; Sousa KH and West SG (2005) Teacher's corner: testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal* 12(3): 471–492.

Coleman JS (1966) *Equality of Educational Opportunity.* Washington, DC: U.S. Government Printing Office.

Coleman JS (1968) Equality of educational opportunity: Reply to Bowles and Levin. *The journal of human resources* 3(2): 237–246.

Coleman JS (1970) Reply to Cain and Watts. *American Sociological Review* 35(2): 242–249.

Coleman JS (1988) Social Capital in the Creation of Human Capital. *American Journal of Sociology* 94: 95–120.

Diamantopoulos A and Winklhofer HM (2001) Index construction with formative indicators: an alternative to scale development. *Journal of Marketing Research* 38(2): 269–277.

Ditton H (2010) Evaluation und Qualitätssicherung. *Handbuch Bildungsforschung* 607–623.

Erikson R and Jonsson JO (1996a) Introduction: Explaining Class Inequality in Education: The Swedish Test Case. In: *Can Education Be Equalized?: The Swedish Case in Comparative Perspective*, 1–63. Boulder, CO / Oxford: Westview Press.

Erikson R and Jonsson JO (1996b) The Swedish Context: Educational Reform and Long-term Change in Educational Inequality. In: *Can Education Be Equalized?: The Swedish Case in Comparative Perspective*, 65–93. Boulder, CO: Westview Press.

Esser H (1999) *Soziologie. Spezielle Grundlagen 1. Situationslogik und Handeln.* Frankfurt / New York: Campus Verlag.

Fox J (2006) Structural Equation Modeling With the sem Package in R. *Structural Equation Modeling* 13(3): 465–486.

Goldthorpe JH (1996) Class analysis and the reorientation of class theory: the case of persisting differentials in educational attainment. *British Journal of Sociology* 47(3): 481–505.

Jaeger MM (2009) Equal Access but Unequal Outcomes: Cultural Capital and Educational Choice in a Meritocratic Society. *Social Forces* 87(4): 1943–1971.

Jencks C (1972) Inequality: A reassessment of the effect of family and schooling in America. .

Jöreskog KG (1969) A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* 34(2): 183–202.

Jöreskog KG (1990) New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality & Quantity* 24(4): 387–404.

Jöreskog KG (1994) On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika* 59(3): 381–389.

Lareau A and Weininger EB (2003) Cultural capital in educational research: A critical assessment. *Theory and Society* 32(5): 567–606.

MacCallum RC and Browne MW (1993) The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin* 114(3): 533–541.

Marsh HW; Chit-Kwong K and Kit-Tai H (2000) Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology* 78(2): 337–349.

Marsh HW; Lüdtke O; Robitzsch A et al. (2009) Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research* 44(6): 764–802.

Marsh HW and Parker JW (1984) Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology* 47(1): 213–231.

Masters GN (1982) A Rasch model for partial credit scoring. *Psychometrika* 47(2): 149–174.

Masters GN (1988) The analysis of partial credit scoring. *Applied Measurement in Education* 1(4): 279–297.

McElroy FW (1967) A necessary and sufficient condition that ordinary least-squares estimators be best linear unbiased. *Journal of the American Statistical Association* 62(320): 1302–1304.

Meyer JW (1970) High school effects on college intentions. *The American Journal of Sociology* 76(1): 59–70.

Müller-Benedict V (2007) Wodurch kann die soziale Ungleichheit des Schulerfolgs am stärksten verringert werden? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59(4): 615–639.

Muthén BO (1984) A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* 49(1): 115–132.

Muthen BO (1994) Multilevel covariance structure analysis. *Sociological methods & research* 22(3): 376–398.

Nash R (2005) Boudon, Realism, and the Cognitive Habitus: Why an Explanation of Inequality / Difference Cannot be Limited to a Model of Secondary Effects. *Interchange* 36(3): 275–293.

Olsson UH (1979) Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* 44(4): 443–460.

Raju NS; Laffitte LJ and Byrne BM (2002) Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology* 87(3): 517–529.

Rasch G (1960) *Probabilistic models for some intelligence and attainment tests*. Koebenhavn: Danmarks paedagogiske Institut.

