EDUCATIONAL AND CLINICAL ASSESSMENT WITHIN COMPETENCE-BASED KNOWLEDGE STRUCTURE THEORY

PASQUALE ANSELMI
UNIVERSITY OF PADOVA, ITALY

JÜRGEN HELLER
UNIVERSITY OF TÜBINGEN, GERMANY

LUCA STEFANUTTI
EGIDIO ROBUSTO
UNIVERSITY OF PADOVA, ITALY

Competence-based knowledge structure theory (CbKST) allows for uncovering the attributes (skills, abilities, symptoms of a disorder) that an individual possesses based on their responses to each of the test items. As such, it represents a sharp departure from classical test theory (CTT) and item response theory (IRT), which aggregate information across items by forming a test score. Although relevant applications of CbKST can be found, CTT and IRT are currently much better known and utilized than CbKST, likely because the latter is more recent. This work aims to contribute to the dissemination of CbKST by illustrating the main characteristics of an assessment conducted within this framework. Two applications are presented that differ in terms of the area of assessment (educational or clinical), the type of test used (conjunctive or disjunctive), and the informativeness of the test regarding the possession of attributes (fully informative or nonfully informative). It is demonstrated how, within CbKST, it is possible to evaluate the confidence with which the attributes possessed by an individual are uncovered and to develop personalized interventions.

Keywords: Competence-based knowledge structure theory; Educational assessment; Clinical assessment; Skills; Symptoms of a disorder.

Correspondence concerning this article should be addressed to Pasquale Anselmi, Department of Philosophy, Sociology, Education and Applied Psychology (FISPPA), University of Padova, Via Venezia 14, 35131 Padova (PD), Italy. Email: pasquale.anselmi@unipd.it

Classical test theory (CTT; Algina & Penfield, 2009; Allen & Yen, 2001; Novick, 1966) and item response theory (IRT; Embretson & Reise, 2020; Hambleton et al., 1991; Lord, 1980) aggregate information across items by forming a test score that reflects an individual's level on the variable under investigation. An assessment conducted within these frameworks might identify a student's level of knowledge in algebra or a patient's level of depression. Competence-based knowledge structure theory (CbKST; Doignon, 1994; Falmagne et al., 1990; Korossy, 1999) represents a sharp departure from these approaches, as it aims to determine whether the individual possesses certain attributes (skills, abilities, symptoms of a disorder) based on their responses to each of the test items. An assessment carried out within CbKST might identify the algebraic concepts mastered by the student or the symptoms of depression exhibited by the patient.

CTT and IRT are well known and widely applied in several areas of assessment (Bacherini et al., 2024; Butcher et al., 1989; Colledani et al., 2018, 2019a, 2019b, 2024; Hubin, 1988; Tassé et al., 2016). In contrast, CbKST is less known and utilized than the other two, likely because it was introduced more recently (the first appearance was in Falmagne et al. in 1990). However, relevant applications of CbKST can be found. Tests have been developed to assess different domains, including basic statistics (de Chiusole et al., 2018), mathematization (Anselmi, Heller, Stefanutti, Robusto, & Barillari, 2024), inductive reasoning (Albert & Held, 1999), fluid intelligence (de Chiusole et al., 2024), and depression (Serra et al., 2015). One notable software implementation in the field of educational assessment is Stat-Knowlab (statistic-knowledge laboratory; de Chiusole et al., 2020), a noncommercial intelligent tutoring system for assessing and learning basic statistics. Developed at the University of Padua, Stat-Knowlab has been used by over 20,000 students in basic statistics courses and has proven useful for monitoring their learning processes. Thanks to an EU grant, Stat-Knowlab has been translated into English and expanded with new courses for international use (http://qhelp-knowlab.eu/). Another software implementation is PsycAssist (psychological assistant; de Chiusole et al., 2024, https://psycassist.fisppa.unipd.it/), a web-based artificial intelligence system designed for neuropsychological assessment and training. Developed as part of an Italian project funded by the Ministry of Research and University, PsycAssist assesses executive functions and fluid intelligence in both Italian and English. Researchers can rely on these and other available tests and software for further applications of CbKST.

This work aims to contribute to the dissemination of knowledge about CbKST by illustrating the main characteristics of an assessment conducted within this framework. To provide a broad picture, two applications are presented that differ in terms of the area of assessment (educational or clinical), the type of test used (conjunctive or disjunctive), and the informativeness of the test regarding the possession of attributes (fully informative or nonfully informative). The application in educational assessment is based on the responses of primary school students to a test assessing the mastery of arithmetic skills. The application in clinical assessment is based on the responses to a test that assesses the presence of social phobia symptoms. Both conjunctive and disjunctive tests are considered. In conjunctive tests, it is necessary to possess all the attributes associated with a certain item to solve it. In disjunctive tests, it is sufficient to possess any one of the attributes associated with an item to solve it. Conjunctive tests are deemed more suitable in educational assessments, where the solution of an item is usually broken into a series of steps that must all be carried out correctly. For example, a student must correctly determine both the probability of the complement of an event and the total probability to solve the probability theory item "Given two events A and B in a sample space S, the following probabilities are known: $P(A \cap B) = .86$; $P(A \cap B') = .02$. Find $P(A')$" (X' denotes the complement of event X; Robusto et al., 2010). Disjunctive tests are considered more suitable in clinical assessments, where a disorder may be diagnosed based on the presence of only one of its symptoms (Roussos et al., 2007; Templin & Henson, 2006). For example, two of the 10 diagnostic criteria for pathological gambling from the Diagnostic and Statistical Manual of Mental Disorders IV Text Revision (American Psychiatric Association, 2000) are "Having committed illegal acts such as forgery, fraud, theft, or embezzlement to finance gambling" (Criterion 8) and "Relying on others to provide money to relieve a desperate financial situation caused by gambling" (Criterion 10). It is sufficient for an individual to meet one of these two criteria to respond affirmatively to the item "Gambling has hurt my financial situation" (Templin & Henson, 2006). A fully informative test allows for uniquely uncovering the attributes an individual possesses based on their responses to test items, whereas a nonfully informative test does not guarantee that these attributes can be uniquely identified. As shown below, the arithmetic test used in the educational assessment application is conjunctive and nonfully informative, while the social phobia test used in the clinical assessment application is disjunctive and fully informative.

This work is organized as follows. First, the basic concepts and characteristics of an assessment within CbKST are illustrated. Next, the two applications in educational and clinical assessment are presented. The potential for planning personalized interventions, which derives from considering the attributes an individual possesses rather than relying solely on their test score (here defined as sum score), is highlighted. Finally, the argument regarding CbKST assessment is extended in several directions. In this work, terminology often refers to an assessment in education. Thus, the attributes assessed are called skills, and the responses to the items are denoted as correct or incorrect. However, in the application regarding social phobia, the attributes are referred to as symptoms, and the responses to the items are denoted as affirmative or negative.

Assessment within Competence-based Knowledge Structure Theory

CbKST is a mathematical theory for assessing the skills an individual masters based on their responses to test items. It extends knowledge structure theory (KST; Doignon & Falmagne, 1985), with the main difference being that KST focuses on the items an individual is capable of solving, whereas CbKST emphasizes the skills that the individual masters, which enable them to solve those items. The collection of all skills mastered by an individual is referred to as the individual's *competence state*. The collection of all competence states existing within a population of individuals constitutes a *competence structure* (Korossy, 1999). An assessment conducted within CbKST aims to identify an individual's competence state based on their responses to test items. This section illustrates the main characteristics of a test based on CbKST, the validation of the competence structure and the test on empirical data, and the identification of an individual's competence state from item responses. Two useful software packages for applying the following methods and analyses are the KST toolbox in MATLAB (Brancaccio et al., 2024) and the *pks* package in R (Heller & Wickelmaier, 2013; Wickelmaier et al., 2024; see also Brancaccio et al., 2024). Other packages in R can be found in Hockemeyer (2024).

