Comparing Pre-service and In-service Teachers' AI Competence Self-Efficacy

-- Measurement Invariance and Latent Mean Differences

Yanfen Chen & Michaela Gläser-Zikuda
School Education and Instructional Research, University of Erlangen-Nürnberg, Germany

Theoretical and Empirical Background

01 Research Background

- AI has reshaped education
- AI potential in education remains underutilized
- AI Literacy Framework (OECD and EU, 2025), primary, secondary education
- Teaches' roles in K-12 stages, cultivate AI-literate citizens and future labor
- Significance of pre- and in-service teachers' AI competence

02 Theoretical Frameworks

- TPACK, DigCompEdu, Teacher Digital Competence (TDC) framework (Mishra and Koehler, 2006; Redecker, 2017; Falloon, 2020)
- UNESCO: AI Competency Framework for Teachers (Miao & Cukurova, 2024)
- Five dimensions: Human-centered Mindset, Ethics of AI, AI Pedagogy, AI Foundations and Applications, AI for Professional Learning

03 Standardized Instruments

- Based on TPACK model: Intelligent/AI-TPACK scales (e.g. Celik, 2023; Ning et al., 2024, 2025) or based on TDC framework (e.g. Falloon, 2020)
- UNESCO: AI competency framework for Teachers (Miao & Cukurova, 2024)
- An extra dimension, AI assessment from DigCompEdu (Redecker, 2017)
- Teacher AI Competence Self-Efficacy (TAICS) scale (Chiu et al., 2024)

04 Research Gaps

- Research on teachers' AI competence is still in its infancy
- Theoretical frameworks just emerged
- Standardized measurements are scarce
- Chinese policy momentum is accelerating (Ministry of Education, 2025)
- Chinese mainland lacks culturally adapted tools for empirical research

Research Questions

RQ 1: To what extent does the TAICS scale demonstrate **a valid factorial structure** in assessing AI competence self-efficacy among pre- and in-service teachers in mainland China?

RQ 2: **Is measurement invariance** between the groups of pre- and in-service teachers supported?

RQ 3: Are there significant **latent mean differences** across TAICS dimensions between the groups of pre- and in-service teachers?

Method

Participants and Procedure

N = 2590 K-12 Teachers

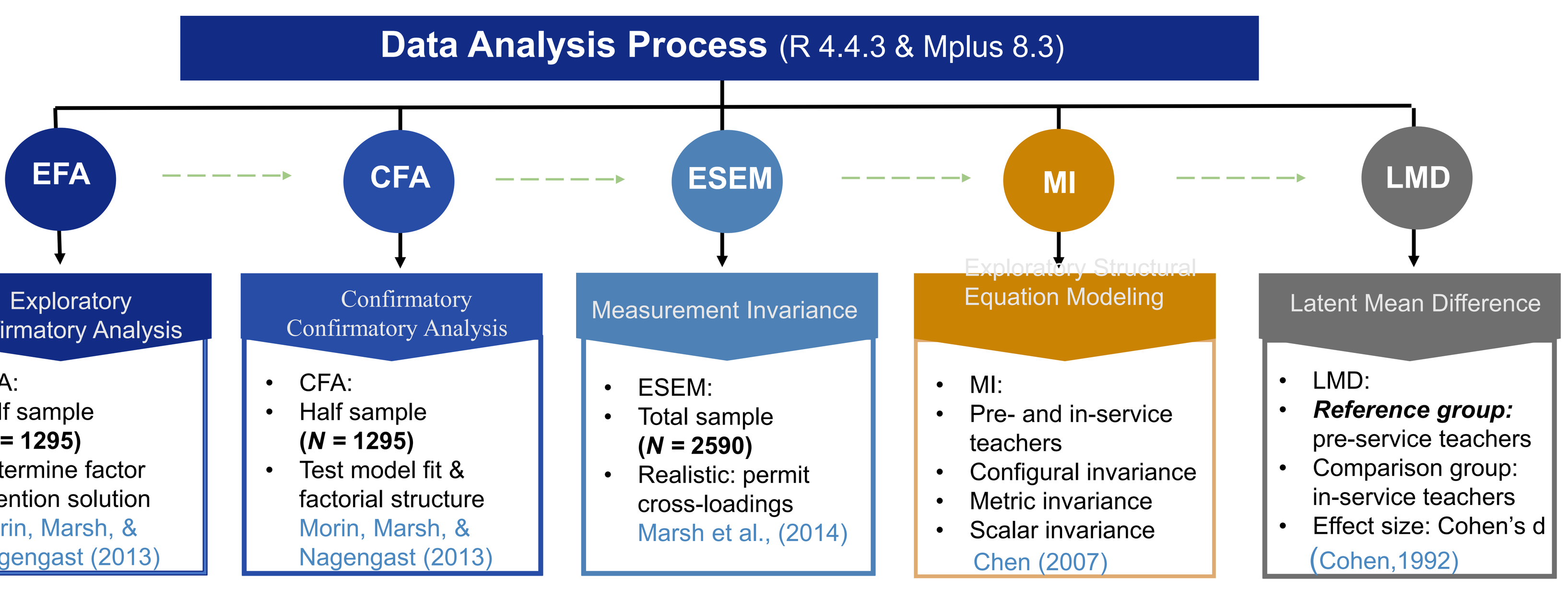
Pre-service teachers (*n* = 770, 29.73%)		In-service teachers (*n* = 1820, 70.27%)	
M = 19.92, *SD* = 1.33	= 81.04%, = 18.96%	*M* = 37.59, *SD* = 9.13	= 75%, = 25%

