

Factor Analysis of a School Context Scale from the TIMSS 2019 Study

Introduction to Factor Analysis and Item Response Theory

Fernando Mora

The TIMSS study on math and science achievement monitors and benchmarks fourth and eighth graders across 64 countries (although my query yielded only 39), along with a multitude of background and context survey items administered to school leaders and content area teachers. I have used the data set for a number of methods courses testing, for example, the generalizability of administering different testlets within classrooms, and Bayesian estimation of teacher background characteristics in students’ final math achievement scores. In the spirit of this class’ topic, and my own curiosity about the properties of the school context scales that teachers and principals answer, I chose to conduct an exploratory factor analysis of one school context scale, and parse whether EFA and CFA models would yield different fit statistics and factor loadings across respondent groups (math, science, or admin) as well as countries.

Table 1: Sample Sizes per group in the international and US datasets

group	International	US
admin	7119	241
math	9643	407
sci	13678	424

Like its civics and social science counterpart PISA, TIMSS attempts to benchmark and contrast international achievement with a standardized instrument, which I find lofty and not a little bit messy given how strongly cultural and material differences may influence across-country variation. Although student achievement scores and covariates are attentively given a number of weighting covariates in TIMSS’ own prescribed analysis recommendations, no such attention is given (to my knowledge) to the context questionnaires measuring teacher and principal attitudes and beliefs.

I am entering this project with the hypothesis that factor analyses using the complete international data set will be noisy and ill-fitting in terms of any latent belief uncovered by the models, and that these should narrow within countries, and more so between respondent groups within a country, representing a more coherent set of shared beliefs across individuals working under similar political and cultural macro-contexts. For this exercise, I will be performing EFAs on a split international dataset, a split international set for admins and for math teachers, and the US and UAE sets. These last two address a concern I have regarding coherence of teacher and admin beliefs between a large heterogeneous country (like the United States), and a smaller, more homogeneous country like the United Arab Emirates.

Table 2: Mean and SD for 3 sample items across all participating countries

country	Teachers' Expectations		Parental Expectations		Students' Desire	
	bg06c_mean	bg06c_sd	bg06g_mean	bg06g_sd	bg06i_mean	bg06i_sd
36	1.96	0.75	2.25	0.90	2.56	0.82
48	2.04	0.76	2.19	0.86	2.38	0.84
152	2.17	0.83	2.51	0.97	2.66	0.85
158	2.08	0.70	2.50	0.91	2.63	0.76
196	2.17	0.79	2.08	0.89	2.66	0.92
246	2.37	0.61	2.55	0.60	2.70	0.64
250	1.98	0.67	2.32	0.84	2.73	0.73
268	2.38	0.63	2.36	0.71	2.74	0.66
344	2.21	0.72	2.41	0.83	2.59	0.79
348	2.29	0.63	2.51	0.82	2.86	0.73
364	2.25	0.74	2.37	0.88	2.59	0.87
372	1.81	0.72	2.20	0.92	2.43	0.76
376	2.09	0.74	2.17	0.75	2.39	0.73
380	2.45	0.69	2.20	0.70	2.73	0.60
392	2.44	0.71	2.27	0.84	2.39	0.76
398	2.12	0.56	2.04	0.59	2.06	0.66
400	2.43	0.81	2.83	0.86	2.62	0.89
410	1.87	0.82	1.94	0.84	2.09	0.85
414	2.24	0.69	2.41	0.79	2.57	0.81
422	2.33	0.73	2.69	0.89	2.50	0.85
440	2.05	0.60	2.24	0.60	2.61	0.62
458	1.66	0.59	1.91	0.72	2.41	0.77
504	2.58	0.73	3.32	0.95	2.98	0.78
512	2.19	0.69	2.53	0.90	2.40	0.89
554	1.87	0.70	2.04	0.77	2.14	0.70
578	2.15	0.59	2.32	0.62	2.36	0.67
620	2.61	0.74	2.45	0.82	2.98	0.72
634	1.88	0.72	1.89	0.73	2.30	0.82
642	2.07	0.71	2.19	0.85	2.64	0.74
643	2.58	0.59	2.30	0.58	2.85	0.59
682	2.07	0.77	2.28	0.90	2.20	0.87
702	2.00	0.67	2.20	0.76	2.44	0.81
710	1.94	0.79	2.47	1.09	3.06	0.95
752	1.96	0.75	2.42	0.82	2.44	0.74
784	1.87	0.77	1.91	0.81	2.23	0.88
792	2.23	0.80	2.30	1.01	2.69	0.83
818	2.29	0.68	2.65	0.83	2.45	0.80
840	1.95	0.77	2.58	0.92	2.77	0.82
926	1.65	0.65	3 2.05	0.85	2.26	0.81

