

Development and Initial Validation of the Computer-Based Orthographic Processing Assessment Short Form: An Application of Cognitive Diagnostic Modeling

Journal of Psychoeducational Assessment
2025, Vol. 43(3) 310–327
© The Author(s) 2024



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/07342829241304165
journals.sagepub.com/home/jpa



Yi-Jui I. Chen¹ , Yi-Jhen Wu² , Yi-Hsin Chen¹ , and Robin Irey³

Abstract

A short form of the 60-item computer-based orthographic processing assessment (long-form COPA or COPA-LF) was developed. The COPA-LF consists of five skills, including rapid perception, access, differentiation, correction, and arrangement. Thirty items from the COPA-LF were selected for the short-form COPA (COPA-SF) based on cognitive diagnostic modeling (CDM) analyses. Both parametric and nonparametric CDM and correlational analyses were conducted with the responses of 74 U.S. second-grade students on the COPA-SF. Results supported five skills of orthographic processing represented on the COPA-SF. Concurrent and convergent validity evidence was found for the COPA-SF. Overall, the findings suggest that the COPA-SF is a time-efficient, reliable, multidimensional measure of orthographic processing.

Keywords

short form, orthographic processing, cognitive diagnostic models, nonparametric classification method

Introduction

Orthographic processing, “the ability to form, store, and access orthographic representations” (Stanovich & West, 1989, p. 404), is a foundational skill for later word reading and spelling

¹University of South Florida, Tampa, FL, USA

²TU Dortmund, Dortmund, Germany

³University of California—San Francisco, San Francisco, CA, USA

Corresponding Author:

Yi-Jhen Wu, Center for Research on Education and School Development, TU Dortmund, Vogelpothsweg 78, Dortmund 44227, Germany.

Email: yi-jhen.wu@tu-dortmund.de

acquisition (Ehri, 2014; Seidenberg & McClelland, 1989; Share, 1995; Stanovich & West, 1989). A series of studies have consistently demonstrated that orthographic processing skills account for the variance of word reading and spelling in monolingual (Chen et al., 2021; Conrad & Deacon, 2016; Cunningham, 2006; Zarić et al., 2021) and bilingual school-age children (Chung et al., 2018; Deacon et al., 2013; Sun-Alperin & Wang, 2011), even after controlling phonological, general verbal and nonverbal abilities. Orthographic processing is critical for word reading and spelling, not only because it accounts for the additional variance in word reading and spelling but also because it affects how learners process word information while learning to read and spell (Chen et al., 2021).

While orthographic processing is an independent and necessary component of word reading and spelling (Apel et al., 2011; Berninger et al., 1991), students' orthographic processing ability is not often considered by teachers when facilitating students' word learning. This may be because of the lack of assessments of this skill and because teachers may have limited knowledge about the concept of orthographic processing and how it links to learning to read, since it has been less studied than other areas of reading. Another reason may be due to the time constraints of the school day, limiting teachers' ability to administer an individualized assessment of orthographic processing. Thus, it is a necessary to develop a short form of an orthographic processing assessment that can be group administered and provides teachers with a quick yet comprehensive measure of students' orthographic processing. Moreover, an orthographic processing assessment could benefit psychologists by helping them to efficiently obtain rich and timely diagnostic information about a student's development. Thus, this study aims to create a short-form version of a computer-based orthographic processing assessment (COPA; Chen et al., 2020) and evaluate the psychometric properties of the short form of COPA (COPA-SF) to determine if it demonstrates similar psychometric properties as the long-form COPA (COPA-LF) by using cognitive diagnostic modeling (CDM).

