EAM2025

XI Conference

23RD - 25TH JULY 2025

Spain Tenerife
Canary Islands

European Association of Methodology

Longitudinal diagnostic models for small sample size contexts

Pablo Nájera, Miguel A. Sorrel, Hyunjee Oh, & Chia-Yi Chiu

Universidad de La Laguna
Cabildo de Tenerife
Tagoro
cajasiete
Gobierno de Canarias
Consejería de Universidades, Ciencia e Innovación y Cultura
Agencia Canaria de Investigación, Innovación y Sociedad de la Información
ICI Instituto Canario de Igualdad
tea
hogrefe

Cognitive diagnosis modelling

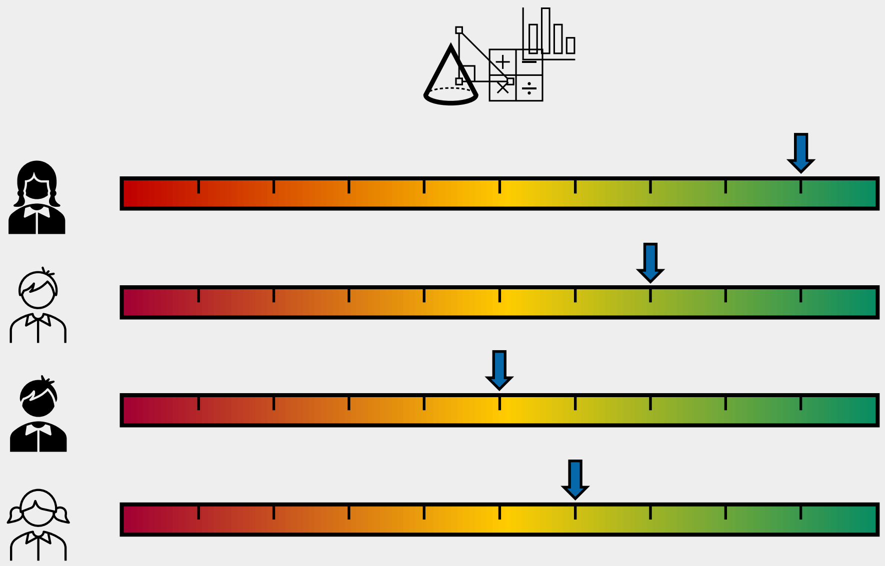
Cognitive diagnosis modelling (**CDM**) is a family of confirmatory statistical models that can be understood either as:

- restricted latent class models, or
- multidimensional item response theory (IRT) models with discrete latent variables

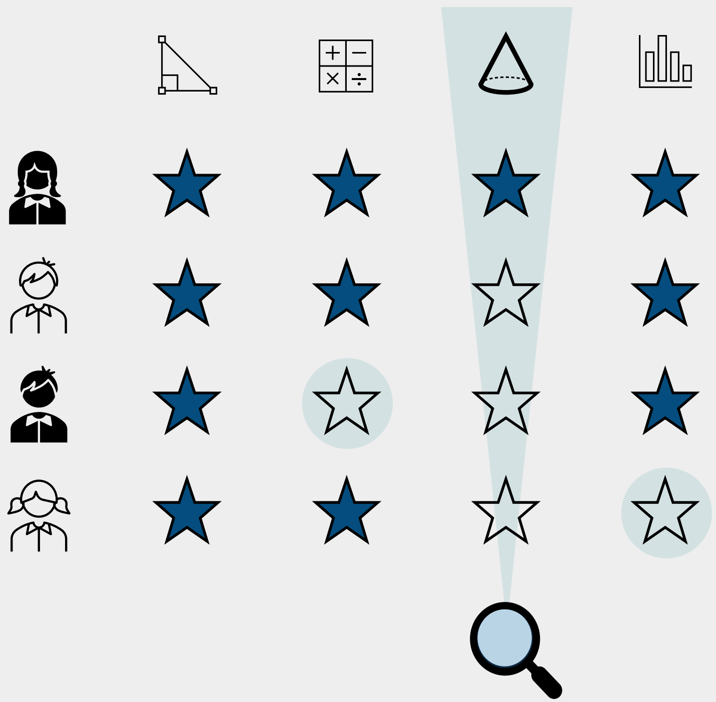
These models have been primarily employed in educational assessments, where the detailed feedback they provide can help identify students' strengths and weaknesses objectively and in a timely manner (de la Torre & Minchen, 2014; Paulsen & Svetina, 2022)

