MEASURES OF SELF-REPORTED READING RESOURCES, ATTITUDES AND ACTIVITIES BASED ON LATENT VARIABLE MODELLING

Kajsa Yang Hansen, Göteborg University, Sweden
Monica Rosén, Göteborg University, Sweden
Jan-Eric Gustafsson, Göteborg University, Sweden

BACKGROUND

In 2001 IEA launched the first round of PIRLS, Progress in International Reading Literacy Study, a comparative study designed to measure grade 4 students reading literacy skills. The study is designed to be repeated every fifth year, and to provide policy makers, researchers, and practitioners with information about educational achievement and learning environments. Along with the reading achievement test data, other indicators of reading related factors were included the study, including information about the schools, classrooms and teachers, and also about home environments and the parents of the students. Compared to previous IEA studies of reading literacy the parental "home questionnaire" is one of the major developments.

Great effort was put into producing valid and reliable reading achievement scales, using matrix sampling and IRT techniques. Furthermore, a number of indices of reading related constructs were introduced in the international report (Mullis, Martin, Gonzalez and Kennedy, 2003). These were based on self-reported information from the students, their parents, their teachers and their school head-master. Examples of indices constructed to measure reading related constructs are the Home Educational Resources (HER), the Early Home Literacy Activities (EHLA), the Parents Attitudes Towards Reading (PATR), Students Attitudes Towards Reading (SATR), and the Student Reading Self-Concept (SRSC). These constructs are presumed to be of major importance for students reading achievement, so participating countries are rank ordered on these. However, from a technical point of view, the indices tend to be quite simple measures compared to the reading achievement scales, most of them being an average across the typically quite few observed indicator variables included in the measure (Gonzalez & Kennedy, 2003). Little is known about the validity of the indices, they may have low reliability and partial non-response may affect the results. One approach to dealing with these methodological problems may be to adopt a multivariate approach, were the questionnaire items are modelled in latent variable models using confirmatory factor analysis. This approach allows tests of fit of the hypothesized model to data, an optimal weighting of the influence of each of the observed variables, and improved possibilities for dealing with partial non-response to items. One way to make latent variables more easy to deal with in regular analysis procedures is to compute individual factor scores from the models, merge them with the data file, and treat them as observed variables in regular analyses, although hopefully with better estimates of the intended construct than is provided by an index variable.

One aim of the present study is to adopt such a latent variable approach and analyse the sets of items constructed to measure the five constructs, investigating dimensionality and model fit. Another aim is to investigate the possibility to produce individual factor scores from these models and to compare their measurement properties with those of the measures reported.

METHOD

In this section, methodological issues in the use of observed indices are discussed, and the variables and methods of analysis applied in the study are presented.

Problems in using observed index variables

As has already been observed there are methodological problems involved in using composite variables:

- Indicators may work differently across countries due to cultural differences. The composite or index of a construct may be unable to capture such differences.
- Usually an assumption of equal weighting is put on measures of an index but this need not hold true since a construct may explain different amounts of variance in different indicators.
- Missing responses for one or more of the components create problems when computing a composite score.
- Measurement errors and unreliability in indicators may bias the relationship of the composite or index of a construct with other constructs, such as various aspects of reading achievement.

Many of these problems may be approached through considering the multivariate nature of the indicators and constructs. Models may be fitted to the data, and evaluated with respect to goodness-of-fit and the measurement properties of the variables. This approach also allows estimation of factor scores for each individual, which may practically useful.

Data

Grade 4 PIRLS samples of six countries (i.e., Canada, England, Germany, Italy, Norway and Sweden) were included in the current study. Variables that were used to create the observed indices were also applied in the current multivariate analysis. In Table 1, the constructs and their indicators are presented. Altogether, 29 variables were involved in measuring the constructs. There were at least 4 indicators for each construct.

Table 1: Reading Related Factors and Their Indicators

Factors	Questions in Questionnaires	Indicators
HER	StQ19. About how many books are there in your home?	ASBGBOOK
Home	StQ20. Do you have any of these things in your home?	
Education	a. Computer	ASBGPS1
Resources	b. Study desk/table for own use	ASBGPS2
	c. Books of your very own	ASBGPS3
	d. Daily newspaper	ASBGPS4
	HQ13. About how many children's books are there in your home?	ASBHCHBK
	HQ14. What is the highest level of education completed by the	ASBHEDUF
	child's father and mother?	ASBHEDUM
SATR	StQ12. What do you think about reading? Tell how much you agree	
Students'	with each of these statements	
Attitudes	a. I read only if I have to	ASBGRST1
Towards	b. I like talking about books with other people	ASBGRST2
Reading	c. I would be happy if someone gave me a book as a present	ASBGRST3
	d. I think reading is boring	ASBGRST4
	e. I need to read very well for the future	ASBGRST5
	f. I enjoy reading	ASBGRST6

SRSC Students' Reading Self Concept	StQ13. How well do you read? Tell how much you agree with each of the following statements. a. Reading is very easy to me b. I do not read as well as other students in my class c. When I am reading by myself, I understand almost everything I read d. Reading aloud is very hard for me	ASBGRAB1 ASBGRAB2 ASBGRAB3 ASBGRAB4
EHLA Early Home Literacy Activities	HQ2. Before your child began primary school, how often did you or someone else in your home do the following activities with him or her? a. Read books b. Tell stories c. Sing songs d. Play with alphabet toys f. Play word games h. Read aloud signs and labels	ASBHAC1 ASBHAC2 ASBHAC3 ASBHAC4 ASBHAC6 ASBHAC8
PATR Parents' Attitudes Toward Reading	HQ11. Please indicate how much you agree with the following statements about reading. a. I read only if I have to b. I like talking about books with other people c. I like to spend my spare time reading d. I read only if I need information e. Reading is an important activity in my home	ASBHSTM1 ASBHSTM2 ASBHSTM3 ASBHSTM4 ASBHSTM5

Note: StQ = Student Questionnaire; HQ = Home Questionnaire. See also (Martin, Mullis & Kennedy, 2003)

In addition to these indicators, the standardized total reading achievement score derived from the mean of the 5 plausible values (INTTOT) was included in the analyses.