An Overview COPA: Design, Items, and Psychometric Properties

The COPA-SF was built from the COPA-LF, so we will first review how the COPA-LF was developed and validated. The COPA-LF was designed to measure students' ability to perceive, access, differentiate, correct, and arrange English letters and spelling patterns (Chen et al., 2020, 2021). The framework of the COPA-LF was based on Stanovich and West's (1989) conceptualization of orthographic processing and Cassar and Treiman's (1997) word list. Cummings's (1988) American English spelling system was used for item design principles. Previous assessments of orthographic knowledge processing have been confounded by word recognition and/or spelling ability (Apel, 2011; Burt, 2006; Castles & Nation, 2006, for review; Chen et al., 2020). In an effort to address this and assess pure orthographic processing, the COPA-LF was developed based on three principles: (a) measure the ability to process single letters, letter combinations, and spelling patterns, (b) include multiple processing skills, and (c) segregate orthographic processing from word recognition or spelling ability (Chen et al., 2020).

The COPA-LF was designed with a simple structure: each of its 60 items was intended to measure only one skill. The psychometric properties of the COPA-LF were examined using a cognitive diagnostic modeling (CDM) approach (Chen et al., 2022). This analysis revealed several key findings. The COPA-LF reflected five dimensions: rapid perception, access, differentiation, correction, and removal. It demonstrated a linear learning sequence: access → differentiation → correction → arrangement. Notably, rapid perception was not included in this sequence based on the CDM analysis. The classification accuracy of learning profiles for both the whole group and individuals exceeded .91. The COPA-LF showed strong correlations with assessments of word recognition, pseudoword reading, and spelling. However, each dimension of the COPA-LF had

relatively lower correlations with sound blending and oral language tasks. The analysis also identified eleven out of 60 items with low item quality, suggesting areas for potential refinement.

A Review of Short-form Development and Validation

From a practical perspective, a long-form assessment can be a challenge to administer in classroom settings and clinical practices. Thus, it is helpful, when possible, to develop a short form measure to reduce response burden and increase a response rate for both academic and research settings (Austin et al., 2018; Lowe, 2021). Common psychometric theories to develop a short-form measure often use confirmatory factor analysis (CFA) and item-response theory (IRT). CFA is typically used to validate a hypothesized factor structure of a short-form measure while IRT is used to validate a factor structure and assess the quality of items (Allen et al., 2015; Sturm et al., 2017). Both CFA and IRT provide continuous scores of a latent ability for individuals. To assess convergent and discriminant validity of a short-form measure, researchers generally examine correlations between latent scores in a short-form measure and relevant external measures.

Like CFA and IRT, CDM can be applied to develop a short-form measure by assessing a factor structure and the quality of items. Furthermore, CDM provides two distinct advantages that do not exist in CFA and IRT: (1) CDM can provide a set of fine-grained discrete skills to describe students' abilities, and (2) CDM can offer detailed information on hierarchical learning sequences (Hsu et al., 2023; Leighton et al., 2004). These two distinct features of CDM offer important information to practitioners because detailed diagnostic information can help them tailor reading instruction and intervention. Surprisingly, despite the distinct advantages that it provides, few studies have used CDM to develop a short-form assessment. However, this approach was best suited for our goal of developing both a short form of the COPA and providing teachers with meaningful, actionable results from that measure.

Cognitive Diagnostic Modeling: the Deterministic Inputs, Noisy “and” Gate (DINA) Model

There are several models that can be used with CDM (Rupp et al., 2010). The COPA-LF was previously validated using the deterministic inputs, noisy and gate (DINA) model (Chen et al., 2022). The COPA-LF comprises items that theoretically measure a single attribute, which is the “simple structure” of the Q-matrix in the CDM framework. The simple structure indicates that when individuals master one attribute of an item, they can answer an item successfully, which is appropriate for the DINA model. Because the COPA-SF has the same structure, we decided to use the DINA model again and to allow for a parallel comparison between both forms. Additionally, it can provide better classification rates, especially when the sample size is small—a common condition in a school context (de la Torre et al., 2015).

In the DINA model, latent response (η_{ij}) to item j for student i is a binary indicator where 1 represents mastery and zero non-mastery and is defined deterministically by $\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}$ where α_{ik} represents the possession of attribute k for student i , q_{jk} is the requirement of attribute k for an item j , and K is the total number of attributes (de la Torre, 2009). The DINA model is one of the simplest CDM models that requires the estimation of only two parameters for each item, the guessing (g_j) and slipping (s_j) parameters. The guessing parameter represents the probability of a correct response X_{ij} for student i who does not possess any or all of the required skills for an item j , denoted $g_j = P(X_{ij} = 1 | \eta_{ij} = 0)$ and the slipping is the probability of an incorrect response for those who have possessed all required attributes for an item j , denoted $s_j = P(X_{ij} = 0 | \eta_{ij} = 1)$.