Cognitive diagnosis modelling

Summative assessments



Diagnostic assessments



Longitudinal CDM

In the last decade, several longitudinal CDMs have been developed with the aim of evaluating changes in respondents' attribute profiles over time

The calibrate-and-score (**CS**; Jurich & Bradshaw, 2014) method involves:

1. Fitting a CDM to the pre-test data

   Calibrating items and obtaining attribute profile classifications

2. Fitting a CDM to the post-test data, constraining the item parameters to the pre-test estimates

   Obtaining attribute profile classifications

This approach does not account for the interdependency at pre- and post-test

Estimation is simplified, with only $2^K - 1$ parameters estimated at post-test

Longitudinal CDM

The transition diagnostic classification model (**TDCM**; Madison & Bradshaw, 2018) is a latent transition analysis (LTA) with a CDM as the measurement model

It estimates three groups of parameters concurrently for pre- and post-test data:

- The probabilities of belonging to each latent class at pre-test
- The probabilities of transitioning from one latent class to another across time
- Item response probabilities at pre- and post-test (via CDM)

This model accounts for local dependency between responses

It enables longitudinal invariance assessment

Estimation is challenging and computationally expensive, with a total of $\sum_{j=1}^{J} 2^{K_j} + 2(2^K - 1) + 2^K(2^K - 1)$ parameters estimated concurrently

Model complexity and sample size

CDM is a family of models with different item response functions and complexities

In order to provide accurate parameter estimates, traditional CDMs require:

- Large sample sizes (*N* > 500; Sen & Cohen, 2021)
- Simple Q-matrices, high-quality items, low dimensionality… (Sorrel et al., 2021)

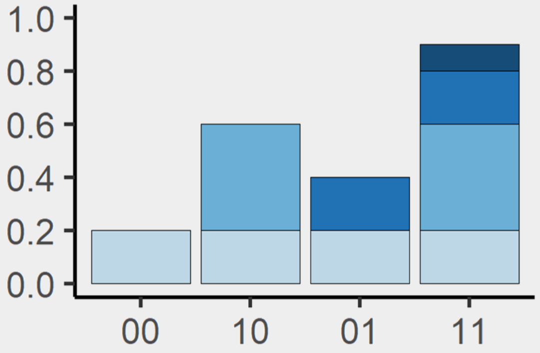
Under suboptimal conditions, a more parsimonious model might provide more accurate parameter estimates and attribute classifications than a more complex one, even when the latter is the generating model (Nájera et al., 2023)

Model complexity and sample size

CDM is a family of models with different item response functions and complexities

G-DINA

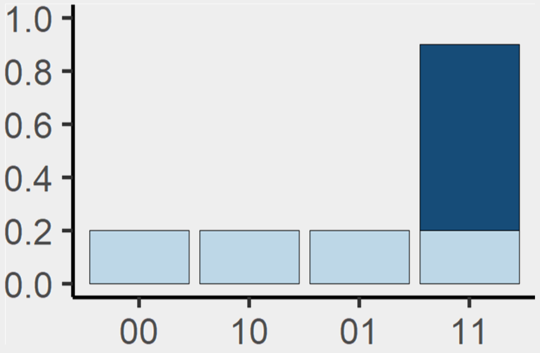
(de la Torre, 2011)



$$(2^K - 1) + \sum_{j=1}^{J} \mathbf{2}^{\boldsymbol{K_j}}$$

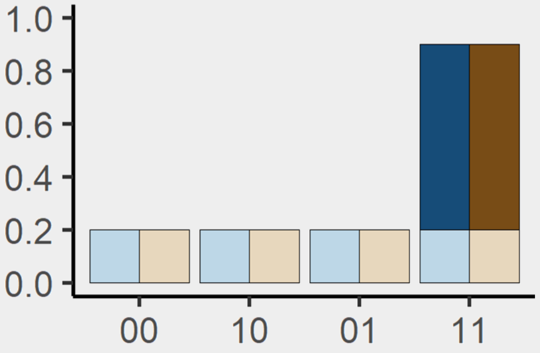
DINA

(Junker & Sijtsma, 2001)