The observed indices were created in such a way that the mean of the indicators of each construct was calculated and categorized into 3 categories (e.g., high, medium and low). Students with partially or completely missing responses on the indicators of the construct were considered missing on the observed index. Table 2 shows the number of valid and missing cases, and the reliability of the observed indices in the six countries.

Table 2: Number of Valid Cases for the Observed Indices and Their Reliability

Observed Indices		CAN	ENG	GER	ITA	NOR	SWE
HER	Valid N	6287	1585	4489	3236	2988	5180
	Missing	1966	1571	3144	266	471	864
	Reliability	.55	.52	.60	.55	.52	.55
EHLA	Valid N	6596	1668	6289	3028	3002	5302
	Missing	1657	1488	1344	474	457	742
	Reliability	.71	.73	.58	.59	.65	.62
PATR	Valid N	6623	1690	6242	2936	3016	5440
	Missing	1630	1466	1391	566	443	604
	Reliability	.81	.84	.83	.79	.83	.84
SATR	Valid N	7842	3053	7081	3331	3258	5732
	Missing	411	103	552	171	201	312
	Reliability	.71	.73	.72	.56	.53	.71
SRSC	Valid N	7943	3102	7256	3435	3330	5833
	Missing	310	54	377	67	129	211
	Reliability	.57	.55	.53	.42	.57	.61

Note: CAN = Canada; ENG = England; GER = Germany; ITA = Italy; NOR = Norway; SWE = Sweden.

There frequency of missing values tended to be greater in the observed indices where items from the Parent Questionnaire are involved, such as HER, EHLA and PATR, which is because in some countries this questionnaire had a low response rate. The amount of missingness varied between countries. For England, for example, about half of the cases lacked information on HER, EHLA, and PATR. Germany also had a very high proportion of missing cases on HER. It is likely that the missingness in the parental indicators is not independent of the parents' attitude towards reading and students' reading achievement. The reliability of the indices was not very high, typically around .50 - .70.

Analytical method

Factor models can be used to examine the dimensionality of a set of indicators. The construct in such a model is a latent variable that causes variation in its indicators (Allen & Yen, 1979; Fornell, 1982; Gustafsson & Balke, 1993; Yang, 2003c). With confirmatory factor analysis (CFA, Bollen, 1989; Kline, 1998), it is possible to test the fit of different hypothesized models for a set of observed variables. The models are evaluated with respect to goodness-of-fit, and with respect to reliability and validity of the measurement provided by the included indicators.

The association of each indicator to its factor or factors is captured by the standardized factor loading. The squared factor loading reflects the amount of the variance in the indicators being explained by the factor. When an observed index is constructed from indicators that have equal variance an equal weight is assigned to each indicator, and this is not optimal when factor loadings do vary. Factor scores estimated from factor analysis may thus be expected to have better measurement properties than an observed index. Factor score estimates can be used as ordinary observed variables "for diagnostic purpose as well as inputs to a subsequent analysis" (Bentler & Yuan, 1997, p. 259; see also, Jöreskog, 2000; Jöreskog, Sörbom, du Toit & du Toit, 2000).

The factor analyses and factor score estimation were carried out with Mplus (Muthén & Muthén, 2001). One advantage of Mplus is that it uses partially missing data in model estimation and in the estimation of factor scores. Mplus detects missing patterns in the data and uses the available information efficiently through use of maximum likelihood estimation (Muthén & Muthén, 2001).

It must be observed that factor scores estimated at the individual level are not the factor, only a projection of factors onto the observed variables, which causes unreliability or lack of determinacy (i.e., correlation between the factor and the estimated factor; Rozeboom, 1988; Vittadini, 1989). There are many different methods, which may be used to estimate factor score. However, factor scores estimated by different methods often are highly correlated (Bollen, 1989), which means the indeterminacy is largely unrelated to the estimation method.

Mplus computes the factor score determinacy for each missing pattern. Rather low factor determinacy was observed for the patterns in which data is available only on few indicators of the factor. To achieve a determinacy at least around .70, cutting point of the maximum number of missing observations was therefore set for all the five factors. Thus, cases that have too few valid indicators are assigned a missing value code.

RESULTS

This section has three parts in which results from analyses of different data sets are presented. The development of latent variable models for each index was done on Swedish data and is presented first. The approach for estimating individual factor scores also was made on Swedish data, and is presented next. In the third part the factor score approach is further evaluated by bringing data from Canada, England, Germany, Italy and Norway into the analysis.

