Proceedings of the 90th Annual International Meeting of the Psychometric Society, Minneapolis, United States (2025).

Diagnosing Cross-Cutting Skills from Existing Assessment: A Cognitive Diagnostic Modeling Approach

Kristin Lansing,[*1] Tessa J. P. van Schijndel,[2] Jeffrey Steedle,[1] Jimmy de la Torre,[3] Alina A. von Davier,[4,5] and L. Andries van der Ark[2¶]

[1] Measurement Research and Development, ACT Inc., Iowa City, IA, USA

[2]Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands

[3]The University of Hong Kong, Pok Fu Lam, Hong Kong

[4]Duolingo, Pittsburgh, PA, USA

[5]EdAstra Tech, Boston, MA, USA

*Corresponding author. Email: lansingkristin@gmail.com

Abstract

Twenty-first century skills (21CS)—including critical thinking, creative thinking and innovation, collaborative problem solving, and information and communication technology (ICT) skills—are widely regarded as essential for success in education and the workforce. Despite their recognized importance, these skills are often embedded only implicitly within domain-specific curriculum and assessments, and therefore remain unmeasured, unreported, and unsupported. In this paper, we propose a general psychometric procedure for making implicitly embedded 21CS explicit using existing educational assessments, without introducing additional testing burden. The procedure integrates (a) a theoretically grounded framework of 21CS attributes, (b) systematic item-level skill coding via a Q-matrix, and (c) cognitive diagnostic modeling (CDM) to estimate student mastery profiles. As a proof of concept, we apply the procedure to a set of course assessments from an introductory academic computing course and demonstrate how CDMs can recover meaningful attribute-level information about student performance. We discuss technical considerations—including identifiability, attribute co-occurrence, and model fit—as well as implications for assessment design, reporting, and future research. The proposed approach illustrates how psychometric tools can be leveraged to surface important but otherwise hidden skill information already present in educational data.

Keywords: cognitive diagnostic models, Q-matrix, twenty-first century skills, cross-cutting skills, assessment design

1. Introduction

Over the past two decades, educational systems worldwide have increasingly emphasized the importance of so-called twenty-first century skills (21CS), such as critical thinking, creativity, collaboration, and digital literacy (OECD, 2018; World Economic Forum, 2025). These skills are commonly described as domain-general, transferable capabilities that enable learners to adapt to rapidly changing social, technological, and economic conditions. While policy frameworks and curricular guidelines frequently highlight 21CS as core educational outcomes, their assessment remains a persistent challenge.

A central difficulty lies in the fact that 21CS are rarely assessed directly. Instead, they are typically embedded implicitly within domain-specific tasks—such as writing assignments, project-based learning activities, or computer-based assessments—where successful performance requires the application of both content knowledge and 21CS. As a result, students may demonstrate 21CS in practice without ever receiving explicit feedback on those skills, and educators may lack actionable data to support skill development.

From a psychometric perspective, this situation represents both a challenge and an opportunity. The challenge is that implicit skills are not explicitly modeled in most assessment designs, rendering them invisible to standard scoring and reporting practices. The opportunity is that existing assessments already contain rich, and valuable, information about these skills. In this paper, we argue that cognitive diagnostic modeling (CDM) provides a principled and practical approach to extracting this information.

We propose a general procedure for making implicit 21CS explicit in existing educational assessments. The procedure does not require the development of new test instruments or additional testing time. Instead, it leverages three components: (1) a theoretically grounded framework that decomposes 21CS into interpretable attributes; (2) a Q-matrix that maps assessment items to these attributes; and (3) a CDM that uses item responses to estimate student mastery profiles. We illustrate this procedure using data from an introductory academic computing course and demonstrate that meaningful diagnostic information can be obtained even from assessments not originally designed for this purpose.

2. Background and Motivation

2. 1 Twenty-First Century Skills in Education

Numerous international organizations have articulated frameworks defining the skills needed for success in modern societies (Ananiadou & Claro, 2009; Griffin & Care, 2014; Lansing-Stoeffler & Daley, 2022). While the specific labels and categorizations vary, most frameworks converge on a set of core competencies, including critical thinking, creativity, collaboration, and technology skills. These competencies are cross-cutting: they apply across domains and contexts rather than being tied to a single subject area.