Tests Based on CbKST

In the tests developed within CbKST, each item is associated with the set of all skills relevant to its solution (Doignon, 1994; Heller et al., 2013). In conjunctive tests, it is necessary to master all the skills associated with an item to solve it. This means that an individual is capable of solving an item if their competence state includes all the skills associated with that item. In disjunctive tests, it is sufficient to master any one of the skills associated with an item to solve it. This means that an individual is capable of solving an item if their competence state includes at least one of the skills associated with that item. The collection of all items that an individual is capable of solving based on their competence state is referred to as the *knowledge state* delineated by the competence state. The collection of the knowledge states delineated by all competence states constitutes a *knowledge structure*.

As an example, consider the three skills a, b, and c, and assume that skill a is a prerequisite for skill b. This means that b cannot be mastered without also mastering a. Figure 1A shows the competence structure defined on the three skills, which contains six competence states. For instance, Ø, $\{a\}$, and $\{a, b\}$ are the competence states of individuals mastering no skills, skill a only, and both skills a and b, respectively. In the figure, any competence state is connected by a path of upward-directed line segments to competence states containing it. The six competence states in the structure are all possible combinations of the three skills that are consistent with a being a prerequisite for b (i.e., where mastery of b implies mastery of a). Consider also a test consisting of five items, which are associated with the three skills as shown in Table 1. In the example, Items 1 and 2 are associated with one skill, while Items 3, 4, and 5 are associated with two skills. If the test is conjunctive, it is necessary to master all the skills associated with a particular item to solve it. For instance, mastering both skills a and c is required to solve Item 4. Individuals with competence state $\{a, b\}$ are capable of solving Items 1 and 3 because each of these items requires a skill included in the competence state, whereas they are not capable of solving Items 2, 4, and 5 because these items require skill c, which does not belong to the competence state. If the test is disjunctive, it suffices to master any of the skills associated with an item to solve it. For instance, mastering either skill a or c is sufficient to solve Item 4. Individuals with competence state $\{a, b\}$ are capable of solving Items 1, 3, 4, and 5 because each of these four items is associated with at least one skill that belongs to the competence state, whereas they are not capable of solving Item 2 because this item requires skill c, which does not belong to the competence state. Table 2 shows the knowledge state delineated by each competence state if the test is conjunctive or disjunctive. Note that individuals mastering no skills (i.e., whose competence state is Ø) are capable of solving no item (i.e., their knowledge state is Ø), while individuals mastering all three skills (i.e., whose competence state is $\{a, b, c\}$) are capable of solving all five items (i.e., their knowledge state is $\{1, 2, 3, 4, 5\}$). Furthermore, an individual mastering all the skills of another individual plus additional skills is capable of solving all items that the first individual is capable of solving and some items more. For instance, suppose the test is conjunctive. An individual with skill a is capable of solving only Item 1, whereas an individual with both skills a and b is capable of solving not only Item 1 but also Item 3. Figures 1B and 1C illustrate the knowledge structures delineated by the competence structure in Figure 1A if the test is conjunctive or disjunctive, respectively. In Figures 1B and 1C, any knowledge state is connected by a path of upward-directed line segments to knowledge states containing it.

TABLE 1
Example of test and associated skills

Item	Associated skills
1	$\{a\}$
2	$\{c\}$
3	$\{a, b\}$
4	$\{a, c\}$
5	$\{b, c\}$

TABLE 2
Knowledge state delineated by each competence state if the test is conjunctive or disjunctive

Competence state	Delineated knowledge state if the test is conjunctive	Delineated knowledge state if the test is disjunctive
Ø	Ø	Ø
$\{a\}$	$\{1\}$	$\{1, 3, 4\}$
$\{c\}$	$\{2\}$	$\{2, 4, 5\}$
$\{a, b\}$	$\{1, 3\}$	$\{1, 3, 4, 5\}$
$\{a, c\}$	$\{1, 2, 4\}$	$\{1, 2, 3, 4, 5\}$
$\{a, b, c\}$	$\{1, 2, 3, 4, 5\}$	$\{1, 2, 3, 4, 5\}$

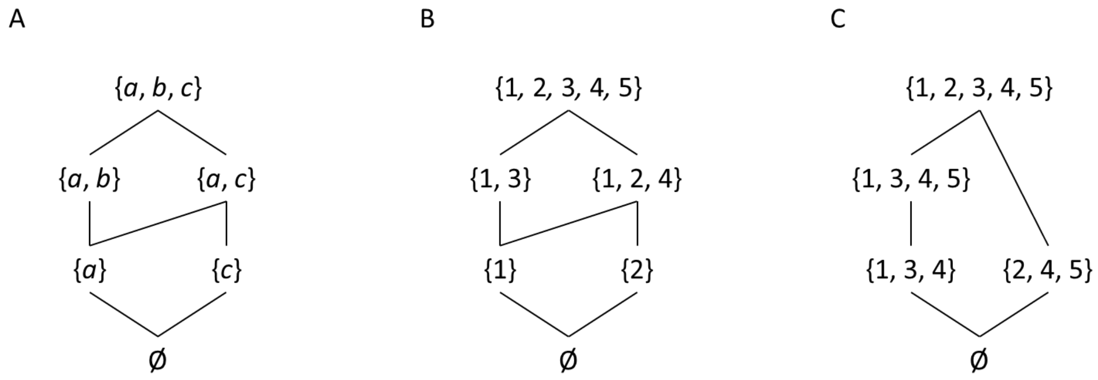

FIGURE 1
Competence structure on the three skills considered in the example (Panel A) and corresponding knowledge structures if the test is conjunctive (Panel B) or disjunctive (Panel C)

Validation of the Competence Structure and the Test

The competence structure and the test are theoretical models of the prerequisite relationship between skills and the association between items and skills, respectively. As such, they need to be empirically validated against individuals' responses to the test items. Within CbKST, a probabilistic model suitable for this purpose is a competence-based extension of the basic local independence model (BLIM; Falmagne & Doignon, 1988a, 1988b), referred to *as the competence-based local independence model* (CBLIM; Heller et al., 2015). The CBLIM takes into account that the competence states and the delineated knowledge states do not occur with the same frequency in the population of interest and that there may not be a perfect correspondence between the items an individual is capable of solving based on the skills mastered (i.e., the knowledge state delineated by the competence state) and the items the individual actually solves (i.e., the response pattern). For example, an individual may fail to solve an item they are capable of due to inattention, or they may solve an item they are not capable of by guessing the correct answer (e.g., when responding to a multiple-choice item). These two conditions are referred to as *careless error* and *lucky guess*, respectively. The CBLIM estimates a probability distribution over the knowledge states, from which a probability distribution over the underlying competence states is derived, as well as a careless error and a lucky guess probability for each item. It is important to note that the terms "careless error" and "lucky guess" are appropriate in educational assessments where responses are coded as correct or incorrect. In clinical assessments, a "no" response from a person who should have answered "yes" is termed a *false negative*, whereas a "yes" response from person who should have answered "no" is termed a *false positive*.