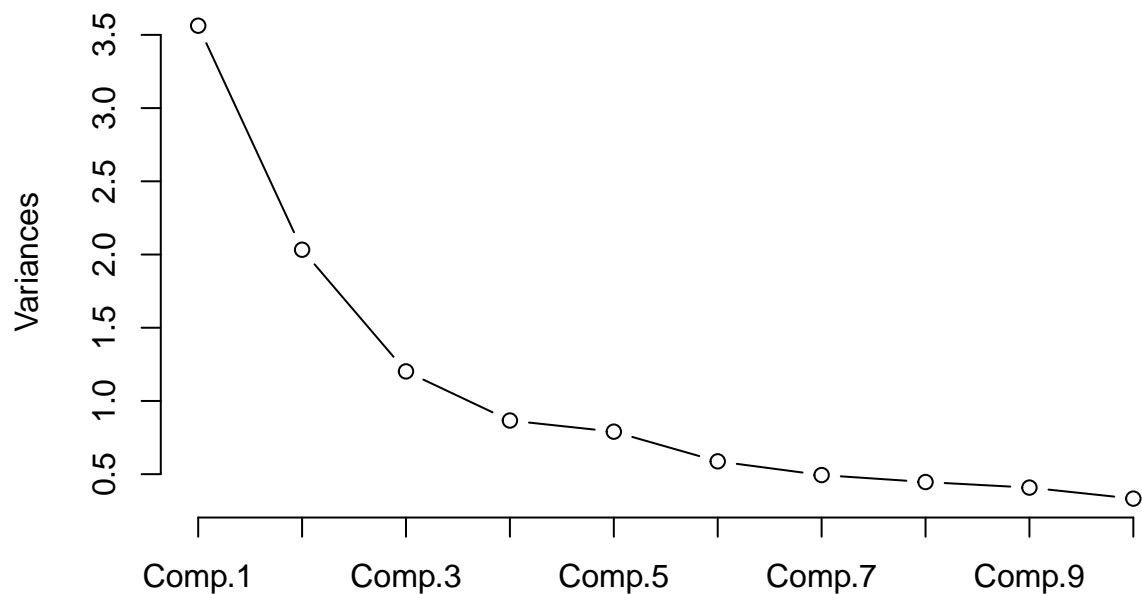
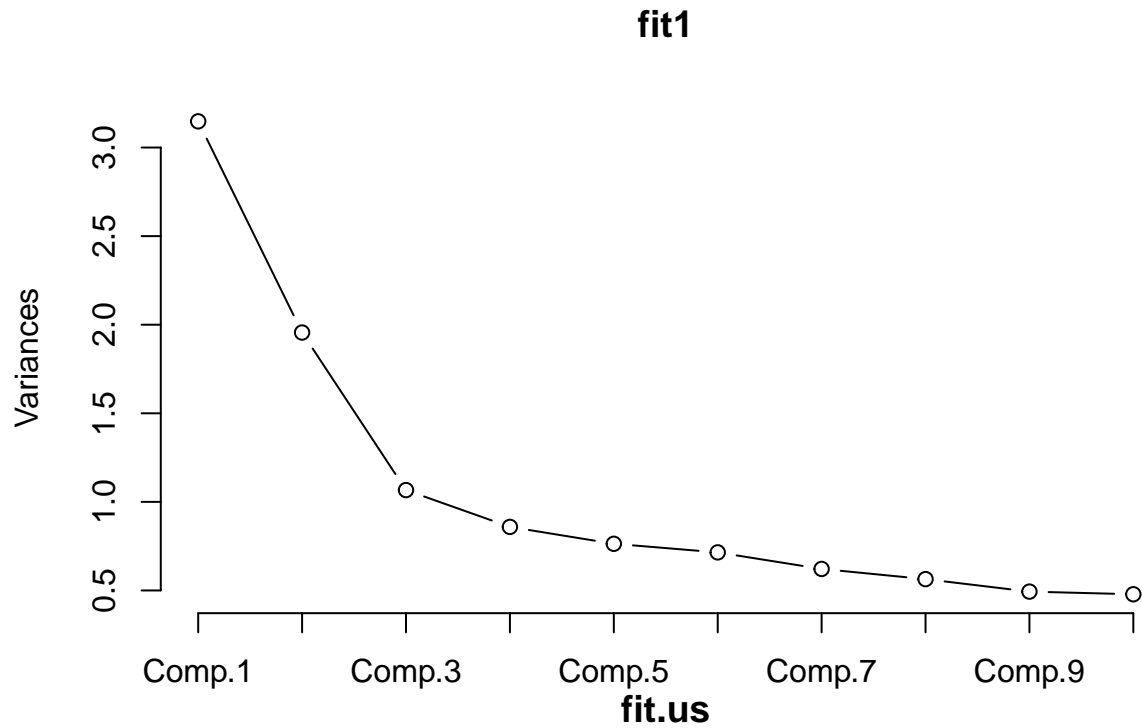
Above, we can appreciate how different countries' scores were from in just three of the items: "Teachers' expectations for student achievement"; Parental expectations for student achievement"; and "Students' desire to do well in school".