Empirical research has consistently linked these skills to positive educational outcomes. Critical thinking and creative thinking are associated with academic performance (Facione, 1991; Sternberg, 2003); collaborative learning improves both cognitive and social outcomes (Laal & Ghodsi, 2012; Terenzini, et. al., 2001; Tolmie et al., 2010); and ICT skills are indispensable in technology-mediated learning environments (Juhaňák, et al., 2019; Makhanu & Kamper, 2012). Despite this evidence, the assessment of 21CS remains fragmented and inconsistent.

2.2 Challenges in Assessing 21CS

Several approaches have been used to assess 21CS, including self-report questionnaires, observational rubrics, performance tasks, and large-scale standardized assessments (Lai & Viering, 2012). Each approach presents trade-offs. Self-reports are scalable but vulnerable to bias; observational methods are rich but resource-intensive; and standalone performance assessments often require substantial development and scoring effort (Lai & Viering, 2012; Long & Wang, 2022).

A further complication is that many 21CS are context-dependent and manifest through interaction with domain-specific content. Attempting to isolate them in decontextualized tasks may undermine their validity. Consequently, embedding 21CS within authentic tasks is often desirable from a construct validity perspective—but doing so makes psychometric modeling more complex.

2.3 Cognitive Diagnostic Models as a Solution

Cognitive diagnostic models were developed to provide fine-grained information about examinee mastery of multiple latent attributes (von Davier & Lee, 2019). Unlike unidimensional item response theory models, CDMs explicitly link item responses to combinations of skills through a Q-matrix. This makes them well-suited for diagnosing complex skill structures, including those involving cross-cutting competencies.

Traditionally, CDMs have been applied to assessments designed explicitly around cognitive attributes (e.g., mathematics subskills). However, recent work has shown that CDMs can also be used to uncover implicit attributes embedded in assessments (e.g., in clinical or psychological contexts) (De La Torre 2011). This insight motivates the present study: if 21CS attributes can be systematically identified in existing assessment items, CDMs can be used to estimate student mastery without altering the assessment itself.

3. A Procedure for Making Implicit Skills Explicit

We propose a three-step procedure for diagnosing implicitly embedded 21CS from existing assessments:

- Select a theoretical framework of 21CS and attributes
- Identify attribute requirements at the item level (Q-matrix construction)
- Estimate student mastery using an appropriate CDM

Each step is described below.

3. 1 Selecting a Theoretical Framework

Because 21CS are often defined at a high level of abstraction (Scoular & Teo, 2021), a suitable framework must decompose each skill into attributes that are specific enough to be meaningfully linked to assessment tasks. The framework should also be internally coherent and applicable across domains.

In our application, we use the ACT Holistic Framework of Education and Workplace Success (Lansing-Stoeffler & Daley, 2022), focusing on its cross-cutting capabilities: Critical Thinking, Creative Thinking and Innovation, Collaborative Problem Solving, and ICT Skills. Each capability is defined in terms of multiple attributes (e.g., inquiry, synthesis, task effectiveness), providing the granularity needed for diagnostic modeling.

3. 2 Q-Matrix Construction

The Q-matrix is a binary matrix indicating which attributes are required for successful performance on each item (Tatsuoka, 1995). Constructing a Q-matrix for implicit skills requires careful judgment, as items were not originally designed with these attributes in mind.

We recommend using multiple raters, structured training, and calibration sessions to improve reliability. Disagreements should be resolved through discussion, with attention to the operational definitions of attributes. Inter-rater agreement statistics can be used to identify problematic attributes or ambiguous items.

3. 3 Assessing Suitability for CDM Analysis and Estimating Student Mastery

Not all assessments are equally suitable for CDMs. Three diagnostics are particularly important:

- Attribute occurrence: Each attribute should appear in a sufficient number of items.
- Attribute co-occurrence: Excessive overlap between attributes can threaten model identifiability.
- Item complexity: Items requiring too many attributes may strain parameter estimation, especially with small samples.