- Data collection: Wis 2024/25, online survey, convenient sampling
- Completion time: 15 mins

Instrument

- Teacher AI Competence Self-efficacy (TAICS) scale (Chiu et al., 2024)
- Six dimensions, 24 items, five-point Likert scale
- AI Knowledge (AIK) (Ω = 0.930, 95%CI: [0.928, 0.932])
- AI Pedagogy (AIP) (Ω = 0.980, 95%CI: [0.980, 0.981])
- AI Assessment (AIA) (Ω = 0.970), 95CI: [0.970, 0.971])
- AI Ethics (AIE) (Ω = 0.947, 95%CI: [0.946, %0.949])
- Human-Centered Education (HCE) (Ω = 0.946, 95%CI: [0.944, 0.947])
- Professional Engagement (PEN) (Ω = 0.940, 95%CI: [0.939, 0.941])
- McDonald Omega: excellent internal consistency (George & Mallery, 2003)
- Standardized forward–backward translation (Brislin, 1970)

Statistical Analysis



Results

RQ1 Table 1 & Table 2. Standardized Factor Loadings of EFA and ESEM

	AIK	AIP	AIA	AIE	HCE	PEN
AIK_1	0.651					
AIK_2	**0.437**	**0.342**				
AIK_3	0.762					
AIK_4	**0.487**	**0.363**				
AIP_1		0.840				
AIP_2		0.767				
AIP_3		0.714				
AIP_4		0.530				
AIA_1		**0.519**				
AIA_2			0.699			
AIA_3			0.670			
AIA_4			0.664			
AIE_1				0.725		
AIE_2				0.889		
AIE_3				0.801		
AIE_4				0.809		
HCE_1					0.730	
HCE_2					0.897	
HCE_3					0.679	
HCE_4					0.694	
PEN_1						0.417
PEN_2						0.605
PEN_3						0.756
PEN_4						0.714

	AIK	AIP	AIA	AIE	HCE	PEN
AIK_1	**0.660**	0.037	-0.001	0.064	0.111	0.003
AIK_2	0.451	0.338	0.048	0.050	-0.002	0.084
AIK_3	**0.817**	-0.087	0.070	0.015	0.038	0.004
AIK_4	0.579	0.304	-0.014	0.046	0.006	0.078
AIP_1	0.048	**0.779**	0.052	0.060	0.066	0.036
AIP_2	0.056	**0.718**	0.102	0.055	0.067	0.043
AIP_3	0.090	**0.612**	0.150	0.035	0.061	0.103
AIP_4	0.150	**0.430**	0.236	0.037	-0.035	0.223
AIA_1	0.027	**0.405**	**0.399**	0.058	0.111	0.042
AIA_2	0.058	0.040	**0.766**	0.022	0.044	0.059
AIA_3	0.043	0.047	**0.758**	0.049	0.060	0.074
AIA_4	0.044	0.120	**0.721**	0.080	0.067	0.007
AIE_1	0.008	0.061	-0.121	**0.740**	0.080	-0.013
AIE_2	0.017	-0.061	0.038	**0.914**	-0.038	-0.011
AIE_3	0.051	-0.002	-0.010	**0.852**	0.003	0.030
AIE_4	-0.049	0.034	0.086	**0.864**	0.015	0.007
HCE_1	0.040	0.115	-0.068	0.099	**0.731**	0.048
HCE_2	0.011	-0.005	0.059	-0.015	**0.931**	-0.053
HCE_3	-0.012	-0.014	0.060	0.053	**0.588**	0.157
HCE_4	0.121	-0.062	0.097	0.013	**0.603**	0.155
PEN_1	0.101	0.143	-0.087	0.143	0.251	**0.415**
PEN_2	0.037	0.050	-0.101	0.191	0.158	**0.571**
PEN_3	0.012	0.077	0.000	0.018	0.043	**0.815**
PEN_4	0.052	-0.052	0.227	-0.006	-0.012	**0.717**