Raudenbush SW; Johnson C and Sampson RJ (2003) A multivariate, multilevel Rasch model with application to self–reported criminal behavior. *Sociological Methodology* 33(1): 169–211.

Rigdon EE and Ferguson Jr CE (1991) The performance of the polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data. *Journal of Marketing Research* 28(4): 491–497.

Rindskopf D and Rose T (1988) Second order factor analysis: Some theory and applications. *Multivariate Behavioral Research* 23(1): 51–67.

Rivkin SG; Hanushek EA and Kain JF (2005) Teachers, schools, and academic achievement. *Econometrica* 73(2): 417–458.

Sammons P (1999) *School effectiveness: Coming of age in the twenty-first century*, vol. 6. Lisse: Swets & Zeitlinger.

Satorra A and Bentler PM (1999) A Scaled Difference Chi-square Test Statistic for Moment Structure Analysis.
URL: *http://preprints.stat.ucla.edu/download.php?paper=260*

Scheerens J (2000) *Improving school effectiveness*. Paris: Unesco, International Institute for Educational Planning.

Sewell WH; Haller AO and Ohlendorf GW (1970) The educational and early occupational status attainment process: Replication and revision. *American Sociological Review* 35(6): 1014–1027.

Sewell WH; Haller AO and Portes A (1969) The educational and early occupational attainment process. *American Sociological Review* 34(1): 82–92.

Skrondal A and Rabe-Hesketh S (2007) Latent Variable Modelling: A Survey*. *Scandinavian journal of statistics* 34(4): 712–745.

Snijders TAB and Bosker RJ (1999) *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications Ltd.

Vermunt JK (2003) Multilevel latent class models. *Sociological methodology* 33(1): 213–239.

Wu M; Adams RJ; Wilson MR et al. (2007) *ACER ConQuest Version 2.0 - Generalized Item Response Modelling Software*. Camberwell, Victoria: ACER Press.

Zhang G and Browne MW (2006) Bootstrap fit testing, confidence intervals, and standard error estimation in the factor analysis of polychoric correlation matrices. *Behaviormetrika* 33(1): 61–74.

```
 +item
-------------------------------------------------------
          |                    |
          |                    |
          |                    |
          |                    |
          |6                   |
          |                    |
          |                    |
 2        |                    |
          |                    |
          |5                   |
          |                    |
          |                    |
          |                    |
         X|                    |
         X|                    |
         X|                    |
         X|                    |
        XX|7                   |
       XXX|12                  |
       XXX|                    |
       XXX|                    |
 1     XXX|                    |
   XXXXXXX|13                  |
    XXXXXX|                    |
    XXXXXX|                    |
  XXXXXXXX|                    |
 XXXXXXXXX|8                   |
  XXXXXXXX|                    |
XXXXXXXXXX|                    |
  XXXXXXXX|                    |
    XXXXXX|                    |
   XXXXXXX|14                  |
   XXXXXXX|10                  |
  XXXXXXXX|                    |
 0 XXXXXXXXX|                  |
   XXXXXXX|                    |
    XXXXXX|                    |
      XXXX|                    |
      XXXX|4                   |
      XXXX|9                   |
       XXX|                    |
        XX|15 16               |
        XX|2                   |
        XX|1                   |
         X|                    |
         X|3                   |
         X|                    |
-1       X|                    |
          |                    |
          |                    |
          |11                  |
          |                    |
          |                    |
          |                    |
          |                    |
          |                    |
          |                    |
          |                    |
          |17                  |
=====================================================
Each 'X' represents  20.7 cases
```

===> item

1	Eltern_3a
2	Eltern_3b
3	Eltern_4_a1
4	Eltern_4_a3
5	ELTERN_5_a2
6	ELTERN_5_c2
7	ELTERN_5_d2
8	ELTERN_5_e2
9	ELTERN_5_f2
10	ELTERN_5_g2
11	ELTERN_5_h2
12	ELTERN_5_i2
13	ELTERN_5_l2
14	ELTERN_7
15	abi
16	ELTERN_12
17	SCHÜLER_4

Figure 5: Distribution of social composition items in the partial credit model.