These diagnostics can inform both modeling decisions and potential assessment redesign.

Once the Q-matrix and item response data are available, a CDM can be fitted to estimate student mastery profiles. Model choice should reflect substantive assumptions (e.g., conjunctive vs. compensatory relationships) and practical constraints such as sample size.

For classroom-level data, restrictive or nonparametric CDMs—such as the R-DINA model (Nájera, Abad et al., 2023)—offer a favorable balance between interpretability and estimation stability.

4. Empirical Application

4.1 Data and Assessment Context

The empirical illustration uses data from six course assessments in an introductory academic computing course within a university preparation program. The assessments comprised 64 items, including individual and group-based tasks, and were scored using rubrics.

Responses from 337 students across 17 classes were available. To facilitate CDM analysis, item scores were dichotomized, with satisfactory performance defined as achieving at least 75% of available points. This cut score reflected the course's existing standard for meeting expectations (i.e., a 'C or better' level of work).

4.2 Q-Matrix and Attribute Diagnostics

Using the ACT framework, two raters coded 12 initial attributes (Figure 1). The percentage of agreement between the non-expert Q-matrices and the expert Q matrix ranged from 67–100% (M = 87%), providing some indication that the coding process could be conducted by non-experts as well as experts. Two attributes fell below 70% agreement between the non-experts and the expert Q-matrix: Analyze & Evaluate (Critical Thinking; 69%) and Transform (ICT; 67%), indicating for example, opportunities to improve training for those attributes, opportunities to improve the clarity of the definitions of those attributes, or ambiguous or unclear item prompts. Based on attribute occurrence diagnostics, three attributes with zero occurrence were removed, yielding a final set of nine attributes. Attribute occurrence, co-occurrence, and item complexity statistics indicated moderate overlap among some attributes but generally acceptable conditions for CDM analysis (Table 1).

	Critical Thinking				Creative Thinking & Innovation				Collaborative Problem Solving		Information & Communication Technology	
Item Number	Inquire	Analyze & Evaluate	Synthesize	Expand	Conventional Ideas	Diverse Ideas	Unconventional Ideas	Evaluate & Improve Ideas	Team Effectiveness	Task Effectiveness	Transform	Share
1	0	0	0	0	1	1	1	1	1	1	1	1
2	0	1	0	0	1	1	1	0	0	1	0	0
3	0	0	0	1	1	0	1	1	0	1	0	0
4	1	0	0	0	0	0	0	0	0	1	0	1
5	0	0	1	0	0	0	0	1	1	0	0	0

Figure 1. Example 21CSs Q-matrix

Figure shows 4 21CSs and their corresponding 12 attributes in columns and item numbers as rows. A 1 indicates that an attribute is required for success with the item, and a 0 indicates that an attribute is not required for success with an item.

Table 1. Attribute occurrence and co-occurrence frequencies (and percentages) in the Q-matrix

		Co-occurrence with other attributes frequencies (and percentages)								
	Occurrence	1	2	3	4	5	6	7	8	9
1. Inquire	22 (34)	(-)	**22 (100)**	**21 (95)**	3 (14)	0 (0)	6 (27)	7 (32)	15 (68)	11 (50)
2. Analyze & Evaluate	50 (80)	22 (43)	(-)	30 (59)	9 (18)	0 (0)	12 (24)	1 (2)	23 (45)	35 (69)
3. Synthesize	30 (47)	21 (70)	**30 (100)**	(-)	8 (27)	0 (0)	7 (23)	8 (27)	17 (57)	17 (57)
4. Expand	9 (14)	3 (33)	**9 (100)**	8 (89)	(-)	0 (0)	0 (0)	1 (11)	4 (44)	8 (89)
5. Unconventional Ideas	2 (3)	0 (0)	0 (0)	0 (0)	0 (0)	(-)	0 (0)	0 (0)	0 (0)	1 (50)
6. Team Effectiveness	16 (25)	6 (38)	12 (75)	7 (44)	0 (0)	0 (0)	(-)	12 (75)	6 (38)	6 (38)
7. Task Effectiveness	13 (20)	7 (54)	9 (59)	8 (62)	1 (8)	0 (0)	**12 (92)**	(-)	5 (38)	4 (31)
8. Acquire	24 (38)	6 (25)	12 (50)	7 (29)	0 (0)	0 (0)	16 (67)	12 (50)	(-)	6 (25)
9. Apply	39 (61)	7 (18)	9 (23)	8 (21)	1 (3)	0 (0)	12 (31)	13 (33)	5 (13)	(-)