The measurement models

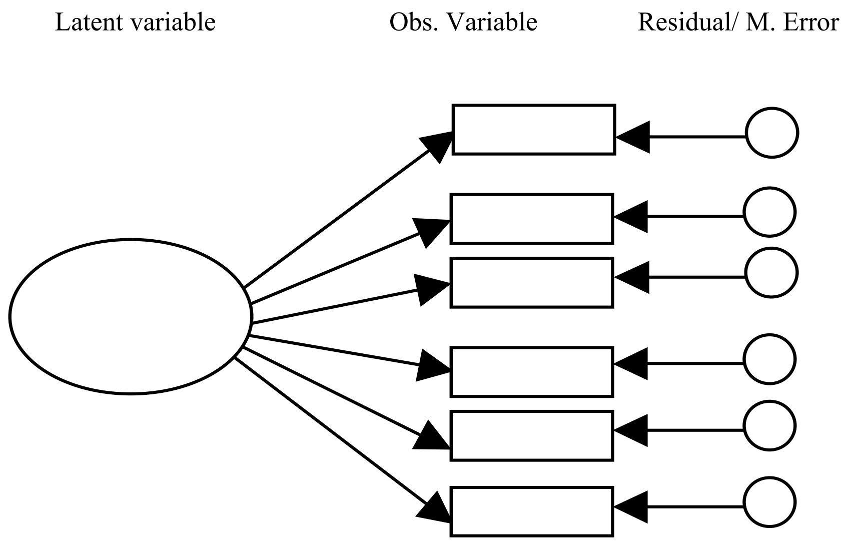
Below the latent variable models fitted for each group of observed variables are described.

The modelling procedure

The analysis started by fitting a basic one-factor measurement model to the observed indicators of Home Educational Recourses (*HER*), Early Home Literacy Activities (*EHLA*), Parents Attitudes Towards Reading (*PATR*), Students Attitudes Towards Reading (*SATR*) and Students Reading Self-Concept (*SRSC*). Each model was evaluated with respect to fit, and further steps to improve the model were taken if necessary to obtain better model fit. Figure 1 shows the general model.

In the basic one-factor measurement model the latent construct affects all the observed variables. No constraints are imposed on the relations, so that the degree of influence from the latent construct may vary across indicators. Each indicator is also affected by measurement error, which is the variance in the observed measure left unexplained by the latent factors. This residual represents a mixture of random error and factors specific to the observed variable. In the basic model the residuals in the observed variables variances are uncorrelated, but the model may be modified to allow covariances among the residuals. Another way to improve a poor-fitting model is to introduce more latent variables, for example as residual factors that are related to a subset of the observed measures. Which approach to chose depends mainly on how well additional factors can be established, but also on the meaningfulness for the purpose of the analysis.

Figure 1: A Basic One-Factor Measurement Model



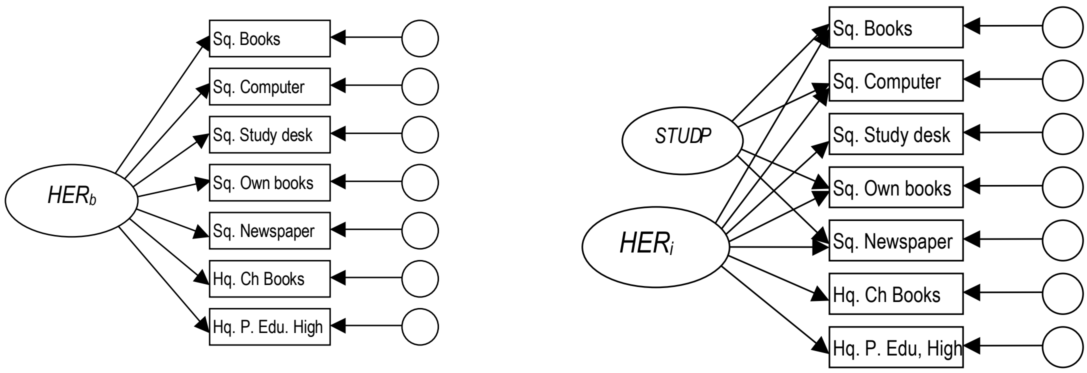
The question of model fit is an interesting one. On the one hand it could be argued that if a model does not fit the data, no meaningful interpretation of factor scores computed from this model may be made. On the other hand it may be argued that modifications of a model to achieve better fit may introduce unnecessary parameters which are not replicable, and which may disturb the estimation of factor scores. In order to investigate this issue comparisons have been made between factor scores computed on the basis of both unmodified and modified models.

Home educational resources

The reported index Home Educational Resources (*HER*) is formed as a sum of three questions from the Home questionnaire and five questions from the Student questionnaire. Parents have reported the estimated number of books and children's books at home, as well as their own educational level. Students' also have reported their estimated number of books at home, whether there is a computer at home, if there is a daily newspaper, if they have books of their own, and a study desk.

The fit of the basic model for *HER* (see Figure 2) was rather poor, which indicates that one factor may not be enough. The model was modified by introducing a second residual factor (*STUDP*) that was related to all the student possession items in the student questionnaire (see Figure 2), and this model fitted much better.

Figure 2: Latent Variable Models of Home Educational Resources



Basic Model
Chi-square = 1627.84; Df = 14; RMSEA = .138

Improved Model
Chi-square = 24.53; Df = 10; RMSEA = .016

The estimated standardized relations between the latent variable and the indicators (i. e., the factor loadings) are presented in Table 3 for the two models. The factor loadings were very similar in the two models. For *HER* the highest loading was found for the number of at books at home. Parents educational level also had a high loading, while the remaining variables of student and home possessions had rather weak relations to the *HER* construct. In the improved model, the *STUDP* factor had its strongest relations to the two items asking about students' own possessions.