Methods for estimating CBLIM parameters from both complete (Heller & Wickelmaier, 2013; Stefanutti & Robusto, 2009) and missing data (Anselmi et al., 2016; de Chiusole et al., 2015a) are available in the literature, as well as methods for testing model assumptions (de Chiusole et al., 2013, 2015b) and local identifiability (i.e., the possibility of obtaining unique estimates for model parameters; Heller, 2017; Heller et al., 2015; Spoto et al., 2013; Stefanutti et al., 2012). The absolute goodness-of-fit of the CBLIM based on a specific competence structure and test can be assessed using Pearson's chi-square statistic (Heller et al., 2024). A p-value greater than .10 indicates satisfactory goodness-of-fit. The estimates of the careless error and lucky guess probabilities of the items are particularly informative about the wording of the items and the skills required to solve them. Specifically, a large careless error probability for an item might suggest that it is poorly worded or confusing, or that it requires additional skills to be solved. Conversely, a large lucky guess probability for an item might indicate that the item somehow prompts the correct response (as may occur, e.g., in the case of a multiple-choice item with implausible distractors), or that some of the skills assumed to be necessary for solving it are indeed superfluous (Anselmi et al., 2012, 2013; Rupp & Templin, 2008). Finally, if the sum of the careless error and lucky guess probabilities of a certain item exceeds 1, then this item provides invalid information for uncovering an individual's competence state based on their response to it (i.e., a correct response to the item suggests that the individual lacks some skills needed to solve it, while an incorrect response suggests that the individual masters all skills needed to solve it; Heller & Repitsch, 2012).

Identification of the Competence State from the Responses to Test Items

Within CbKST, the assessment of an individual's competence state from their response pattern is a two-step process. The first step is to infer the individual's knowledge state from the response pattern. Based on the individual's response pattern and the knowledge state, careless error and lucky guess probabilities estimated from empirical data using the CBLIM, a likelihood is calculated for each knowledge state, expressing the plausibility of it being the individual's knowledge state. The knowledge state with the highest likelihood is regarded as the individual's *uncovered knowledge state*. The second step is to derive the competence state from the knowledge state, and it is based on the skills associated with each test item. The competence state delineating the uncovered knowledge state is the individual's *uncovered competence state*.

Two elements influence the accuracy of the assessment. The first element is the likelihood of the individual's uncovered knowledge state. The higher the value of this likelihood, the greater the confidence that the uncovered knowledge state corresponds to the individual's actual knowledge state. The second element is the possibility of uniquely identifying the competence state underlying the uncovered knowledge state. If distinct competence states delineate distinct knowledge states, then the competence state of any individual can be uniquely uncovered from their knowledge state. Conversely, if distinct competence states delineate the same knowledge state, there is no guarantee that the individual's competence state can be uniquely uncovered. A test is said to be *fully informative* in the first case and *nonfully informative* in the second case (Anselmi et al., 2022).

Consider again the example of the five-item test measuring three skills. If the test is conjunctive, then distinct competence states delineate distinct knowledge states (Table 2, Figure 1B). Assuming that {1, 2, 4} is an individual's uncovered knowledge state, then $\{a, c\}$ would be that individual's uncovered competence state. If the test is disjunctive, then competence states Ø, $\{a\}$, and $\{c\}$ delineate distinct knowledge states (Ø, {1, 3, 4}, and {2, 4, 5}, respectively), whereas competence states $\{a, c\}$ and $\{a, b, c\}$ delineate the same knowledge state {1, 2, 3, 4, 5} (Table 2, Figure 1C). If {1, 3, 4} is an individual's uncovered knowledge state, then $\{a\}$ is their uncovered competence state. If {1, 2, 3, 4, 5} is an individual's uncovered knowledge state, then their competence state could be either $\{a, c\}$ or $\{a, b, c\}$, indicating that while the individual certainly masters skills a and c, their mastery of skill b is uncertain. In this case, if $\{a, c\}$ is taken to be the individual's uncovered competence state, we can confidently identify which skills they master for sure, but we may underestimate their mastery. A high likelihood for the uncovered knowledge state, along with the presence of a unique competence state that delineates it, leads to a high likelihood that the uncovered competence state is the individual's actual competence state.

APPLICATION IN EDUCATIONAL ASSESSMENT

The application described in this section pertains to the assessment of arithmetic skills, which are fundamental mathematical abilities necessary for performing basic operations with numbers and essential for everyday problem solving and decision making. The data analyzed (available at https://osf.io/kgzsw/?view_only=19a7af3d9f214675909c236942f4e395) were collected as part of a study on the assessment of skill attainment in CbKST by Anselmi et al. (2017). A total of 129 children in the third year of primary school completed two equivalent paper-and-pencil versions of an arithmetic test two months apart. Only the pretest data are considered here.

The Arithmetic Test

The arithmetic test consists of 12 open-response items, and the responses were coded as correct or incorrect prior to analysis. Six skills are needed to solve the items: (*a*) addition without carrying, (*b*) subtraction without column borrowing, (*c*) addition with carrying, (*d*) subtraction with column borrowing, (*e*) multiplication, and (*f*) division. Table 3 shows the skills associated with each item. The test is conjunctive because all the skills associated with an item are assumed to be necessary to solve it. For example, Item 12 reads: "Francis has a total of 72 coloured pencils and wants to store them in nine pencil cases. He then decides to put another 17 pencils in each pencil case. How many pencils will each pencil case contain?", and it is assumed to require skills *a* ("addition without carrying"), *c* ("addition with carrying"), and *f* ("division"). Skill *a* is assumed to be a prerequisite for skill *c*, and skill *b* is assumed to be a prerequisite for skill *d*, which means that *c* cannot be mastered without also mastering *a*, and *d* cannot be mastered without also mastering *b*. Applying the considered prerequisite relations to the six skills results in the competence structure shown in Figure 2. This structure consists of 36 competence states, representing 36 different profiles of the six skills that are expected to characterize primary school students. It should also be noted that, consistent with the defined prerequisite relations, any competence state that includes *c* also includes *a*, and any competence state that includes *d* also includes *b*.

TABLE 3
Arithmetic test items and associated skills

Item	Associated skills	Careless error probability	Lucky guess probability	Careless error plus lucky guess probability
1	$\{a, b, c, d\}$	.337	.000	.337
2	$\{a, e, f\}$	.039	.216	.255
3	$\{a, e, f\}$	.066	.108	.173
4	$\{a, b, c, d, e\}$	.397	.046	.443
5	$\{a, b, c, d\}$	.137	.071	.208
6	$\{b, d, e\}$	.072	.340	.412
7	$\{a, b, c, e\}$	.143	.252	.395
8	$\{f\}$	.095	.337	.432
9	$\{e\}$	.066	.555	.620
10	$\{a, c, f\}$	.500	.000	.500
11	$\{a, f\}$	.026	.151	.177
12	$\{a, c, f\}$	.068	.000	.068

Note. *a* = addition without carrying; *b* = subtraction without column borrowing; *c* = addition with carrying; *d* = subtraction with column borrowing; *e* = multiplication; *f* = division.