The lettered list below shows verbatim the questions that participating teachers, as well as school principals, are asked regarding their beliefs about their school's teacher, parent, and student attitudes regarding, for lack of a better cohering description, academic *excellence*. The wording is identical across the two questionnaires administered to either principals or math and science teachers, except for an extra item (6L), which asks teachers about collaboration between leadership and instructional staff. Administrators do not receive this question, so that item was omitted from the analyses. The questions appear in a single block of the survey, and respondents are given a five point likert-like scale. Their choices range from 'Very High', coded as 1, to 'Very Low', coded 5:

- a) Teachers' understanding of the school's curricular goals
- b) Teachers' degree of success in implementing the school's curriculum
- c) Teachers' expectations for student achievement
- d) Teachers' ability to inspire students
- e) Parental involvement in school activities
- f) Parental commitment to ensure that students are ready to learn
- g) Parental expectations for student achievement
- h) Parental support for student achievement
- i) Students' desire to do well in school
- j) Students' ability to reach school's academic goals
- k) Students' respect for classmates who excel academically

From the list and description provided above, you may guess that I hypothesized a 3 factor structure pertaining to respondents' beliefs about: 1) Teachers at their school; 2) Parents at their school; and 3) Students at their schools. All the questions generally inquire about these 3 groups' attitudes towards academic excellence, as I mentioned before. However, wording is tricky in these 11 items where we see ability used differently regarding students, teachers, and parents.

In either case, I used a screeplot derived from polychoric correlations of the 11 items and, using a cutoff score above 1 on variance, found that 3 components did seem to be a reasonably parsimonious factor structure. The plot shows eigenvalues on components from both the full international dataset, and the United States subset, with similar results.



Exploratory Factor Analysis

The next step was to perform exploratory factor analysis on a split dataset, which I was able to do thanks to the large number of participants in each study. This would allow a second confirmatory step after assessing fit and best possible model/factor structure. For the factor analysis, I included the argument to use Maximum Likelihood Estimation, as well as specifying a polychoric correlation method.