$$(2^K - 1) + \boldsymbol{2J}$$

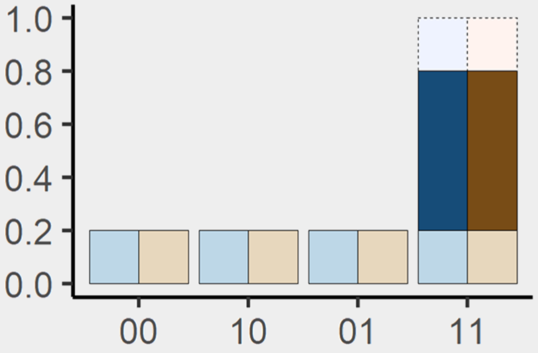
S-DINA



$$(2^K - 1) + \mathbf{2}$$

R-DINA

(Nájera et al., 2023)



$$(2^K - 1) + \mathbf{1}$$

The present study

Up to date, studies about longitudinal CDM have worked with relatively optimal contexts (e.g., simple Q-matrices) and large sample sizes ($N > 1000$)

The main objective of the present study is to elucidate the most optimal way of estimating longitudinal CDM in small sample size scenarios

- Is the simplicity of the **CS** method enough to overcome its limitations and outperform the **TDCM**?
- Can simpler models (e.g., **R-DINA**) provide more accurate parameter estimates and attribute classifications than the **G-DINA** model, even when the G-DINA is the true, generating model?

Simulation study design

Simulation factors

Generating model = **G-DINA**

Sample size = 25, 50, 100

Number of attributes = 3, 5

Q-matrix complexity = simple, complex (.45, .35, .20)