① EFA (half-sample)

- Supported six-factor
- *AIK_2 & AIK_4*: cross-loadings on *AIP*
- *AIA_1*: misloaded on *AIP*
- Excluded in later CFA

② CFA (half-sample)

- Revised six-factor: good fit (χ^2/df = 4.229, robust CFI = 0.965, robust TLI = 0.958, robust RMSEA = 0.068, SRMR = 0.030)
- Factors: inter-correlated
- Factor loadings: >0.75

③ ESEM (full sample)

- Six-factor: good fit (χ^2/df = 3.684, robust CFI = 0.989, robust TLI = 0.979, robust RMSEA = 0.032, SRMR = 0.008)
- *AIK_2 and AIK_4*: primary loadings on *AIK*; secondary loadings on *AIP*
- *AIA_1*: high cross loadings on *AIP*; removed
- Revised ESEM model: slightly improved fit (χ^2/df = 3.652, robust CFI = 0.990, robust TLI = 0.980, robust RMSEA = 0.032, SRMR = 0.008)

④ Concurrent Validity (full sample)

- Adapted DigCompEdu Scale (Quast et al., 2025)
- Seven factors: *ADMIN, LESS, TEACH, ASSES, EMPOW, COMP, COPY*
- 0.5 < r < 0.7, all factors inter-correlated

RQ1 Table 3. Concurrent Validity with Adapted DigCompEdu Scale

	AIK	AIP	AIA	AIE	HCE	PEN
ADMIN	.564***	.539***	.488***	.539***	.541***	.574***
LESS	.576***	.559***	.492***	.589***	.581***	.600***
TEACH	.563***	.576***	.539***	.579***	.571***	.604***
ASSES	.563***	.620***	.660***	.482***	.542***	.598***
EMPOW	.618***	.638***	.609***	.597***	.602***	.640***
COMP	.612***	.675***	.694***	.556***	.593***	.656***
COPY	.593***	.584***	.548***	.664***	.590***	.576***

➢ **Final model (RQ1)**

- Supported a six-factor, 23-item TAICS model
- excluded AIA_1 for the mainland Chinese context

RQ2: Table 4. Measurement Invariance

Model	χ^2/df	CFI	TLI	RMSEA	SRMR	Model	ΔCFI	ΔRMSEA	Decision
Overall	6.872	0.965	0.959	0.064	0.029	-	-	-	-
Pre-service Teachers	1.778	0.956	0.949	0.053	0.042	-	-	-	-
In-service Teachers	2.971	0.972	0.967	0.065	0.027	-	-	-	-
M1: Configural Invariance	4.007	0.965	0.959	0.064	0.033	-	-	-	-
M2: Metric Invariance	3.970	0.965	0.960	0.063	0.035	M2-M1	0.000	-0.001	**Accept**
M3: Scalar Invariance	4.342	0.960	0.956	0.066	0.037	M3-M2	-0.005	0.003	**Accept**
M4: Strict Invariance	5.564	0.940	0.938	0.079	0.041	M4-M3	-0.020	0.013	**Reject**

⑤ Measurement (RQ2) Invariance

- Configural, metric & scalar invariance were supported

⑥ Latent Mean (RQ3) Differences

- pre-service teachers scored significantly higher than in-service teachers across all dimensions
- most substantial gaps:

AI Pedagogy (d = −0.585)
AI Knowledge (d = −0.517)
AI Assessment (d = −0.478)

RQ2: Table 5. Latent Mean Differences

	ΔM	P	Cohen's d	Size effect
AIK	-0.344	< 0.001	-0.517	**medium**
AIP	-0.462	< 0.001	-0.585	**medium-to-large**
AIA	-0.406	< 0.001	-0.478	**near-medium**
AIE	-0.163	< 0.001	-0.275	small
HCE	-0.213	< 0.001	-0.290	small-to-medium
PEN	-0.191	< 0.001	-0.265	small

Discussion

- Culturally adapted scale: replicated 6-factor-structure, 23 items, removed item AIA_1
- Potential reasons: a) contextual & sampling differences b) model-related assumptions
- Latent mean differences: pre-service scored significantly higher than in-service teachers
- Potential reasons: a) exposure to generative AI tools b) motivational and career-stage differences c) AI-integrated teacher preparation programs

Reference

Scan for Reference!

Contact

Yanfen Chen
yanfen.chen@fau.de
Prof. Dr. Michaela Gläser-Zikuda
michaela.glaeser-zikuda@fau.de

Friedrich-Alexander-Universität
Faculty of Humanities, Social Sciences, and Theology