Keeping factor loading results from an oblique rotation in Table 3, we see some evidence

for three latent factors across all the potential samples used. However, there were examples of cross-loading between factors on several items. Most noticeably, items 6C, 6G and 6H. G and H crossloaded between the teacher student factors, and the parent student factor, respectively. 6H, parental support, similarly cross loaded across parent and student factors, as well. All these cross-loadings make some sense, semantically speaking, as the items refer to one group action (“expectation” and “support”) into a student behavior, “achievement”.

Table 3: Factor Loadings on Full Sample, and US, Math Teacher, and Admin Subsamples

Item	Intl Sample			US Sample			UAE Sample			Math Teachers			Admins		
	ML1	ML2	ML3	ML2	ML3	ML1	ML1	ML2	ML3	ML1	ML2	ML3	ML1	ML2	ML3
bg06a		0.884		0.875				0.922			0.875			0.893	
bg06b		0.907		0.924				0.926			0.887			0.956	
bg06c		0.467	0.235	0.771	0.112		0.149	0.600	0.140	0.108	0.441	0.228		0.532	0.210
bg06d	0.126	0.606		0.658	0.107		0.105	0.622	0.180	0.124	0.570	0.108	0.136	0.650	
bg06e	0.894					0.965			0.883	0.879			0.899		
bg06f	0.934					0.872			0.942	0.950			0.923		
bg06g	0.453		0.291		0.507	0.453	0.332		0.317	0.443		0.281	0.498		0.300
bg06h	0.706		0.235		0.459	0.562	0.401		0.575	0.681		0.258	0.728		0.229
bg06i			0.910		0.891		0.968					0.919			0.934
bg06j			0.823	0.140	0.789		0.861					0.794			0.811
bg06k			0.640		0.687		0.760				0.104	0.632		0.117	0.616

Assessing fit for the models, I provide each one’s uniquenesses and fit statistics in Table 4. Depending on which fit statistic you want to use, there are a number of conclusions. First, that there was relatively worse model fit for the entire US sample than the international sample, if we base our decision on RMSEA, TLI, and CFI. Might the united states inconsistency of curricular goals, resources, and policies across states and districts be a larger influence on respondents’ beliefs than differences across the rest of the world? This is hard to interpret. However, BIC is lowest in the national samples, despite RMSEA, TLI, and CFI. I will also assume the massive sample size for the international sample plays a strong role in the inflated χ^2 and BIC statistics.

Regardless, TLI and CFI was lowest in the US sample, while the same indices were quite consistent in the world sample, math teachers, principals, and even the UAE sample. RMSEA never went past a .05 threshold, though I realize this is a subjective cutoff. Model fit for the EFA was, overall, fair but not *great*, and I wonder if TIMSS researchers designed the instrument with latent variable modeling in mind. No research from the organization itself seems to indicate a concern for the reliability of the scales as they pertain to psychological constructs, beyond descriptive use or for use as regression covariates!

I continue on below to explore whether this three factor model is adequate in explaining different respondents’ beliefs about their school using the split half of the full international sample, borrowing strength from numbers and the relative good fit compared country-specific indices.

Table 4: Model Fit and Uniqueness Indices

	Intl Sample	US Sample	UAE Sample	Math Teachers	Admin
Uniquenesses					
bg06a	0.283	0.275	0.229	0.309	0.253
bg06b	0.172	0.133	0.169	0.181	0.120
bg06c	0.516	0.326	0.341	0.551	0.439
bg06d	0.453	0.419	0.310	0.492	0.371
bg06e	0.291	0.114	0.226	0.291	0.295
bg06f	0.115	0.074	0.062	0.119	0.113
bg06g	0.491	0.251	0.552	0.503	0.429
bg06h	0.217	0.135	0.218	0.230	0.191
bg06i	0.181	0.135	0.099	0.178	0.135
bg06j	0.227	0.304	0.155	0.246	0.210
bg06k	0.492	0.441	0.349	0.510	0.490
Fit Indices					
Chi-Squared	970.939	75.349	107.019	335.914	220.138
BIC	2361.768	326.487	331.634	581.882	551.155
RMSEA	0.083	0.133	0.093	0.080	0.090
TLI	0.953	0.907	0.953	0.953	0.951
CFI	0.978	0.958	0.979	0.979	0.978