Note: Table shows attribute occurrence by frequency (and percentage) of items in which the attribute was indicated as required. Table also shows the co-occurrence of attributes with other attributes by count (and percentage). Instances with co-occurrence >90% are indicated with bold. Asymmetry of attribute percentages is due to the change in the occurrence attribute as the denominator

4.3 Assessing Suitability for CDM Analysis and Estimating Student Mastery

Given the conjunctive nature of many tasks and the moderate sample size, we fitted the Restricted Deterministic Input, Noisy "And" gate (R-DINA) model (Nájera, Abad et al., 2023). R-DINA is a more parsimonious, parametric variation of the standard DINA model, specifically designed for use in small-scale or classroom-level assessments. Model fit, classification accuracy, and identifiability were evaluated using established diagnostics.

Test-level classification accuracy was approximately 71%, with attribute-level accuracy exceeding 78% for all attributes. Some attribute profiles were partially unidentifiable, primarily due to high co-occurrence among certain attributes.

Aggregate mastery estimates were high for most attributes, suggesting either strong student performance or limited discrimination in the assessments. This pattern is consistent with what is known about students enrolled in the Global Assessment Certificate (GAC) program, who are typically high-performing, university-bound students selected into an academically rigorous curriculum. Consequently, relatively high levels of performance on course assessments were not unexpected. Individual mastery profiles, however, revealed observable variation across students and attributes. This variation should be interpreted cautiously given that some profiles are less precisely estimated than others due to design constraints (e.g., high attribute co-occurrence and a limited number of items per attribute). Individual mastery profiles illustrate how our proposed procedure could support targeted feedback and instructional decision-making.

The CDM provided insights regarding aggregate attribute mastery and attribute level classification accuracy. Aggregate attribute mastery ranged from 70.2% to 99.7% (Table 2). These relatively high percentages of attribute mastery either indicate that the exams discriminated poorly on most attributes, or that almost all students mastered 21CSs as required by the course assessment. Attribute level classification accuracy is also included in Table 2 as an aggregate indication of the accuracy of attribute mastery classifications. Attribute level classification accuracies ranged from 78.8% to 99.9%.

Table 2. Estimated Percentage of Aggregate Attribute Mastery and Attribute Level Classification Accuracy by Attribute

	Inq	A&E	Syn	Exp	UnI	TeE	TaE	Acq	App
EPAAM	93.8	99.6	95.5	85.5	70.2	98.7	93.2	95.9	99.7
ALCA	96.8	99.9	98.8	94.0	78.8	98.8	97.5	98.4	99.6

Note: EPAAM = Estimated percentage of aggregate attribute mastery; ALCA = Attribute level classification accuracy; Inq = Inquire; A&E = Analyze and Evaluate; Syn = Synthesize; Exp = Expand, UnI = Unconventional Ideas; TeE = Team effectiveness; TaE = Task effectiveness; Acq = Acquire; App = Apply

5. Discussion

The results demonstrate that implicit 21CS embedded in existing assessments can be made explicit using a principled psychometric procedure. From a measurement perspective, the approach offers a way to extend the interpretive value of assessments beyond their original design intent.

Several technical considerations warrant attention. High attribute co-occurrence can compromise identifiability, highlighting the importance of deliberate assessment design if diagnostic reporting is an intended use. Formal sensitivity analysis was not conducted for this study. Dichotomization of polytomous scores simplifies modeling but entails information loss; future work may explore polytomous CDMs or Bayesian approaches.

From a practical standpoint, the procedure supports more transparent reporting of cross-cutting skills and aligns with calls for assessment systems that integrate learning and measurement. Importantly, it does so without increasing testing burden.