In addition to these two item parameters in the DINA model, de la Torre (2009) suggests an item discrimination index (IDI; δ_j) that considers both the slipping and guessing parameters, which

is $\delta_j = 1 - s_j - g_j$. The discrimination index is maximized as the slipping and guessing parameters approach zero and can be used as an index of item quality. The higher IDI value the item has, the better quality that item possesses.

Nonparametric Cognitive Diagnostic Modeling

To deal with a small sample size under the CDM framework, a nonparametric classification approach is a promising solution because it demonstrates a highly accurate classification rate in a small sample size (Chiu & Douglas, 2013; Chiu et al., 2018). A nonparametric classification approach does not require any assumption of a model, and it does not have a procedure to estimate item parameters of any CDM models to classify individuals. Instead, it classifies individuals directly by minimizing the distance between observed item-response vectors and the ideal item-response vectors derived from each possible proficiency profile. For example, suppose there are three skills of orthographic processing that yield eight possible proficiency profiles (2^3). In a nonparametric approach, the distance between an individual item-response vector and each of the ideal item-response vector from the eight possible proficiency profiles will be calculated. When the smallest distance for a proficiency profile is determined, an individual will be assigned to this proficiency profile.

To calculate the distance between an individual item-response vector and an ideal item-response vector, a Hamming distance method was utilized in this study (Chiu & Douglas, 2013). For example, if an individual item-response vector for three items is (0,1,1) and an ideal item-response vector is (0,0,1), then the distance is $|0-0|+|1-0|+|1-1| = 1$. The value of 1, therefore, represents the distance between an individual item-response vector and an ideal item-response vector for a proficiency profile. If an individual has more than one proficiency profile with the same smallest distance, one of the proficiency profiles will be randomly assigned to an individual (Zheng & Chiu, 2019). More details about the distance calculation can be found in the Analysis section of this paper.

The Present Study

In the current study, we created and provided initial evidence of validity of a short form, 30-item COPA (COPA-SF) that can be quickly administered by teachers and psychologists in less than 10 minutes. Our research questions were:

1. What is the dimensionality of the COPA-SF and does it represent the same five skills as the COPA-LF?
2. What are the learning sequence and mastery probabilities of the COPA-SF and are they similar to the COPA-LF?
3. What are the correlations between mastery probabilities of skills (i.e., rapid perception, access, differentiation, correction, and arrangement) and external measures of word recognition, pseudoword reading, phonological processing, spelling ability, orthographic knowledge, and oral language?
4. Does the COPA-SF maintain the high classification accuracy and consistency of the COPA-LF?

Method

Participants

Participants were 74 second-grade students (51% boys) from two schools in one suburban city and one urban city located in northern California. The mean age in years at the time of data collection

was 8.03 years ($SD = .33$), and the majority of the students were White, non-Hispanic (68%). The remaining participants were Asian (15%), African American (6%), Hispanic/Latino (4%), or multiracial (3%), with 4% unreported. All participants completed the COPA-LF, and all but three participants completed the assessments for orthographic knowledge, oral vocabulary, word recognition, pseudoword reading, spelling, and phonological awareness. The participants' performances on these assessments are summarized in [Table S1](#) (see [Supplemental Materials](#)). The study was approved by the IRB at the first authors' affiliation. Three weeks before data collection, parent consent was obtained and a child participant's verbal consent was obtained prior to testing.

Measures

Orthographic Processing. The COPA-LF ([Chen et al., 2020](#)) was used to assess students' orthographic processing ability. The computer-based assessment includes 60 test items and 10 practice items. The first 51 items require a fixed response, while the last nine items are open-ended.