Item discrimination = 0.1, 0.2, 0.3

Probability of attribute mastery growth = 0.20

Probability of attribute mastery loss = 0.10

Item parameter drift = none, random

6 items per attribute, attribute correlations of 0.5

Methods

TDCM CS

×

G-DINA DINA S-DINA R-DINA

Dependent variables

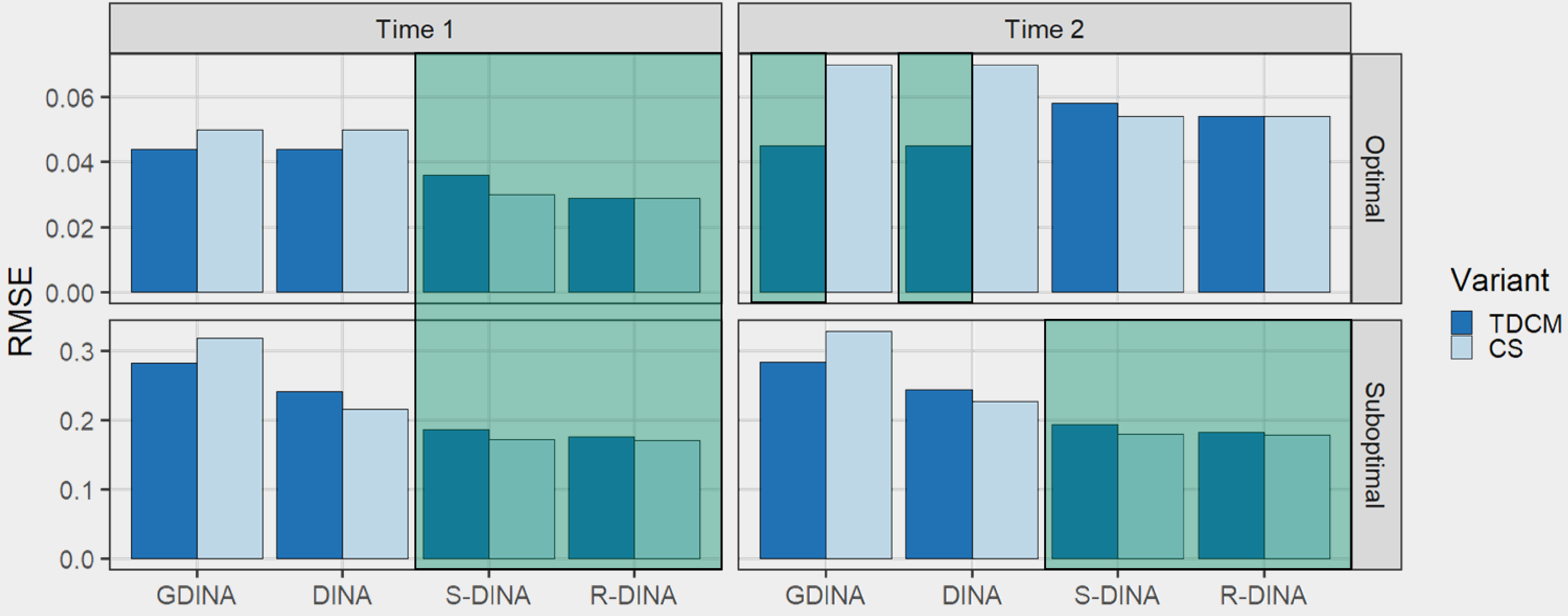
Item parameter recovery

Transition probs. Recovery

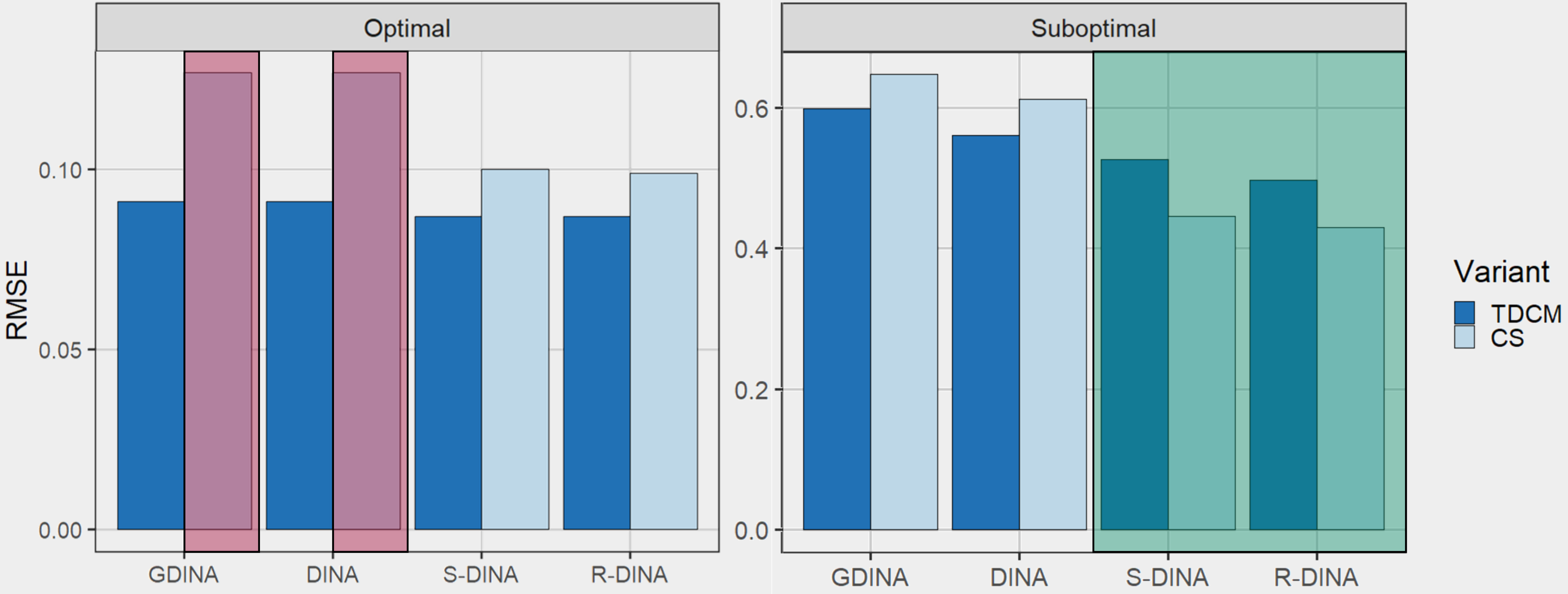
Classification accuracy

Model fit

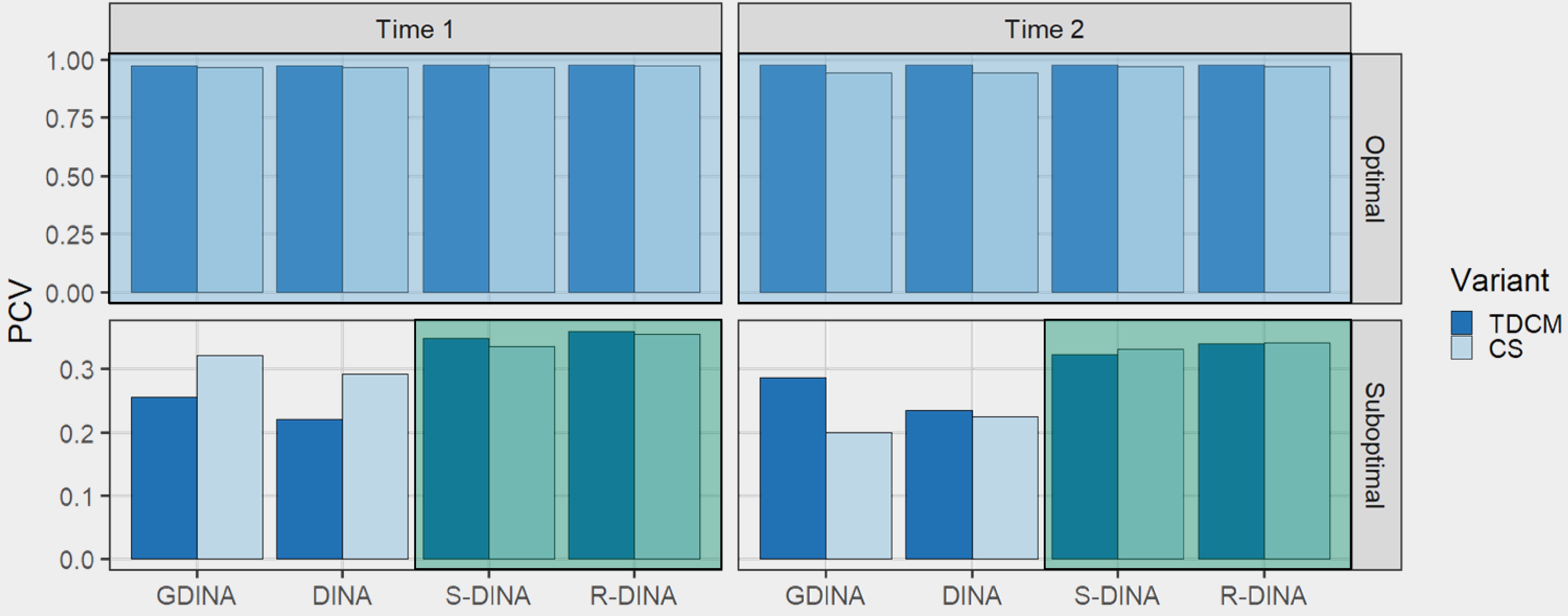
Results (I): Item parameter recovery



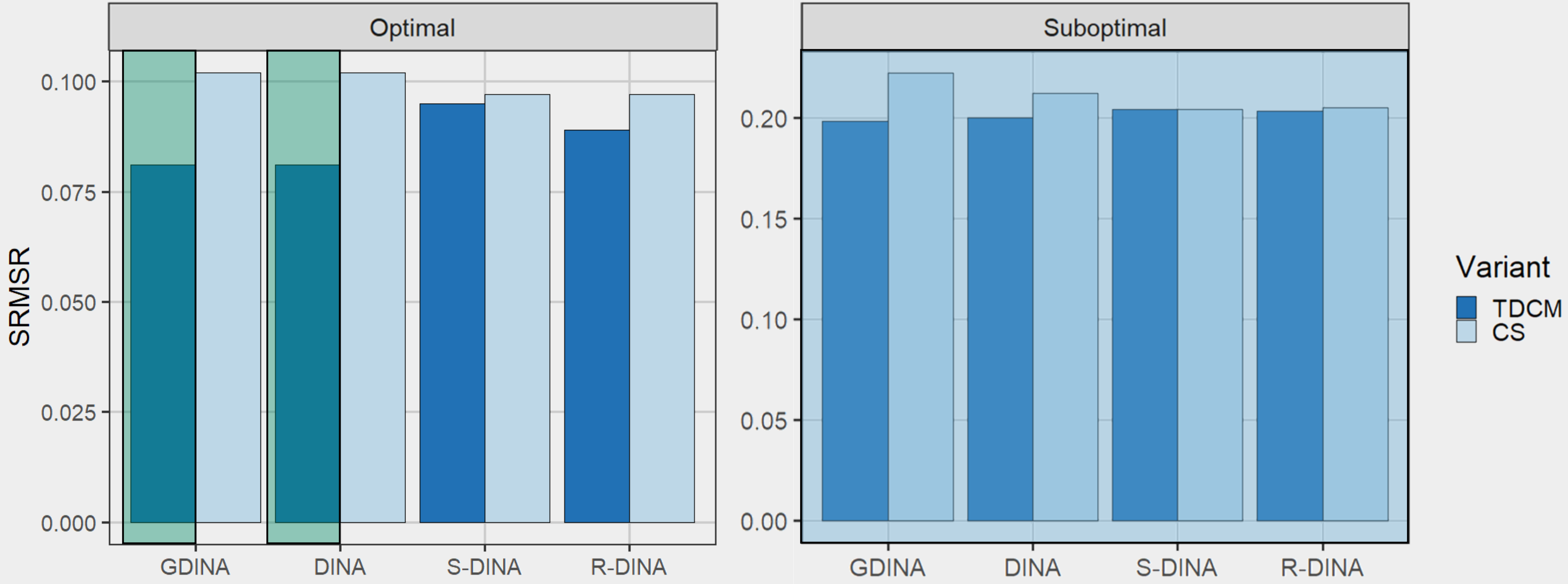
Results (II): Transition probabilities recovery



Results (III): Attribute classification accuracy