Table 3: Factor Loadings in Basic and Improved Model of Home Educational Resources

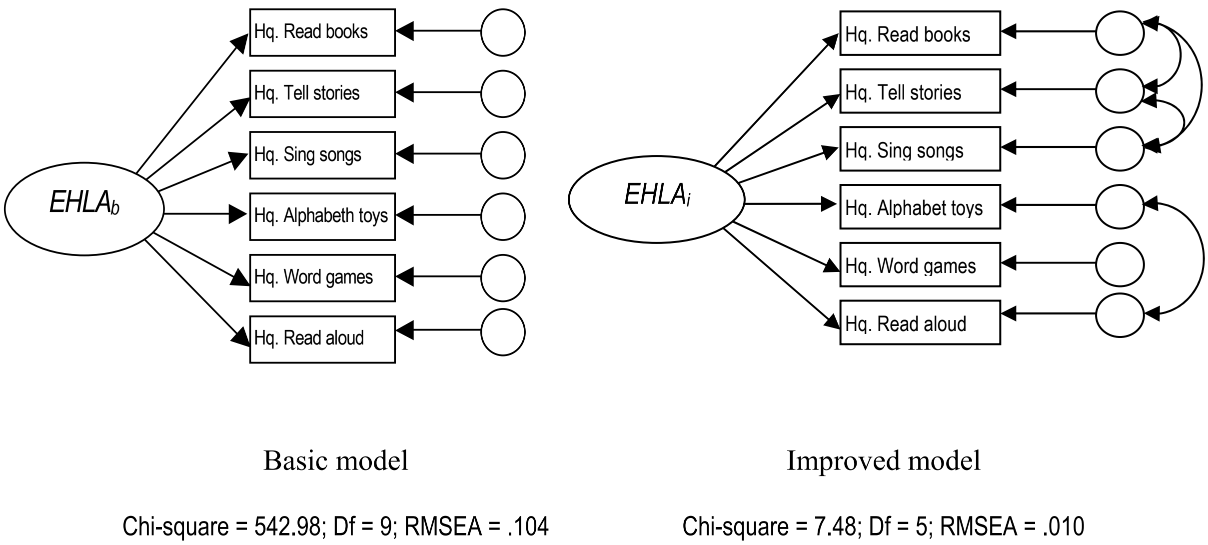
Obs variable	HER_b Basic model	HER_i Improved model	*STUDP*
Hq. N of children's books at home	0.65	0.68	
Hq. Parents' highest education level	0.48	0.49	
Sq. Books at home	0.66	0.68	0.03
Sq. Computer at home	0.28	0.28	
Sq. Study desk at home	0.26	0.16	0.53
Sq. Own books	0.26	0.14	0.82
Sq. Daily newspaper at home	0.26	0.19	0.33

It may thus be concluded that the *HER* index is not uni-dimensional since at least two latent variables could be identified from this set of indicators. However, the general *HER*-factor was quite similarly defined when uni-dimensionality was assumed, and when a second factor was allowed.

Early home literacy activities

The Early Home Literacy Index (*EHLA*) was based on six questions in the Home questionnaire, which asked how often the parents were engaged in various literacy activities with the child before grade 1. The observed index EHLA was an average across these six items.
The basic *EHLA* model did not fit well (see Figure 3). However, there were too few variables to allow identification of an additional latent variable. The model therefore was modified through introduction of covariances among residuals, which caused model fit to be acceptable according to standard criteria (Figure 3).

Figure 3: Latent Variable Models of Early Home Literacy Activities



The factor loadings are presented in Table 4. The loadings were of similar size for all indicators in the basic model, while they tended to vary somewhat more in the improved model. Play word games and play with alphabet toys had somewhat lower loadings in the improved model. *EHLA* accounted for 14 - 26% of the variance in the observed variables.

Table 4: Factor Loadings in Basic and Improved Models of Early Home Literacy Activities

Observed variable	$EHLA_b$ Basic model	$EHLA_i$ Improved model
Hq. Read books	0.43	0.33
Hq. Tell stories	0.45	0.37
Hq. Sing songs	0.45	0.38
Hq. Play w alphabet toys	0.54	0.49
Hq. Play word games	0.51	0.45
Hq. Read aloud signs and labels	0.43	0.50

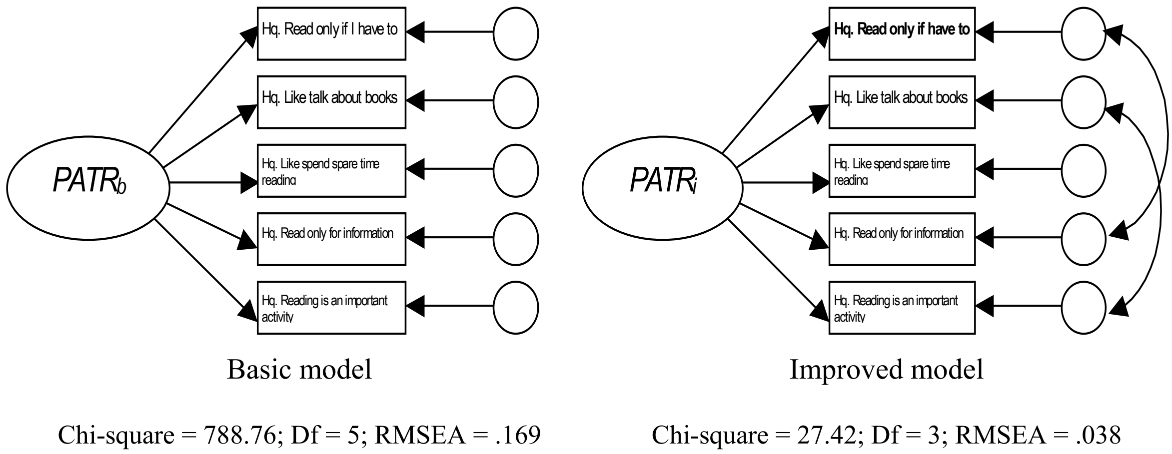
It should be noted that all the indicators of EHLA were retrospective questions, and retrospective questions are known to have poor reliability and validity.