Items 1–10 were designed to measure ability to count the number of letters among a group of symbols within 5 seconds (e.g., *gšcykh*), representing *rapid perception* skill. Items 11–20 measure *access* ability by having students identify a letter combination among four options after seeing the target letter combination for 3 seconds followed by a 4 second break (e.g., *tplkmn*, *tpikmn*, *tplhmn*, and *tplkwn*). Items 21–31 measure ability to identify the permissible combination among three other impermissible combinations (e.g., *tiught*, *cuught*, *tought*, *tyught*, and *cyught*), representing *differentiation* ability. Items 32–41 measure ability to *correct* a spelling by adding a letter. Items 42–51 measure ability to *correct* a spelling by removing a letter. Lastly, items 52–60 measure *arrangement* ability by asking students to rearrange letters to create a real word. The Cronbach's α is .86, suggesting good reliability. [Table 1](#) summarizes how these items link to the skills and provide corresponding definitions and related item features. A binary scoring system (0 = incorrect, 1 = correct) was employed to score the participants' responses. A detailed description of the item design can be found in [Supplemental Materials](#).

Orthographic Knowledge. To measure students' orthographic knowledge, the non-normed Orthographic Choice Task ([Olson et al., 1985](#)) was used. This task included 23 items and was designed to assess children's ability to pick the correct spelling between two homophones (e.g., *rain* vs. *rane*) in English. Students were asked to circle the correct spelling on paper. Responses were scored as either correct or incorrect. The Cronbach's alpha is .83. However, it should be noted that orthographic knowledge is different but related to orthographic processing. Orthographic processing represents procedural operations, whereas orthographic knowledge is declarative knowledge processing ([Burt, 2006](#); [Castles & Nation, 2006](#); [Chen et al., 2020](#)).

Oral Receptive Vocabulary. The Peabody Picture Vocabulary Test-Fourth Edition (PPVT; [Dunn & Dunn, 2012](#)) was used to assess students' receptive oral vocabulary. Students were asked to point to the picture that best described an orally provided vocabulary word. The test developers report test–retest reliabilities ranged from .92 to .96, and the split-half reliability ranged from .89 to .97.

Word Recognition. The Letter and Word Identification subtest of the Woodcock-Johnson III (WJ-III; [Woodcock et al., 2007](#)) was used to measure students' word recognition ability. The median reliability of this test is .94 ([McGrew et al., 2007](#)).

Pseudoword Reading. This was assessed with the Word Attack subset of the WJ-III Test of Achievement. This subtest assesses children's ability to read aloud pseudowords, such as *gusy* or *sluke*. The median reliability of the test is .87 ([McGrew et al., 2007](#)).

Table 1. Summary of COPA-LF.

Items	Skill	Definition	Response Type	Time Limitation	Example
1–10	Rapid perception	The ability to identify letters within a very short time	Fixed	5 seconds to respond	g\$cykh
11–20	Access: Storing and retrieval	The ability to store letter combinations and then retrieve them from short term memory	Fixed	3 seconds to see stimuli 4 seconds to pause unlimited time to respond	tplkmn tpikmn tplhmn tplkwn
21–31	Differentiation	The ability to differentiate permissible and impermissible letter combinations	Fixed	Unlimited time	tiught cuught tought cyught
32–41	Correction: Addition	The ability to add a letter to make spelling correct	Fixed	Unlimited time	detal: l h i o
42–51	Correction: Removal	The ability to delete a letter to make a spelling correct	Fixed	Unlimited time	lauyout: y u o a
51–60	Arrangement	The ability to arrange letter to make a real word	Open	Unlimited time	a, c, e, e, r, t -----

Spelling. The spelling subtest of the WJ-III was used to measure students' spelling ability. This paper-and-pencil subtest was designed to assess children's ability to spell orally presented words used in a sentence. The median reliability of the test is .90 (McGrew et al., 2007).

Phonological Awareness. The Elision and Blending subtests of the Comprehensive Test of Phonological Processing (CTOPP; Wagner et al., 1999) were used to assess participants' phonological awareness. The Elision reliability ranged from .89 to .91. The Blending reliability ranged from .79 to .87 (Wagner et al., 1999).