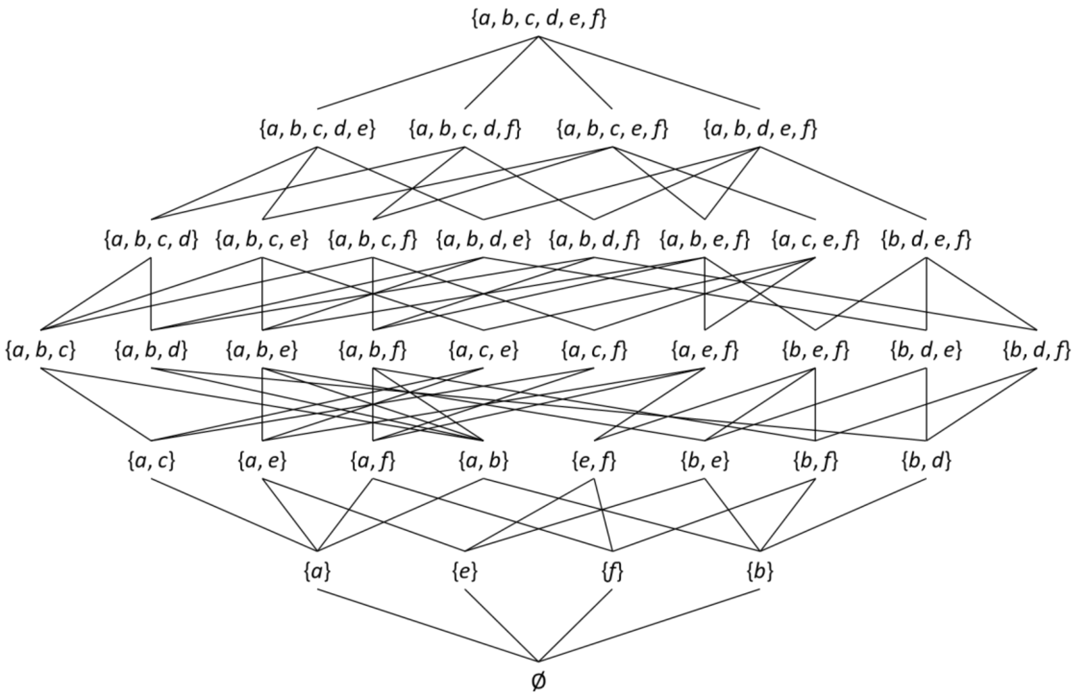

FIGURE 2
Competence structure on the six skills considered in the arithmetic test
Note. *a* = addition without carrying; *b* = subtraction without column borrowing; *c* = addition with carrying; *d* = subtraction with column borrowing; *e* = multiplication; *f* = division.

Results

Validation of the Competence Structure and the Test

The goodness-of-fit of the CBLIM to the data is satisfactory: Pearson's chi-square statistic = 4420.6, df = 4055, bootstrapped p = .115. With the exception of the careless error probability of Item 10 and the lucky guess probability of Item 9, all other lucky guess and careless error probabilities are quite low (Table 3). For no item is the sum of the careless error and lucky guess probabilities greater than 1. Taken together, these results suggest that the prerequisite relationships among the skills are valid and that the items are well formulated, actually require the skills they are supposed to require, and provide valid information for evaluating skill mastery.

Identification of the Competence States from the Responses to Test Items

The median of the likelihoods of the students' uncovered knowledge states is .976 (first and third quartiles = .867, .999). This result indicates that, overall, there is a high level of confidence that the uncovered knowledge states correspond to the students' actual knowledge states.

The 36 competence states in Figure 2 delineate only 17 distinct knowledge states, indicating that the test is not fully informative. For 87 of the 129 students, there is a unique competence state that delineates the uncovered knowledge state. Thus, given the high level of confidence that the uncovered knowledge states correspond to the students' actual knowledge states, we can be confident about the skills mastered by these 87 students (i.e., that their uncovered competence states reflect their actual competence states). For the remaining 42 students, there is no unique competence state delineating their uncovered knowledge state. In this study, the uncovered competence state of each of these 42 students is taken to be the one that consists only of those particular skills that the students master for sure. Eight competence states are identified based on the student's responses to the 12 items: Ø, $\{a, c, f\}$, $\{a, b, c, d\}$, $\{a, b, c, d, f\}$, $\{a, b, c, d, e\}$, $\{a, b, d, e, f\}$, $\{a, b, c, e, f\}$, and $\{a, b, c, d, e, f\}$. Of these, the uncovered competence states of the 42 students without unique skill assessment are either Ø or $\{a, c, f\}$. It is possible that these 42 students master additional skills. However, their mastery cannot be uniquely identified from the test items.

Table 4 shows the number of students with each uncovered competence state and the number of students with a given score and uncovered competence state. The most frequently uncovered competence states are $\{a, b, c, d, e, f\}$, Ø, $\{a, b, c, d, e\}$, and $\{a, b, c, d, f\}$, which characterize 53, 40, 15, and 12 students, respectively.

Due to careless errors and lucky guesses, students with the same competence state may have different test scores. For example, the test scores of the 12 students with competence state $\{a, b, c, d, f\}$ range from 4 to 8 (Table 4). The knowledge state delineated by competence state $\{a, b, c, d, f\}$ is {1, 5, 8, 10, 11, 12}. Since these students are capable of solving six items, their expected test score is 6 if neither careless errors nor lucky guesses occur. The observed test scores of these students range from 4 to 8. Thus, some of them responded carelessly to up to two items, while others guessed the responses to up to two items.

Students with the same test score may have different competence states. For example, the uncovered competence state of the six students with a test score of 6 is either $\{a, b, c, d, f\}$ or $\{a, b, c, d, e\}$ (Table 4).

TABLE 4
Number of students with a given score at the arithmetic test and uncovered competence state

Test score	Freq. (Prop.)	Uncovered competence state							
		Ø 40 (.310)	$\{a, c, f\}$ 2 (.016)	$\{a, b, c, d\}$ 1 (.008)	$\{a, b, c, d, f\}$ 12 (.093)	$\{a, b, c, d, e\}$ 15 (.116)	$\{a, b, d, e, f\}$ 1 (.008)	$\{a, b, c, e, f\}$ 5 (.039)	$\{a, b, c, d, e, f\}$ 53 (.411)
0	8 (.062)	8	0	0	0	0	0	0	0
1	11 (.085)	11	0	0	0	0	0	0	0
2	6 (.047)	6	0	0	0	0	0	0	0
3	9 (.070)	8	0	1	0	0	0	0	0
4	11 (.085)	5	2	0	2	2	0	0	0
5	7 (.054)	2	0	0	3	2	0	0	0
6	6 (.047)	0	0	0	2	4	0	0	0
7	8 (.062)	0	0	0	2	4	1	0	1
8	11 (.085)	0	0	0	3	2	0	3	3
9	12 (.093)	0	0	0	0	1	0	2	9
10	12 (.093)	0	0	0	0	0	0	0	12
11	18 (.140)	0	0	0	0	0	0	0	18
12	10 (.078)	0	0	0	0	0	0	0	10

Note. a = addition without carrying; b = subtraction without column borrowing; c = addition with carrying; d = subtraction with column borrowing; e = multiplication; f = division.

Discussion

Based on the results of the assessment, the 129 students are characterized by eight different competence states, and we can be quite confident that these represent the students' actual competence states. However, there are 42 students for whom the test does not provide enough information to unambiguously identify the mastered skills. Students with the same competence state may have different test scores and students with the same test score may have different competence states. In educational assessment, considering an individual's competence state rather than relying solely on their test score provides powerful information for developing personalized didactic interventions. For example, an intervention for a student with competence state $\{a, b, c, d, f\}$ should aim to promote the attainment of skill e ("multiplication"), whereas an intervention for a student with competence state $\{a, b, c, d, e\}$ should aim to promote the attainment of skill f ("division").