Parents' attitudes towards reading

Parents' Attitudes Towards Reading (*PATR*) was indicated by four questions in the Home questionnaire, where the parent marked to what extent she or he agreed with different statements. Some of the statements were negatively phrased and the coding was therefore reversed before the average across items was computed.

The basic one-factor model did not have acceptable model fit (see Figure 4). An improved model, which included a covariance between two pairs of residuals, fitted the data quite well (see Figure 4).

Figure 4: Latent Variable Models of Parent's Attitude Towards Reading



In Table 5 the factor loadings of *PATR* are presented. It may be noted that the loadings were quite sizable in both models. The highest loading was observed for the statement whether the parents' enjoy reading on spare time. Three of the statements had loadings around .70, while for the statement "reading is an important activity in my home" the lowest loading was observed.

Table 5: Factor Loadings In Basic and Improved Model of Parents' Attitudes Towards Reading

Observed variable	$PATR_b$ Basic model	$PATR_i$ Improved model
Hq. Read only if have to	0.74	0.67
Hq. Like talking about books	0.70	0.69
Hq. Like reading on spare time	0.83	0.90
Hq. Read only for info seek	0.76	0.68
Hq. Reading valued activity at home	0.59	0.57

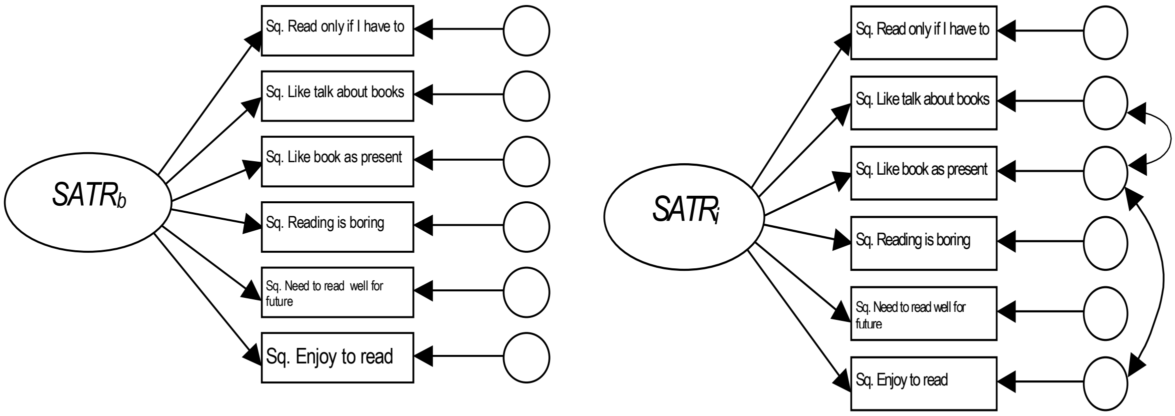
The correlated residuals showed that there was more information to extract from these data. However, in order to properly identify additional factors more indicators would be needed.

Students' attitudes towards reading

The indicators of Students' Attitudes Towards Reading (*SATR*) were quite similar to the indicators of *PATR*, even though there were six statements instead of five. Two negatively phrased statements were coded in reverse.

Figure 5 presents the models of *SATR*. The basic one-factor model had relatively good fit, although the ratio between chi square and degrees of freedom indicated that there was some room for improvement. In the improved model a covariance for one pair of residuals was introduced.

Figure 5: Latent Variable Models of Students' Attitude Towards Reading



Basic model
Chi-square = 312.11; Df = 9; RMSEA = .076

Improved model
Chi-square = 87.51; Df = 7; RMSEA = .044

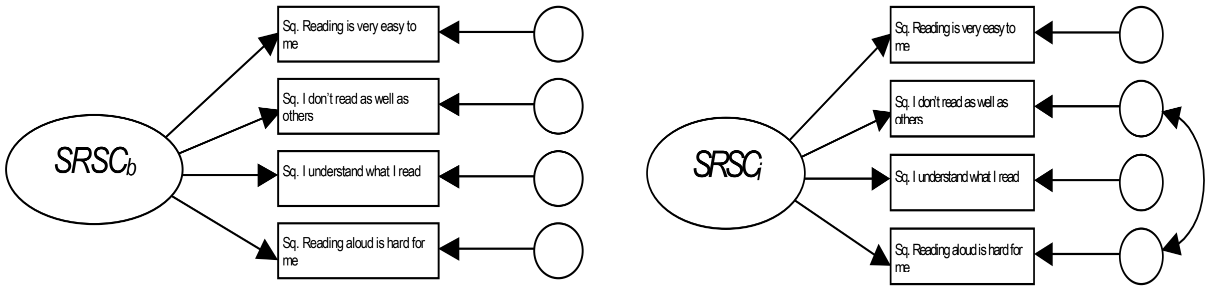
The factor loadings of *SATR* varied considerably over indicators but were highly similar for the two models (see Table 6). The highest loadings were observed for "reading is boring" and "I enjoy reading", where SATR accounted for about 70 % of the variance in the indicators. The lowest loading (.10) was observed for the statement "I need to read well for the future." This is probably because this statement holds true for most students regardless of their attitude towards reading.