Statistical Analyses

Test Development. To select items for the short-form COPA (COPA-SF), the DINA model and the simple six-skill Q-matrix (i.e., an item requires only one skill) were used to obtain IDI values as an index of item quality. A six-skill Q-matrix was used because the original development of the 60-item COPA included six subtests, including rapid perception, access, differentiation, addition correction, removal correction, and arrangement subtests. Five items with the highest IDI values in each subtest were selected for the COPA-SF. The minimum IDI for item selection was .20 (Chen et al., 2019; Lee et al., 2012).

Dimensionality Models. After constructing the 30-item COPA-SF (i.e., 5 items for each of the 6 skills), three models were analyzed: one-skill, five-skill, and six-skill models. We choose these models because they were validated with the COPA-LF (Chen et al., 2022). A one-skill model assumes that only one general orthographic processing skill measures students' performance across all test items. In the Q-matrix, the one-skill model assumes one skill is measured by all 30 test items. The five-skill model comprises five skills, including rapid perception (items 1–5),

access (items 6–10), differentiation (items 11–15), correction (items 16–25), and arrangement (items 26–30). The six-skill model was constructed to test if the correction skill is best measured by two subcategories: addition and removal.

Linear Attribute Hierarchy Models. After confirming the best-fitting dimensionality model, four linear skill hierarchy models (H1-H4) were further explored and compared with the non-hierarchical model (H0). Hierarchical models were used to assume the various learning sequences of skills in this study. Hierarchical model 1 (H1) assumes the correction skill is a prerequisite of the arrangement skill (A4→A5). Hierarchical model 2 (H2) adds an assumption to H1, that is, differentiation is a prerequisite of correction (A3→A4→A5). Hierarchical model 3 (H3) adds one more assumption to H2, that is, access is a prerequisite of differentiation (A2→A3→A4→A5). Hierarchical model 4 (H4) adds one more assumption to H3, that is, rapid perception is a prerequisite of access (A1→A2→A3→A4→A5). These models were considered based on the information processing theory: more sophisticated skills (e.g., arrangement) must rely on mastery of the foundational skills (e.g., correction) (Chen et al., 2020). As a result, it was assumed these skills were developed based on a specific sequence: rapid perception, access, differentiation, correction, and arrangement.

CDM Analyses and Evaluation Criteria. In this study, the dimensionality of the COPA-SF was examined first, followed by linear attribute hierarchy model comparisons. After obtaining the best-fitting model of dimensionality along with a linear hierarchy model, the final CDM analysis was conducted for skill diagnostic information for individuals and the whole group. All CDM analyses were conducted using parametric DINA and nonparametric DINA models.

All parametric CDM analyses with the DINA model were implemented by the “CDM” R package (Robitzsch et al., 2020) using marginal maximum likelihood. To investigate model fit, Akaike’s information criterion (AIC; Akaike, 1987), Bayesian information criterion (BIC; Schwarz, 1978), MAX (χ^2) (Robitzsch et al., 2020), and the root mean square error approximation (RMSEA) were used to choose the best-fitting model. Additionally, the likelihood ratio (LR) test was used for model comparison.

Lower values of AIC or BIC indicate a better model fit. MAX (χ^2) is an absolute fit index that can be tested for statistical significance. A significant p -value indicates that the statistical independence of the item pair is violated, and thus the model does not fit the data well (Hu et al., 2016). RMSEA values below .05 indicate good fit, values below .10 indicate moderate fit, and values above .10 indicate poor fit (Kunina-Habenicht et al., 2009). The likelihood ratio (LR) test was used for model comparison. A significant p -value of the LR test favors a more complex model, whereas a non-significant p -value favors a simpler model, based on the parsimonious rule. For example, when comparing the non-hierarchical model (H0) and the hierarchical model 1 (H1: correction skill→arrangement), the non-hierarchical model is a more complex model because it has more estimated parameters compared to hierarchical model.