Application in Clinical Assessment

The application described in this section pertains to the assessment of social phobia, which is defined as an excessive and unreasonable fear of social situations. Social phobia is one of the most common anxiety disorders and one of the most common psychiatric disorders overall (Kessler et al., 2005; National Collaborating Centre for Mental Health, 2013). The data analyzed (available in Robitzsch et al., 2022) consist of the responses of 863 individuals to a social phobia test, which were simulated based on the results of an empirical study conducted by Fang et al. (2019) using the same test.

The Social Phobia Test

The test consists of 13 yes/no items from the National Epidemiological Survey on Alcohol and Related Conditions (Grant et al., 2003), designed according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (American Psychiatric Association, 1994). Fang et al. (2019) analyzed the responses to these items and identified three symptoms of social phobia that are responsible for an affirmative answer: (a) fear of public performance, (b) fear of scrutiny, and (c) fear of interaction with other people. Table 5 shows the symptoms associated with each item. The test is disjunctive because it is assumed that any one of the symptoms associated with an item is sufficient for an affirmative response. For example, it is sufficient for an individual to be afraid of performing in public (symptom a) or being scrutinized (symptom b) to respond affirmatively to Item 3 "[Have you ever had a strong fear or avoidance of] taking part or speaking at a meeting." Additionally, the authors assumed that symptom a is a prerequisite for both symptoms b and c, meaning that a person must be afraid of performing in public (symptom a) in order to be afraid of being scrutinized (symptom b) or of interacting with other people (symptom c). Applying the considered prerequisite relations to the three symptoms results in the competence structure shown in Figure 3, which consists of five competence states.

TABLE 5
Social phobia test items and associated symptoms

Item	Associated symptoms	False negative probability	False positive probability	False negative plus false positive probability
1	$\{a\}$	.052	.488	.540
2	$\{a, b\}$	.116	.250	.366
3	$\{a, b\}$	.214	.122	.336
4	$\{a, b, c\}$	.187	.183	.370
5	$\{b\}$	.243	.138	.382
6	$\{b\}$	.644	.047	.691
7	$\{b, c\}$	.349	.144	.493
8	$\{b, c\}$	.417	.179	.596
9	$\{c\}$	.632	.019	.651
10	$\{c\}$	.076	.171	.247
11	$\{c\}$	.099	.182	.281
12	$\{b, c\}$	.515	.150	.665
13	$\{b, c\}$	.768	.003	.771

Note. *a* = fear of public performance; *b* = fear of scrutiny; *c* = fear of interaction with other people.

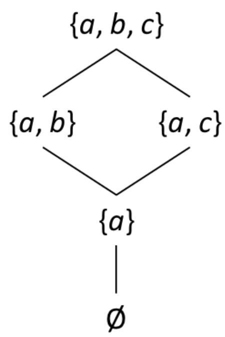

FIGURE 3
Competence structure on the three symptoms considered in the social phobia test
Note. *a* = fear of public performance; *b* = fear of scrutiny; *c* = fear of interaction with other people.

Results

Validation of the Competence Structure and the Test

Investigating the goodness-of-fit of the CBLIM to the data is not relevant in this study because the competence structure (Figure 3) and the disjunctive test (Table 5) are the same as those used to simulate the data. If the estimates of the false negative and false positive probabilities are similar to those observed in empirical data, they provide valuable information about the test and its assessment capabilities. One possible interpretation is that the estimates of the false negative and false positive probabilities may reveal tendencies to dissimulate or simulate the symptoms evaluated by the items, respectively. Under this interpretation, dissimulation may be particularly likely for Items 6, 8, 9, 12, and 13, which have high false negative probabilities, while simulation may occur for Item 1, which has a high false positive probability (Table 5). Another possible reason for high false negative probabilities could be individuals' poor awareness of their own symptoms. For no item does the sum of the false negative and false positive probabilities exceed 1 (Table 5), suggesting that the items provide valid information for evaluating the presence of symptoms.

Identification of the Competence States from the Responses to Test Items

The median of the likelihoods of the individuals' uncovered knowledge states is .871 (first and third quartiles = .728, .982), indicating a high level of confidence that these knowledge states correspond to the individuals' actual knowledge states. The five competence states in Figure 3 delineate five distinct knowledge states, which means that the test is fully informative. Thus, the competence state of any individual can be uniquely uncovered from their knowledge state. Given the high level of confidence that the uncovered knowledge states correspond to the individuals' actual knowledge states, along with the test's full informativeness, there is a strong level of confidence that the uncovered competence states reflect the individuals' actual competence states.

Table 5 shows the number of students with each uncovered competence state and the number of individuals with a given score and uncovered competence state. Competence state Ø is the most frequent, followed by $\{a\}$, $\{a, b, c\}$, $\{a, b\}$, and $\{a, c\}$.

Due to false negative and false positive responses, individuals with the same competence state may have different test scores. For example, the test scores of the 243 individuals with competence state $\{a\}$ range from 3 to 7 (Table 6). The knowledge state delineated by competence state $\{a\}$ is $\{1, 2, 3, 4\}$. Since these individuals are expected to respond affirmatively to four items, their expected test score is 4 if neither false negatives nor false positives occur. It is worth noting that 4 is the most frequently observed test score among these 243 individuals, with 81 of them achieving it. However, since the observed test scores of these individuals range from 3 to 7, some had a false negative on one item, while others had false positives on up to three items.

Additionally, individuals with the same test score may have different competence states. For example, the uncovered competence state of the 48 individuals with a test score of 10 is $\{a, b\}$, $\{a, c\}$, or $\{a, b, c\}$, although $\{a, b, c\}$ is the most frequent among them (40 out of 48 individuals; Table 6).

Discussion

Based on the results of the assessment, we can conclude that the 863 individuals are characterized by five different competence states, and we can be fairly confident that these represent the individuals' actual competence states. In clinical assessment, considering an individual's competence state rather than relying solely on their test score provides powerful information for developing personalized therapeutic interventions. For example, intervention for an individual with competence state $\{a, b\}$ should focus on resolving symptoms *a* and *b* of social phobia (fears of public performance and scrutiny, respectively), whereas intervention for an individual with competence state $\{a, c\}$ should focus on resolving symptoms *a* and *c* (fears of public performance and interaction with other people, respectively).

TABLE 6
Number of individuals with a given score at the social phobia test and uncovered competence state

Test score	Freq. (Prop.)	Uncovered competence state				
		Ø 309 (.358)	{*a*} 243 (.282)	{*a*, *b*} 100 (.116)	{*a*, *c*} 68 (.079)	{*a*, *b*, *c*} 143 (.166)
0	43 (.050)	43	0	0	0	0
1	76 (.088)	76	0	0	0	0
2	81 (.094)	81	0	0	0	0
3	111 (.129)	67	44	0	0	0
4	123 (.143)	37	86	0	0	0
5	86 (.100)	5	71	6	2	2
6	88 (.102)	0	37	21	21	9
7	56 (.065)	0	5	33	10	8
8	54 (.063)	0	0	18	19	17
9	60 (.070)	0	0	18	11	31
10	48 (.056)	0	0	3	5	40
11	28 (.032)	0	0	1	0	27
12	9 (.010)	0	0	0	0	9
13	0 (.000)	0	0	0	0	0

Note *a* = fear of public performance; *b* = fear of scrutiny; *c* = fear of interaction with other people.