Results (IV): Absolute fit



Conclusions

More accurate **parameter estimates** and **attribute profile classifications** are provided by **simple CDMs** (S-DINA and R-DINA), especially under suboptimal conditions

Model fit is similar for **all CDMs**

The **TDCM** provides slightly better results with **complex CDMs** (G-DINA and DINA), especially under optimal conditions

The **CS** method provides slightly better results with **simple CDMs** (S-DINA and R-DINA), especially under suboptimal conditions

Overall…

1. S-DINA + CS
2. R-DINA + CS/TDCM
3. G-DINA + TDCM

Conclusions

Future extensions:

1. Anchor item design (e.g., different items on pre- and post-test)

   CDM classification invariance (Madison & Bradshaw, 2018)

2. Different modelling approaches of latent class transition

   Multilevel CDM (Huang, 2017)

   Hidden Markov models (Wang et al., 2018)

   Bayesian networks (Lee & Gu, 2024)

Conclusions

Research on CDM for small sample sizes

- Model comparison to reduce complexity (Sorrel et al., 2021)
- Reliability estimation via multiple imputation (Kreitchmann et al., 2023)
- Estimation approaches for small samples (Sorrel et al., 2023; Nájera et al., 2023)
- Computerized adaptive testing with no calibration sample (Nájera et al., 2025)

Psychometric Models & Applications Lab.