Due to the small sample size ($N = 74$) of this study, we applied a nonparametric cognitive diagnosis method to provide further evidence for choosing the best-fitting attribute hierarchical model by calculating the average loss function values, an approach that is not affected by small sample size. The nonparametric analyses were conducted using the “NPCD” R package (Zheng & Chiu, 2019) and Hamming distance estimation. The average value of a loss matrix across all individuals with the smallest loss function values from the NPCD package was computed for each attribute hierarchical model and compared to select the best-fitting model. The loss matrix provides the values of the distance between the observed and ideal item-response vectors using the Hamming distance estimation. In the loss matrix, rows

represent candidate attribute profiles, and columns represent individuals (Zheng & Chiu, 2019). The values in a matrix represent the distances between the observed and ideal item-response vectors with a weight assigned to guessing and a weight for slipping. Using the Hamming distance estimation, weights for guessing and slipping parameters were 1s (Zheng & Chiu, 2019). To compute the average loss function value for each attribute hierarchical model, we followed these steps: (1) we calculated the distances between the observed and ideal item-response vectors for all candidate attribute profiles across all individuals, (2) for each individual, we assigned the candidate attribute profile with the smallest loss function value as their attribute profile (Zheng & Chiu, 2019), and (3) we then calculated the average of these smallest values across all individuals for each attribute hierarchy. The attribute hierarchical model with the smallest average loss value was considered the best-fitting model. This is because the smallest average loss value indicates the least prediction error between the observed data and the hypothesized model, suggesting a good fit between the data and the model.

For linear attribute hierarchy model comparisons, we revised the AlphaNP function in the NPCD package because the original function is not designed for the DINA model with a linear hierarchical structure. We modified it by revising the number of attribute profiles based on the feature of the hierarchal structure. The R code is available upon request.

For the final analysis with the best-fitting hierarchical model, a Bayesian maximum a posteriori estimation with the parametric DINA model was used to obtain all diagnostic information, including individuals' skill patterns and profiles, skill mastery probabilities, and proportions of skill patterns. Classification accuracy and consistency based on estimators of Johnson and Sinharay (2018) and simulation-based estimation (Robitzsch et al., 2020) were computed. The best-fitting hierarchical model was also used to obtain the diagnostic information with nonparametric CDM approaches. The diagnostic information from parametric and nonparametric CDM models was compared by computing the profile agreement rate and the attribute agreement rate (Chiu & Douglas, 2013). The profile agreement rate represents the agreement of all skills as a whole, and the attribute agreement rate denotes the agreement of individual skills within a diagnostic profile. The high agreement rate indicates the valid diagnostic information from the parametric DINA model with a small sample.

Correlation Analyses. Pearson product-moment correlations were computed between each attribute mastery probability and each external measure (i.e., orthographic knowledge, oral receptive vocabulary, letter and word recognition, pseudoword reading, spelling, and phonological awareness) to examine criterion-related validity.

Results

The Dimensionality of the COPA-SF

Table 2 reports model fit for the one-skill, five-skill, and six-skill models for the COPA-SF. The absolute fit index, Max (χ^2), indicates that the one-skill model did not fit the data well ($p < .01$), but both the five-skill and six-skill models fit the data well ($p = .71$ and $.68$, respectively). AIC and BIC were lowest for the five-skill model. To assess model fit for each attribute hierarchical model, we provided the mean of RMSEA across items. All of the models had a mean RMSEA above .10. Given that our sample size is small, the values of RMSEA might not be sensitive to reflect item fit across models.

Comparing models, the LR test demonstrated that the five-skill and six-skill models fit significantly better than the one-skill model, $\chi^2(30) = 185, p < .001$ and $\chi^2(62) = 206, p < .001$,

Table 2. Fit Statistics for One-Skill, Five-Skill, and Six-Skill Non-Hierarchical Models.