FINAL REMARKS

CbKST allows for uncovering the attributes (skills, abilities, symptoms of a disorder) that an individual possesses based on their responses to each of the test items. As such, it represents a sharp departure from CTT and IRT, which aggregate information across items by forming a test score. This work has outlined the main characteristics of an assessment conducted within CbKST, exploring different areas of assessment (educational and clinical), types of tests (conjunctive and disjunctive), and the informativeness of the tests (fully informative and nonfully informative). Within CbKST, it is possible to evaluate the confidence with which the uncovered competence state represents the individual's actual competence state. Based on the individual's competence state, personalized interventions can be developed. This section builds on the previous sections and extends the argument in several directions.

Some works within CbKST have focused on the properties of tests that allow for uniquely uncovering the skills an individual masters from their item responses (Heller et al., 2015, 2017). Methods have been proposed for developing tests that are as informative as possible about the skills mastered and, if desired, also minimal (i.e., no item can be removed without reducing the informativeness of the tests; Anselmi et al., 2022; Anselmi, Heller, Stefanutti, & Robusto, 2024). These methods have been successfully used to develop new tests from scratch and to shorten and improve existing tests (Anselmi, Heller, Stefanutti, Robusto, & Barillari, 2024). An application of these methods to the 12-item arithmetic test considered in this work would indicate which specific items should be added to make the test fully informative. For example, adding an item that requires skills *a* and *c* would allow for unambiguously determining whether an individual masters both skills *a* and *c* or skill *a* only. This distinction is not possible with the original 12-item arithmetic test. Similarly, an application of these methods to the 13-item social phobia test would indicate which specific items could be removed without making the test less informative. For example, a shortened form consisting of Items 4, 5, and 9 is as informative about the symptoms of social phobia experienced by the individuals as the original 13-item test.

The CbKST assessment illustrated in this work is based on the administration of all test items. However, CbKST also enables computerized adaptive assessments that aim at accurately uncovering an individual's competence state by presenting them with a minimum number of items (Anselmi, de Chiusole, & Heller, 2024; Anselmi et al., 2016; Dowling & Hockemeyer, 2001; Falmagne & Doignon, 1988a, 1988b; Heller & Repitsch, 2012). The items presented next are appropriately selected based on responses to previously administered items. The assessment is tailored to the individual (i.e., they are presented with items that are neither too difficult nor too easy for them) so that it is neither demotivating nor boring. Implementations of CbKST include Stat-Knowlab (de Chiusole et al., 2020), PsycAssist (de Chiusole et al., 2024), and ATS-PD (acronym for Adaptive Testing System for Psychological Disorders; Donadello et al., 2017), which are designed for the adaptive assessment of basic statistics, neuropsychological functioning, and psychological disorders, respectively.

CbKST is not the only framework that can be used to identify the attributes that an individual possesses from their responses to test items. Cognitive diagnostic models (CDMs; DiBello et al., 2007; Rupp et al., 2010; Tatsuoka 1990) also serve this purpose. CbKST and CDMs have developed more or less independently in parallel for decades, with essentially no crosstalk between them. Each uses its own notation and defines its own concepts. However, there is a close correspondence between these two frameworks, ranging from very basic concepts to specific models. The interconnections between CbKST and CDMs were first established by Heller et al. (2015, 2016) and further elaborated by Anselmi, Noventa, and Heller (2024), Heller (2022), Noventa et al. (2024), and Stefanutti et al. (2025).

The CbKST assessment illustrated in this work relies on conjunctive or disjunctive tests consisting of dichotomous items. However, the theory allows for the assessment to be extended in at least two ways. On the one hand, it allows for handling tests where multiple strategies can be used to solve the items, each involving the application of one or more skills (Heller et al., 2013; Heller & Stefanutti, 2024). For example, consider the item "A bag contains 5-cent, 10-cent, and 20-cent pieces. The probability of drawing a 5-cent piece is .35, that of drawing a 10-cent piece is .20 and that of drawing a 20-cent piece is .45. What is the probability that the coin randomly drawn is a 5- cent or a 20-cent piece?" A student could solve this item using either the union of mutually exclusive events or the probability of the complement of an event (Anselmi et al., 2012). On the other hand, recent developments in KST (Heller, 2021; Stefanutti, Anselmi, et al., 2020; Stefanutti, de Chiusole, et al., 2020; see also Schrepp, 1997) and CbKST (Stefanutti et al., 2023) provide the necessary tools to assess the skills an individual possesses based on their responses to polytomous items. In some cases, polytomous items are preferable to dichotomous items because they allow for a more nuanced or refined assessment. Developing CbKST assessments for polytomous items along the lines sketched above will be the aim of future work.

REFERENCES

Albert, D., & Held, T. (1999). Component-based knowledge spaces in problem solving and inductive reasoning. In D. Albert & J. Lukas (Eds.), *Knowledge spaces: Theories, empirical research, and applications* (pp. 15-40). Lawrence Erlbaum Associates.

Algina, J., & Penfield, R. D. (2009). Classical test theory. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology* (pp. 93-122). Sage Publications Ltd. https://doi.org/10.4135/9780857020994.n5

Allen, M. J., & Yen, W. M. (2001). *Introduction to measurement theory*. Waveland Press.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). American Psychiatric Pub Inc.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision ed.). American Psychiatric Pub Inc.

Anselmi, P., de Chiusole, D., & Heller, J. (2024). Adaptive assessment. In J. Heller & L. Stefanutti (Eds.), *Knowledge structures: Recent developments in theory and application* (pp. 159-180). World Scientific. https://doi.org/10.1142/9789811280481_0007

Anselmi, P., Heller, J., Stefanutti, L., & Robusto, E. (2022). Constructing, improving, and shortening tests for skill assessment. *Journal of Mathematical Psychology*, *106*, Article 102621. https://doi.org/10.1016/j.jmp.2021.102621

Anselmi, P., Heller, J., Stefanutti, L., & Robusto, E. (2024). Constructing tests for skill assessment with competence-based test development. *British Journal of Mathematical and Statistical Psychology*, *77*(3), 429-458. https://doi.org/10.1111/bmsp.12335

Anselmi, P., Heller, J., Stefanutti, L., Robusto, E., & Barillari, G. (2024). *Real life applications of competence-based test development to the construction, improvement, and shortening of tests* [Manuscript submitted for publication].