Table 6: Factor loadings in basic and improved models of Students' Attitudes Towards Reading

Observed variable	$SATR_b$ Basic model	$SATR_i$ Improved model
Sq. Read only if have to	0.52	0.53
Sq. Like talking about books	0.38	0.37
Sq. Appreciates book as present	0.55	0.54
Sq. Reading is boring	0.84	0.85
Sq. Need to read well for the future	0.10	0.09
Sq. Enjoy reading	0.88	0.88

The index of Students' Reading Self-Concept (*SRSC*) was based on four statements about the students' own reading abilities. Two items (no 2 and 4) were coded in reverse because of the phrasing of the statements.
Two one-factor models were fitted to the data (see Figure 6). Like for the other indices, the basic model did not fit the data very well. However, adding a covariance between the residuals of the two negatively phrased statements, the model fit improved dramatically. Since the chi-square was smaller than the degrees of freedom in the improved model, RMSEA could not be computed for this model.

Figure 6: Latent Variable Models of Students' Reading Self-Concept



Basic model
Chi-square = 124.24; Df = 2; RMSEA = .102

Improved model
Chi-square = 0.29; Df = 1

The highest loadings were observed for items asking about how easy the student experienced reading to be (Table 7). The item "I understand almost everything I read" had the lowest loading.

Table 7: Factor Loadings In Basic And Improved Models Of Students' Reading Self-Concept

Observed variable	$SRSC_b$ Basic model	$SRSC_i$ Improved model
Sq. Reading is easy	0.76	0.66
Sq. I read not as well as others	0.64	0.72
Sq. I understand almost everything I read	0.40	0.28
Sq. Read aloud is hard	0.46	0.48

The amount explained variance in the observed variables by the latent *SRSC* factor varied between 8-52%, so we can conclude that the indicators are quite unevenly related to the latent variable.
Summing up this first phase of modelling, it can be concluded that the basic one-factor model did not have acceptable fit for any of the constructs. However, by modifying the models, mainly by introducing covariances among residuals, good fit was obtained. Another observation is that the factor loadings were in general not very affected by the improvements made to the model.

Factor scores

Before factor scores were estimated, the measurement models for each construct were gathered in an oblique model (i. e. a model with covariances among the latent variables). This was done separately for the basic models and the improved models, to allow comparison between factor scores estimated the two groups of models. There were several reasons for gathering the one-factor models into an oblique model and use this model for estimating the factor scores. One was that more information was made available for estimating the factor scores, which improves the precision of the individual factor score estimates. Another reason was the possibility to estimate factor scores for those cases that had partial missing data in the observed variables, due to the more effective use of the information available. Yet another reason was the practical convenience compared to the alternative procedure to estimate factors scores for one factor at a time.

Table 8: Latent Variable Correlations in the Oblique Models

	Basic models						Improved models				
	HER_b	$EHLA_b$	$PATR_b$	$SATR_b$	$SRSC_b$		HER_i	$EHLA_i$	$PATR_i$	$SATR_i$	$SRSC_i$
HER_b	1					HER_i	1				
$EHLA_b$	.50	1				$EHLA_i$	.59	1			
$PATR_b$	.52	.45	1			$PATR_i$	.52	.50	1		
$SATR_b$	.20	.15	.15	1		$SATR_i$	.19	.15	.15	1	
$SRSC_b$	.26	.23	.16	.49	1	$SRSC_i$	.26	.23	.16	.49	1
Model fit: Chi-2= 5884.13; df = 368; RMSEA = .050						Model fit: Chi-2= 2578.11; df = 350; RMSEA = .032					

The estimated correlations between the factors (see Table 8) ranged between .14 (*EHLA* and *SATR*) and .58 (*HER* and *PATR*). As can be noted, the estimates did not differ much between models. High correlations were found between Home Educational Resources, Early Home Literacy Activities and Parents Attitudes Towards Reading, which factors all reflect characteristics of the home. High correlations were also found between Students Attitudes Towards Reading and Students' Reading Self-Concept. Although further analyses of the interrelationships between these latent variables are warranted, such analyses are outside the scope of the current paper.
As was described in the methodology section the missing data modelling allows estimation of factor scores from partially missing data. In fact, when individual factor scores were estimated from the models, factor estimates was obtained for all cases regardless of the amount of information available for each case. Thus, even when a student had not responded to any of the items related to a latent variable an estimate was assigned. In this case, however, the determinacy was 0.0, and when few items had been responded to the determinacy was low. In order not to jeopardize the reliability of individual estimates, the factor score estimate was replaced with a missing code for those cases that had less than a certain number of valid responses for a factor.

Table 9 presents the correlations between the two estimated factor scores for each index. The factor labels are the same as those for the latent variables, although no longer in italics as their latent counterpart. The correlations were all close to unity, the lowest correlation (.91) being observed for EHLA. Thus, the factor scores derived from the basic and the improved models seem to be more or less interchangeable. Results from further analyses involving the two sets of factor scores are, however, reported below.

Table 9. Correlations of Factor Scores Based On Basic and Improved Oblique Models

Measures	Correlation
$\mathrm{HER_b}$ with $\mathrm{HER_i}$	.976
$\mathrm{PATR_b}$ with $\mathrm{PATR_i}$	.990
$\mathrm{EHLA_b}$ with $\mathrm{EHLA_i}$	.970
$\mathrm{SATR_b}$ with $\mathrm{SATR_i}$	.998
$\mathrm{SRSC_b}$ with $\mathrm{SRSC_i}$	.979

It should be observed that while the latent variables of a CFA model are error-free, this is not true for the factor scores estimated for each individual. This is because the factor scores are estimated from the error-laden observed variables. Still, however, factor score estimates may be expected to be better compared to the corresponding observed index, since in this approach the information in the data is used more effectively.