6. Conclusion

This paper presents a general, replicable procedure for diagnosing 21CS that are implicitly embedded in existing educational assessments. By combining theoretical skill frameworks, Q-matrix methodology, and cognitive diagnostic models, the approach enables psychometricians and educators to surface valuable information already present in assessment data.

Future research should explore applications across domains, investigate automated or AI-assisted Q-matrix construction, and examine how diagnostic feedback on 21CS influences teaching and learning. As educational systems increasingly emphasize transferable skills, psychometric methods that make the implicit explicit will become ever more important.

References

ACT (2019). Global Assessment Certificate (GAC) Overview. ACT. https://www.act.org/content/dam/act/unsecured/documents/global/GAC-Overview.pdf

Ananiadou, K., & Claro, M. (2009). 21st century skills and competences for new millennium learners in OECD countries. (OECD Education Working Papers, No. 41.) OECD Publishing. https://dx.doi.org/10.1787/218525261154

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

Facione, P. A. (1991). Using the California critical thinking skills test in research, evaluation, and assessment. California Academies Press. https://files.eric.ed.gov/fulltext/ED337498.pdf

Griffin, P., & Care, E. (2014). The ATC21S method. In P. Griffin & E. Care (Eds.), Assessment and teaching of 21st Century Skills (pp. 3–33). Springer.

Juhaňák, L., Zounek, J., Záleská, K., Bárta, O., & Vlčková, K. (2018). The relationship between students' ICT use and their school performance: Evidence from PISA 2015 in the Czech Republic. Orbis Scholae, 12(2), 37–64. https://doi.org/10.14712/23363177.2018.292

Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. Procedia-Social and Behavioral Sciences, 31, 486–490. https://doi.org/10.1016/j.sbspro.2011.12.091

Lai, E. R., & Viering, M. (2012). Assessing 21st Century skills: Integrating research findings. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Vancouver, BC.

Lansing-Stoeffler, K., & Daley, N. (2022). Cross-Cutting Capabilities: Transferable Skills for the 21st Century. ACT Research Report. 2022-07. ACT, Inc.

Long, H., & Wang, J. 2022. Dissecting reliability and validity evidence of subjective creativity assessment: A literature review. Educational Psychology Review, 34(3), 1399–1443. https://doi.org/10.1007/s10648-022-09679-0

Makhanu, E., & Kamper, G. (2012). The relationship between principals' access to information and communication technology (ICT) and school performance in Kenya. Education and General Studies, 1(1), 38-47. http://kerd.ku.ac.ke/123456789/1494

OECD (2018). The future of education and skills: Education 2030. OECD Publishing. https://www.oecd.org/education/2030/E2030%20Position%20Paper%20(05.04.2018).pdf

Scoular, C. & Teo, I. (2021). Developing strategic plans for an aligned approach to 21st century skills integration. Australian Council for Educational Research. https://doi.org/10.37517/978-1-74286-626-0

Sternberg, R. J. (2003). Creative thinking in the classroom. Scandinavian Journal of Educational Research, 47(3), 325–338. https://doi.org/10.1080/00313830308595

Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern and recognition approach. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.). Cognitively diagnostic assessment (pp. 327–359). Erlbaum

Terenzini, P. T., Cabrera, A. F., Colbeck, C. L., Parente, J. M., & Bjorklund, S. A. (2001). Collaborative learning vs. lecture/discussion: Students' reported learning gains. Journal of Engineering Education, 90(1), 123-130. https://doi.org/10.1002/j.2168-9830.2001.tb00579.x

Tolmie, A. K., Topping, K. J., Christie, D., Donaldson, C., Howe, C., Jessiman, E., ... & Thurston, A. (2010). Social effects of collaborative learning in primary schools. Learning and instruction, 20(3), 177–191. https://doi.org/10.1016/j.learninstruc.2009.01.005

von Davier, M., & Lee, Y. S. (Eds.) (2019). Handbook of diagnostic classification models. Springer

World Economic Forum. 2025. New Economy Skills: Unlocking the Human Advantage. Centre for the New Economy and Society, World Economic Forum. https://www.weforum.org/publications/new-economy-skills-unlocking-the-human-advantage/