Confirmatory Factor Analysis

For the Confirmatory Factor Analyses, I chose to stick with the international sample in order to confirm how well the 3 factor model explains respondent's beliefs about their school environment. Additionally, I tested invariance across multiple groups, to address my curiosity over whether the latent variable is accounted for equally between principals, math teachers, and science teachers. Were this the case, a good argument could be made for the potential use of this scale for contextualizing school performance or student achievement measures, regardless of country and inclusive of math, science, and school indicators of the three latent *excellence* factors.

Table 5: Model Fit Indices by Model

	Strict	Strong	Partial
Free parameters	186.000	175.000	197.000
Chi-Squared statistic	3773.193	3773.193	3661.564
Degrees of freedom	111.000	122.000	100.000
p-value	0.000	0.000	0.000
Comparative Fit Index	0.995	0.995	0.995
Tucker-Lewis Index	0.992	0.993	0.991
RMSEA	0.080	0.076	0.083
SRMR	0.045	0.045	0.044

Table 6 indicates how varying constraints on model parameters affects the models' structures. Surprisingly, and I will assume my code had a fatal flaw, restricting slopes, intercepts

and variances across groups did virtually nothing across partial, strong, nor strict invariance models. Table 5 above features the different fit indices used, and across the board, these did not change meaningfully as models became less restrictive. Using a (deprecated) `measurementInvarianceCat` function also showed non-significant differences across invariance limits all the way from configural to strict invariance.

Although I am pretty sure something was coded improperly, were my coding correct these results would indicate that TIMSS researchers may have some reason to more freely make mean comparisons in this specific scale. Specifically, with regard to teacher and principal beliefs regarding teacher, parent, or students' orientation towards "excellence".

Table 6: Fit Indices by Model

	Strict Invariance		Strong Invariance		Partial Invariance	
Principals						
Parent factor mean a	0	-	0	-	0	-
Teacher factor mean a	0	-	0	-	0	-
Student factor mean a	0	-	0	-	0	-
Parent factor variance a	1	-	1	-	1	-
Teacher factor variance a	1	-	1	-	1	-
Student factor variance a	1	-	1	-	1	-
Covariance P:T a	0.709	0.01	0.709	0.01	0.706	0.01
Covariance P:S a	0.848	0.006	0.848	0.008	0.846	0.006
Covariance T:S a	0.72	0.01	0.72	0.01	0.719	0.01
Math Teachers						
Parent factor mean m	-0.015	0.017	-0.015	0.017	0	0.015
Teacher factor mean m	0	0.032	0	0.032	0	0.032
Student factor mean m	0	0.02	0	0.02	0	0.02
Parent factor variance m	0.975	0.009	0.975	0.009	0.702	0.003
Teacher factor variance m	0.869	0.012	0.869	0.012	0.609	0.005
Student factor variance m	0.961	0.012	0.961	0.012	0.656	0.003
Covariance P:T m	0.562	0.012	0.562	0.012	0.399	0.007
Covariance P:S m	0.815	0.01	0.815	0.01	0.571	0.005
Covariance T:S m	0.58	0.013	0.58	0.013	0.401	0.007
Science Teachers						
Parent factor mean s	-0.015	0.015	-0.015	0.015	0	0.013
Teacher factor mean s	0	0.031	0	0.031	0	0.028
Student factor mean s	0	0.018	0	0.018	0	0.016
Parent factor variance s	0.97	0.008	0.97	0.008	0.729	0.002
Teacher factor variance s	0.897	0.012	0.897	0.012	0.647	0.004
Student factor variance s	0.966	0.011	0.966	0.011	0.706	0.003
Covariance P:T s	0.58	0.01	0.58	0.01	0.428	0.006
Covariance P:S s	0.8	0.009	0.8	0.009	0.594	0.004
Covariance T:S s	0.614	0.011	0.614	0.011	0.446	0.006