In the next step of the analyses several sets of correlations were computed between different ways to measure the constructs and reading achievement. One set of correlations was computed with the observed index variables, another with factor scores from the basic oblique model, yet another with factor scores from the improved models, and finally one set of correlations were estimated with the latent variables, through adding the reading achievement variable to the oblique (improved) model. The results are presented in Table 10.

Table 10: Correlations With Reading Achievement

Construct	No model *obs. scores*	Basic model *Factor scores$_b$*	Improved model *factor scores$_i$*	Improved model *Latent variables*
Home Educational Resources	.29	.52	.40	.46
Early Home Literacy Activities	.15	.27	.28	.31
Parents' Attitudes Towards Reading	.20	.27	.26	.25
Students' Attitudes Towards Reading	.32	.40	.37	.36
Students' Reading Self-Concept	.38	.67	.51	.55

The highest correlations were obtained with the factor scores, and generally the results were highly similar for factor scores estimated from the basic and improved models. However, for EHLA the factor score estimated from the improved model was higher. This may have to do with the specification of the improved model for EHLA, in which introduction of the covariance between two of the residuals caused the factor loading of one of the indicators (play word games) to drop considerably.

It is interesting to note that the correlations with the observed indices generally were considerably lower than the correlations with the factor scores. This is because data is used less efficiently in the observed indices.

The correlations obtained with the latent variables and Reading Achievement in the oblique models generally were a bit lower than those obtained with individual factor scores, which is somewhat surprising. However, in this case the comparison may not have been entirely correct due to differences in the treatment of missing data. In the latent variable model all the cases that are lacking information were included, while for the individual factor scores some of them were assigned a missing code according to the rules described in the method part. Nevertheless, the general conclusion is that the estimated factor scores seemed to work excellently, and that the factor score approach seems to be a very promising alternative to the observed indices for these five constructs.

Whether this factor score approach can be equally fruitful for comparative data was investigated next.

Cross country comparisons

Next the two oblique models for each index were applied on the data from five other countries in the PIRLS 2001 database. For comparative purposes the results from the analysis of the Swedish data are also included. In Table 11 results from the tests of model fit are reported.

Table 11: Model Fit Indices For The Oblique Models of Self Reported Literacy Factors In Six Countries

Country	Oblique basic factor model			Oblique improved factor model		
	Chi-square	df	RMSEA	Chi-square	Df	RMSEA
Canada	6067.25	368	.043	3693.57	350	.034
England	2009.87	368	.038	1348.59	350	.030
Germany	6251.91	368	.046	4327.11	350	.039
Italy	2807.50	368	.044	1773.21	350	.034
Norway	3634.60	368	.051	2195.80	350	.039
Sweden	5884.13	368	.050	2578.11	350	.032

For both models, the RMSEA indices were acceptable and quite similar across countries for both models. For all countries the improved model had better fit, even though the fit advantage was greater for Sweden than for the other countries. This is because the model modifications were done on the Swedish data and all the modifications do not replicate for the other countries. The factor loadings were in both models with very few exceptions quite similar to each other and across countries. The correlations between the two factor scores for each index also were close to unity for the five indices in all six countries.

In the next step, correlations were computed between reading achievement and the different measures of the intended construct.

Table 12: Correlations With Reading Achievement

Read Ach With Home Educational Resources (HER)	No model obs. scores	Basic model factor scores$_b$	Improved model factor scores$_i$	Improved model *Latent variable*
Canada	.29	.53	.39	.37
England	.31	.60	.47	.52
Germany	.34	.64	.54	.56
Italy	.23	.52	.40	.43
Norway	.30	.57	.41	.46
Sweden	.29	.52	.40	.46
Read Ach with Early Home Literacy Activities (EHLA)	**No model obs. scores**	**Basic model factor scores$_b$**	**Improved model factor scores$_i$**	**Improved model *Latent variable***
Canada	.17	.31	.30	.30
England	.18	.36	.34	.32
Germany	.13	.27	.37	.34
Italy	.15	.26	.27	.27
Norway	.22	.33	.31	.33
Sweden	.15	.27	.28	.31
Read Ach with Parents' Attitudes Towards Reading (PATR)	**No model obs. scores**	**Basic model factor scores$_b$**	**Improved model factor scores$_i$**	**Improved model *Latent variable***
Canada	.20	.28	.26	.25
England	.25	.31	.29	.28
Germany	.26	.34	.32	.30
Italy	.24	.31	.28	.28
Norway	.20	.25	.23	.21
Sweden	.20	.27	.26	.25

Read Ach with Students' Attitudes Towards Reading (SATR)	No model obs. scores s	Basic model factor scores$_b$	Improved model factor scores$_i$	Improved model *Latent variable*
Canada	.33	.42	.39	.39
England	.30	.41	.37	.36
Germany	.33	.41	.38	.38
Italy	.23	.31	.27	.26
Norway	.25	.41	.37	.42
Sweden	.32	.40	.37	.36
Read Ach with Students' Reading Self-Concept (SRSC)	No model obs. scores	Basic model factor scores$_b$	Improved model factor scores$_i$	Improved model *Latent variable*
Canada	.39	.66	.48	.58
England	.39	.69	.50	.60
Germany	.38	.64	.47	.52
Italy	.30	.56	.37	.42
Norway	.45	.73	.56	.66
Sweden	.38	.67	.51	.55

As may be seen in Table 12 the results already reported for Sweden generalized to the other countries as well. Thus, for almost all constructs the correlations with observed indices were considerably lower than the correlations with the estimated factor scores. The estimates of the correlations of a particular construct with reading achievement generally were highly similar over countries, even though it may be noted that in several instances somewhat lower correlations were observed for Italy than for the other countries. Another general finding was that the latent variable correlation with reading achievement was either equal or lower than the factor score correlations, but higher than the observed score correlation.