Anselmi, P., Noventa, S., & Heller, J. (2024). Knowledge structures and related theories. In J. Heller & L. Stefanutti (Eds.), *Knowledge structures: Recent developments in theory and application* (pp. 53-84). World Scientific. https://doi.org/10.1142/9789811280481_0003

Anselmi, P., Robusto, E., & Stefanutti, L. (2012). Uncovering the best skill multimap by constraining the error probabilities of the gain-loss model. *Psychometrika*, *77*(4), 763-781. https://doi.org/10.1007/S11336-012-9286-0

Anselmi, P., Robusto, E., & Stefanutti, L. (2013). A procedure for identifying the best skill multimap in the gain-loss model. *Electronic Notes in Discrete Mathematics*, *42*, 9-16. https://doi.org/10.1016/j.endm.2013.05.140

Anselmi, P., Robusto, E., Stefanutti, L., & de Chiusole, D. (2016). An upgrading procedure for adaptive assessment of knowledge. *Psychometrika*, *81*(2), 461-482. https://doi.org/10.1007/s11336-016-9498-9

Anselmi, P., Stefanutti, L., de Chiusole, D., & Robusto, E. (2017). The assessment of knowledge and learning in competence spaces: The gain-loss model for dependent skills. *British Journal of Mathematical and Statistical Psychology*, *70*(3), 457-479. https://doi.org/10.1111/bmsp.12095

Bacherini, A., Anselmi, P., Havercamp, S. M., & Balboni, G. (2024). Psychometric properties of the Beliefs About Adults with ID Scale in American physicians: Application of classical test and Rasch measurement theories. *Journal of Intellectual Disability Research*, *68*(10), 1129-1145. https://doi.org/10.1111/jir.13143

Brancaccio, A., de Chiusole, D., & Wickelmaier, F. (2024). Software packages for knowledge structure theory. In J. Heller & L. Stefanutti (Eds.), *Knowledge structures: Recent developments in theory and application* (pp. 297-322). World Scientific. https://doi.org/10.1142/9789811280481_0012

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for administration and scoring. Minnesota Multiphasic Personality Inventory-2: MMPI-2*. University of Minnesota Press.

Colledani, D., Anselmi, P., & Robusto, E. (2019a). Development of a new abbreviated form of the Eysenck Personality Questionnaire-Revised with multidimensional item response theory. *Personality and Individual Differences*, *149*, 108-117. https://doi.org/10.1016/j.paid.2019.05.044

Colledani, D., Anselmi, P., & Robusto, E. (2019b). Using multidimensional item response theory to develop an abbreviated form of the Italian version of Eysenck's IVE questionnaire. *Personality and Individual Differences*, *142*, 45-52. https://doi.org/10.1016/j.paid.2019.01.032

Colledani, D., Boragno, P., Fiabane, E. M., Setti, I., & Gabanelli, P. (2024). Further validation of the persecutory ideation questionnaire in the Italian context: Results from classical and modern test theory. *Assessment*. Advance online publication. https://doi.org/10.1177/10731911241257012

Colledani, D., Robusto, E., & Anselmi, P. (2018). Development of a new abbreviated form of the Junior Eysenck Personality Questionnaire-Revised. *Personality and Individual Differences*, *120*, 159-165. https://doi.org/10.1016/j.paid.2017.08.037

de Chiusole, D., Spinoso, M., Anselmi, P., Bacherini, A., Balboni, G., Mazzoni, N., Brancaccio, A., Epifania, O. M., Orsoni, M., Giovagnoli, S., Garofalo, S., Benassi, M., Robusto, E., Stefanutti, L., & Pierluigi, I. (2024). PsycAssist: A web-based artificial intelligence system designed for adaptive neuropsychological assessment and training. *Brain Sciences*, *14*(2), Article 122. https://doi.org/10.3390/brainsci14020122

de Chiusole, D., Stefanutti, L., Anselmi, P., & Robusto, E. (2013). Assessing parameter invariance in the BLIM: Bipartition models. *Psychometrika*, *78*(4), 710-724. https://doi.org/10.1007/S11336-013-9325-5

de Chiusole, D., Stefanutti, L., Anselmi, P., & Robusto, E. (2015a). Modeling missing data in knowledge space theory. *Psychological Methods*, *20*(4), 506-522. https://doi.org/10.1037/met0000050

de Chiusole, D., Stefanutti, L., Anselmi, P., & Robusto, E. (2015b). Naïve tests of basic local independence model's invariance. *The Spanish Journal of Psychology*, *18*, Article E26. https://doi.org/10.1017/sjp.2015.24

de Chiusole, D., Stefanutti, L., Anselmi, P., & Robusto, E. (2018). Testing the actual equivalence of automatically generated items. *Behavior Research Methods*, *50*, 39-56. https://doi.org/10.3758/s13428-017-1004-5

de Chiusole, D., Stefanutti, L., Anselmi, P., & Robusto, E. (2020). Stat-Knowlab. Assessment and learning of statistics with competence-based knowledge space theory. *International Journal of Artificial Intelligence in Education*, *30*, 668-700. https://doi.org/10.1007/s40593-020-00223-1

DiBello, L. V., Roussos, L. A., & Stout, W. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. Rao & S. Sinharay (Eds.), *Handbook of statistics* (Vol. 26, pp. 970-1030). North-Holland Publications.

Doignon, J.-P. (1994). Knowledge spaces and skill assignments. In G. H. Fisher & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics and methodology* (pp. 111-121). Springer-Verlag.

Doignon, J. P., & Falmagne, J. C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, *23*(2), 175-196. https://doi.org/10.1016/S0020-7373(85)80031-6

Donadello, I., Spoto, A., Sambo, F., Badaloni, S., Granziol, U., & Vidotto, G. (2017). ATS-PD: An adaptive testing system for psychological disorders. *Educational and Psychological Measurement*, *77*(5), 792-815. https://doi.org/10.1177/0013164416652188

Dowling, C., & Hockemeyer, C. (2001). Automata for the assessment of knowledge. *IEEE Transactions on Knowledge and Data Engineering*, *13*(3), 451-461. https://doi.org/10.1109/69.929902

Embretson, S. E., & Reise, S. P. (2020). *Item response theory for psychologists.* Lawrence Erlbaum Associates.

Falmagne, J.-C., & Doignon, J.-P. (1988a). A class of stochastic procedures for the assessment of knowledge. *British Journal of Mathematical and Statistical Psychology*, *41*(1), 1-23. https://doi.org/10.1111/j.2044-8317.1988.tb00884.x

Falmagne, J.-C., & Doignon, J.-P. (1988b). A Markovian procedure for assessing the state of a system. *Journal of Mathematical Psychology*, *32*(3), 232-258. https://doi.org/10.1016/0022-2496(88)90011-9

Falmagne, J.-C., Koppen, M., Villano, M., Doignon, J.-P., & Johanessen, L. (1990). Introduction to knowledge spaces: How to build, test and search them. *Psychological Review*, *97*(2), 204-224. https://doi.org/10.1037/0033-295X.97.2.201

Fang, G., Liu, J., & Ying, Z. (2019). On the identifiability of diagnostic classification models. *Psychometrika*, *84*(1), 19-40. https://doi.org/10.1007/s11336-018-09658-x

Grant, B. F., Kaplan, K., Shepard, J., & Moore, T. (2003). *Source and accuracy statement for wave 1 of the 2001-2002 national epidemiologic survey on alcohol and related conditions*. National Institute on Alcohol Abuse and Alcoholism.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Press.