cdmTools cdcatR cdcatS


Miguel A. Sorrel
UAM, Spain

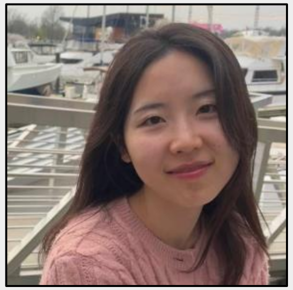

Francisco J. Abad
UAM, Spain


Rodrigo S. Kreitchmann
UNED, Spain


Jimmy de la Torre
HKU, Hong Kong


Scarlett Escudero
UMN, USA


Chia-Yi Chiu
CU, USA


Hyunjee Oh
CU, USA

Longitudinal diagnostic models for small sample size contexts

Pablo Nájera, Miguel A. Sorrel, Hyunjee Oh, & Chia-Yi Chiu

Thank you!

pnajera@comillas.edu

COMILLAS UNIVERSIDAD PONTIFICIA ICAI ICADE CIHS

This work was supported by the 2024 Call for Funding of Research Projects of Universidad Pontificia Comillas under the project "Advancements in Cognitive Diagnosis Models for Formative Assessments"

References

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*(2), 179–199. https://doi.org/10.1007/s11336-011-9207-7

de la Torre, J., & Minchen, N. (2014). Cognitively Diagnostic Assessments and the Cognitive Diagnosis Model Framework. *Psicología Educativa*, *20*(2), 89-97.

Huang, H. Y. (2017). Multilevel cognitive diagnosis models for assessing changes in latent attributes. *Journal of Educational Measurement*, *54(4*), 440-480.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258-272. https://doi.org/10.1177/01466210122032064

Jurich, D. P., & Bradshaw, L. P. (2014). An illustration of diagnostic classification modeling in student learning outcomes assessment. *International Journal of Testing*, *14*, 49–72.

Kreitchmann, R. S., de la Torre, J., Sorrel, M. A., Nájera, P., & Abad, F. J. (2022). Improving reliability estimation in cognitive diagnosis modeling. Behavior Research Methods, *55*, 3446-3460.

Lee, S., & Gu, Y. (2024). Latent conjunctive Bayesian network: Unify attribute hierarchy and Bayesian network for cognitive diagnosis. *The Annals of Applied Statistics*, *18*(3), 1988-2011.

Madison, M. J., & Bradshaw, L. P. (2018). Assessing growth in a diagnostic classification model framework. *Psychometrika*, *83*(4), 963-990.

Nájera, P., Abad, F. J., Chiu, C.-Y., & Sorrel, M. A. (2023). The restricted DINA model: A comprehensive cognitive diagnostic model for classroom-level assessment. *Journal of Educational and Behavioral Statistics*, *48*(6), 719-749.

Nájera, P., Sorrel, M. A., Chiu, C.-Y., & Abad, F. J. (2025). Variable-Length Cognitive Diagnostic Computerized Adaptive Testing in Small-Scale Assessments. [Manuscript submitted for publication].

Paulsen, J., & Svetina, D. (2022). Examining cognitive diagnostic modeling in classroom assessment conditions. *The Journal of Experimental Education*, *90*(4), 916–933.

Sen, S., & Cohen, A. S. (2021). Sample Size Requirements for Applying Diagnostic Classification Models. *Frontiers in Psychology*, *11*, 621251.

Sorrel, M. A., Abad, F. J., & Nájera, P. (2021). Improving Accuracy and Usage by Correctly Selecting: The Effects of Model Selection in Cognitive Diagnosis Computerized Adaptive Testing. *Applied Psychological Measurement*, *45(2*), 112–129.

Sorrel, M. A., Escudero, S., Nájera, P., Kreitchmann, R. S., & Vázquez-Lira, R. (2023). Exploring Approaches for Estimating Parameters in Cognitive Diagnosis Models with Small Sample Sizes. *Psych*, *5*, 336-349.

Wang, S., Yang, Y., Culpepper, S. A., & Douglas, J. (2018) Tracking Skill Acquisition with cognitive diagnosis models: a higher-order, hidden Markov model with covariates. Journal of Educational and Behavioral Statistics, *43*(1), 57–87.