Discussion and Conclusions

The results reported here indicate that the factor score estimation approach may be a very useful alternative to the observed score indices that are currently used in the analyses of questionnaire data in the IEA studies and other surveys. The main advantage is that the severe attenuation of relations between the constructs and other variables caused by the low reliability (0.50 – 0.70) of the observed indices is avoided. The systematic downward bias of estimates of correlations caused by errors of measurement certainly has had the implication that many constructs measured with questionnaires have been dismissed or underestimated as explanatory variables.

Another advantage of the procedures used here is that they allow estimation of factor scores from cases who have not responded to all the items. This allows fuller use of the available data, and reduces the risk of bias caused by list-wise deletion of cases with missing data. It should be stressed, however, that care must be taken to assign missing data codes to those cases for which the available information is insufficient to allow reliable estimation of the factor score, since otherwise noise will be introduced into the data.

It is interesting to note that the factor score estimation method deals with missing data in a similar way as does the IRT procedure used to estimate achievement from the matrix-sampling designs used in many large-scale assessments. A possible future development may be to apply matrix-sampling design ideas to the design of questionnaires as well, allowing for broader coverage of the constructs.

Yet another advantage of the factor score estimation procedure is that the resulting scores may be used in standard statistical analyses, such as regression analysis or multi-level analysis. Latent variable models, which are another alternative, tend to become unwieldy when they include measurement models for large sets of variables and many groups of cases. Analyses based on factor scores may bring many of the advantages of such models, but with greater ease and simplicity.

One of the somewhat surprising findings of the study was that the estimated factor scores tended to have higher correlations with reading achievement than did the error-free latent variables in the measurement model. However, since this finding is based on comparisons between the results from analyses based upon a somewhat different number of cases care should be exercised in interpreting this result.

According to the model fit criteria employed in CFA most of the basic one-factor models fitted the data extremely poorly, while the improved models fitted the data very well. Still the factor scores computed from the basic and the improved models correlated close to unity for most of the constructs, and they had similar correlations with reading achievement. These patterns of results suggest that model fit is not important for the usefulness of the factor scores. One reason for this may be that the estimated factor scores mainly are functions of the un-standardized factor loadings, and these did not differ much between the basic and improved models. It is, however, conceivable that other types of modifications than those employed here may affect the factor loadings more strongly, which in turn may cause less agreement between factor scores estimated from unmodified and modified models. Further research seems to be needed on this issue.

References

Adams, M. J. (1990). *Beginning to read: thinking and learning about print*. Cambridge, Mass.: MIT Press.

Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, Calif.: Brooks/Cole Pub. Co.

Bentler, P. M. & Yuan, K.-H. (1997). Optimal Conditionally Unbiased Equivariant Factor Score Estimators. In M. Berkane (Ed.), *Latent variable modeling and applications to causality*. New York: Springer.

Bollen, A. K. (1989). *Structural Equations with Latent Variables*. New York: Wiley & Sons.

Campbell, J. R., Kelly, D. L., Mullis, I. V. S., Martin, M. O. & Sainsbury, M. (2001). *Framework and Specifications for PIRLS Assessment 2001*. PIRLS International Study Center, Lynch School of Education, Boston College, Chestnut Hill MA: the United States.

Fornell, C. (Ed.). (1982). *A Second Generation of Multivariate Analysis* (Vol. 1). New York: Praeger Publishers.

Gustafsson, J.-E. & Balke, G. (1993). General and Specific Abilities as Predictors of School Achievement. *Multivariate Behavioral Research, 28*(4), 407-434.

Jöreskog, K.-G. (2000). *Latent Variable Scores and Their Uses*, from
http://www.ssicentral.com/lisrel/column6.htm

Jöreskog, K.-G., Sörbom, D., du Toit, S. & du Toit, M. (2000). *LISREL 8: New Statistical Features*. Chicago: Scientific Software International.

Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling*. New York: Guildford Press.

Martin, M. O., Mullis, I. V. S. & Kennedy, A. M. (2003). *PIRLS 2001 technical report*. Chestnut Hill, MA: IEA International Association for the Evaluation on educational Achievement.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J. & Kennedy, A. M. (2003). *PIRLS 2001 International report : IEA' s study of reading literacy achievement in primary schools*. Chestnut Hill, MA: IEA International Association for the Evaluation of Educational Achievement.

Muthén, L. K. & Muthén, B. O. (2001). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.

Rozeboom, W. W. (1988). Factor Indeterminacy: The Saga Continuous. *British Journal of Mathematical and Statistical Psychology, 41*, 209-226.

Vittadini, G. (1989). Indeterminacy Problems in the LISREL Model. *Multivariate Behavioral Research, 24*, 397-414.

Yang, Y. (2003c). *Measuring Socioeconomic Status and its Effects at Individual and Collective Levels: A Cross-Country Comparison*. Gothenburg: Acta Universitatis Gothonburgenis, Gothenburg Studies in Educational Sciences 193.