Heller, J. (2017). Identifiability in probabilistic knowledge structures. *Journal of Mathematical Psychology*, *77*, 46-57. https://doi.org/10.1016/j.jmp.2016.07.008

Heller, J. (2021). Generalizing quasi-ordinal knowledge spaces to polytomous items. *Journal of Mathematical Psychology*, 101, Article 102515. https://doi.org/10.1016/j.jmp.2021.102515

Heller, J. (2022). Complete *Q*-matrices in conjunctive models on general attribute structures *British Journal of Mathematical and Statistical Psychology*, *75*(3), 522-549. https://doi.org/10.1111/bmsp.12266

Heller, J., Anselmi, P., Stefanutti, L., & Robusto, E. (2017). A necessary and sufficient condition for unique skill assessment. *Journal of Mathematical Psychology*, *79*, 23-28. https://doi.org/10.1016/j.jmp.2017.05.004

Heller, J., & Repitsch, C. (2012). Exploiting prior information in stochastic knowledge assessment. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *8*(1), 12-22. https://doi.org/10.1027/1614-2241/a000035

Heller, J., & Stefanutti, L. (2024). Knowledge structures and their competence-based extension. In J. Heller & L. Stefanutti (Eds.), *Knowledge structures: Recent developments in theory and application* (pp. 3-26). World Scientific. https://doi.org/10.1142/9789811280481_0001

Heller, J., Stefanutti, L., Anselmi, P., & de Chiusole, D. (2024). Parameter estimation and model validation. In J. Heller & L. Stefanutti (Eds.), *Knowledge structures: Recent developments in theory and application* (pp. 105-132). World Scientific. https://doi.org/10.1142/9789811280481_0005

Heller, J., Stefanutti, L., Anselmi, P., & Robusto, E. (2015). On the link between cognitive diagnostic models and knowledge space theory. *Psychometrika*, *80*(4), 995-1019. https://doi.org/10.1007/s11336-015-9457-x

Heller, J., Stefanutti, L., Anselmi, P., & Robusto, E. (2016). Erratum to: On the link between cognitive diagnostic models and knowledge space theory. *Psychometrika*, *81*(1), 250-251. https://doi.org/10.1007/s11336-015-9494-5

Heller, J., Ünlü, A., & Albert, D. (2013). Skills, competencies and knowledge structures. In J.-C. Falmagne, D. Albert, C. Doble, D. Eppstein, & X. Hu (Eds.), *Knowledge spaces: Applications in education* (pp. 229-242). Springer-Verlag.

Heller, J., & Wickelmaier, F. (2013). Minimum discrepancy estimation in probabilistic knowledge structures. *Electronic Notes in Discrete Mathematics*, *42*, 49-56. https://doi.org/10.1016/j.endm.2013.05.145

Hockemeyer, C. (2024). *R packages for knowledge space theory*. Retrieved from https://resources.cord-hockemeyer.info/techreports/Hockemeyer_KST-R_v2.pdf

Hubin, D. R. (1988). *The Scholastic Aptitude Test: Its development and introduction, 1900-1948*. University of Oregon.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593-602. https://doi.org/10.1001/archpsyc.62.6.593

Korossy, K. (1999). Modeling knowledge as competence and performance. In D. Albert & J. Lukas (Eds.), *Knowledge spaces: Theories, empirical research, and applications* (pp. 103-132). Lawrence Erlbaum Associates.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Lawrence Erlbaum Associates.

National Collaborating Centre for Mental Health). (2013). *Social anxiety disorder: Recognition, assessment and treatment.* British Psychological Society.

Noventa, S., Heller, J., & Kelava, A. (2024). Toward a unified perspective on assessment models, part I: Foundations of a framework. *Journal of Mathematical Psychology*, *122*, Article 102872. https://doi.org/10.1016/j.jmp.2024.102872

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, *3*(1), 1-18. https://doi.org/10.1016/0022-2496(66)90002-2

Robitzsch, A., Kiefer, T., George, A. C., & Ünlü, A. (2022). *Cognitive diagnosis modeling* (Version 8.2-6). https://cran.r-project.org/web/packages/CDM/CDM.pdf

Robusto, E., Stefanutti, L., & Anselmi, P. (2010). The Gain-Loss Model: A probabilistic skill multimap model for assessing learning processes. *Journal of Educational Measurement*, *47*(3), 373-394. https://doi.org/10.1111/j.1745-3984.2010.00119.x

Roussos, L. A., Templin, J. L., & Henson, R. A. (2007). Skill diagnosis using IRT-based latent class models. *Journal of Educational Measurement*, *44*(4), 293-311. https://doi.org/10.1111/j.1745-3984.2007.00040.x

Rupp, A. A., & Templin, J. (2008). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, *68*(1), 78-96. https://doi.org/10.1177/0013164407301545

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.

Schrepp, M. (1997). A generalization of knowledge space theory to problems with more than two answer alternatives. *Journal of Mathematical Psychology*, *41*(3), 237-243. https://doi.org/10.1006/jmps.1997.1169

Serra, F., Spoto, A., Ghisi, M., & Vidotto, G. (2015). Formal psychological assessment in evaluating depression: A new methodology to build exhaustive and irredundant adaptive questionnaires. *PLoS One*, *10*(4), Article e0122131. https://doi.org/10.1371/journal.pone.0122131

Spoto, A., Stefanutti, L., & Vidotto, G. (2013). Considerations about the identification of forward- and backward-graded knowledge structures. *Journal of Mathematical Psychology*, *57*(5), 249-254. https://doi.org/10.1016/j.jmp.2013.09.002

Stefanutti, L., Anselmi, P., de Chiusole, D., & Spoto, A. (2020). On the polytomous generalization of knowledge space theory. *Journal of Mathematical Psychology*, *94*, Article 102306. https://doi.org/10.1016/j.jmp.2019.102306

Stefanutti, L., Anselmi, P., de Chiusole, D., Spoto, A., & Heller, J. (2025). The BLIM, the DINA and their polytomous extensions. Rejoinder to the Commentary by Chiu, Köhn and Ma. *Psychometrika*. Advance online publication. https://doi.org/10.1017/psy.2024.24

Stefanutti, L., de Chiusole, D., Anselmi, P., & Spoto, A. (2020). Extending the basic local independence model to polytomous data. *Psychometrika*, *85*(3), 684-715. https://doi.org/10.1007/s11336-020-09722-5

Stefanutti, L., Heller, J., Anselmi, P., & Robusto, E. (2012). Assessing the local identifiability of probabilistic knowledge structures. *Behavior Research Methods*, *44*, 1197-1211. https://doi.org/10.3758/s13428-012-0187-z

Stefanutti, L., & Robusto, E. (2009). Recovering a probabilistic knowledge structure by constraining its parameter space. *Psychometrika*, *74*(1), 83-96. https://doi.org/10.1007/S11336-008-9095-7

Stefanutti, L., Spoto, A., Anselmi, P., & de Chiusole, D. (2023). Towards a competence-based polytomous knowledge structure theory. *Journal of Mathematical Psychology*, *115*, Article 102781. https://doi.org/10.1016/j.jmp.2023.102781

Tassé, M. J., Schalock, R. L., Thissen, D., Balboni, G., Bersani Jr, H., Borthwick-Duffy, S. A., Spreat, S., Widaman, K. F., Zhang, D., & Navas, P. (2016). Development and standardization of the diagnostic adaptive behavior scale: Application of item response theory to the assessment of adaptive behavior. *American Journal on Intellectual and Developmental Disabilities*, *121*(2), 79-94. https://doi.org/10.1352/1944-7558-121.2.79

Tatsuoka, K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Safto (Eds.), *Monitoring skills and knowledge acquisition* (pp. 453-488). Lawrence Erlbaum Associates.

Templin, J., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*(3), 287-305. https://doi.org/10.1037/1082-989X.11.3.287

Wickelmaier, F., Heller, J., Mollenhauer, J., Anselmi, P., de Chiusole, D., Brancaccio, A., & Stefanutti, L. (2024). *pks: Probabilistic Knowledge Structures* (Version 0.6-1). https://cran.r-project.org/web/packages/pks/index.html