Models	Number of Attribute Profiles	Max (χ^2), p	AIC	BIC	Mean of RMSEA	Loss	Model Comparison	χ^2	df	p-Value
One-skill model	2	21.29, $p < .01$	2566	2706	.17	17.82	One versus five skills	185	30	<.001
Five-skill model	32	9.91, $p = .71$	2441	2651	.17	5.90	One versus six skills	206	62	<.001
Six-skill model	64	10.00, $p = .68$	2484	2767	.18	5.67	Five versus six skills	21	32	= .93

Note. Loss = Average value for the lose function based on nonparametric approach with the hamming method.

respectively. However, comparing the five-skill model and the six-skill model, the result was not statistically significant ($\chi^2(32) = 21.21, p = .93$), suggesting the simpler model (i.e., the five-skill model) was the best-fitting model. Although the six-skill model showed the smallest average loss value (Loss = 5.67), the difference between the average loss values of the five-skill and six-skill models were very small. As most results of model fit favored the five-skill model, we decided the five-skill model as the final model to examine its hierarchical structure.

Hierarchical Models: Learning Sequences of Five Attributes

Table 3 shows model fit for four linear attribute hierarchy models and the baseline model (i.e., non-hierarchical model) for the COPA-SF. Max (χ^2) did not show statistically significant results at $p = .05$ level, indicating all of the models fit the data well. Based on AIC and BIC, H2 (A3→A4→A5) was the best model with the lowest value of AIC, whereas H3 (A2→A3→A4→A5) was best with the lowest BIC. Similar to the results of validating dimensionality, the mean of the RMSEA for H0-H3 were above .10, except for H4.

We further conducted model comparisons among the baseline model and four competitive hierarchical models (see Table 4). When comparing the non-hierarchical model (H0: the most complex model) with all linear hierarchical models (i.e., the simpler models with fewer estimated parameters), most LR tests showed statistically non-significant results, except for H4 (Perception → Access → Differentiation → Correction → Arrangement). Note that the χ^2 value was negative when comparing H0 and H1. Thus, the meaning of the comparison could not be interpreted. These results suggested that H2–H3 models fit the data better than H0. Based on the parsimonious rule, models H2, and H3 were preferable to H0 and H4. Results from the nonparametric approach showed that the average loss value among H0–H2 was small (Table 3). H4 showed the largest loss value (Loss = 8.19), suggesting that H4 did not fit the data well. Combining this with the results from the parametric and nonparametric approaches, H2 and H3 represented the best hierarchical models for the data.

As the model comparison between H0 and H1 could not be interpreted, we further compared H2 and H3 with H1. The LR tests showed statistically non-significant results when comparing H1 and H2 and H1 and H3 models. This suggests that the H2 and H3 models fit the data better than H1. Lastly, we compared H2 and H3 models to decide on the final model. The results were statistically significant ($\chi^2(6) = 17.09, p < .05$), suggesting the H2 model was better.

Because of inconsistent findings based on various fit indices and small difference between the average of the loss value between H2 and H3, we took classification consistency and classification accuracy of the H2 and H3 models into consideration. The H2 model yielded

Table 3. Fit Statistics for Five-Skill Non-Hierarchical and Five-Skill Hierarchical Models.

Models	Numbers of		AIC	BIC	Mean of	
	Attribute Profiles	Max (χ^2), p			RMSEA	Loss
H0: Non-hierarchical	32	9.91, $p = .71$	2441	2651	.17	5.90
H1: Correction (A4) → Arrangement (A5)	24	9.95, $p = .70$	2425	2616	.17	5.91
H2: Differentiation (A3) → Correction (A4) → Arrangement (A5)	16	9.96, $p = .70$	2410	2583	.16	5.93
H3: Access (A2) → Differentiation (A3) → Correction (A4) → Arrangement (A5)	10	8.03, $p = 1.00$	2415	2574	.12	6.64
H4: Perception (A1) → Access (A2) → Differentiation (A3) → Correction (A4) → Arrangement (A5)	6	10.02, $p = .68$	2472	2621	.05	8.19

Note. Loss = Average value for the lose matrix based on a nonparametric approach with the hamming method.

Table 4. Model Comparisons Among the Non-Hierarchical and Hierarchical Models.

Model	(Number of Estimated Parameter)		Model	(Number of Estimated Parameter)	χ^2	df	p
H0	(31)	versus	H1	(23)	-.06 ^a	8	1
H0	(31)	versus	H2	(15)	.80	16	1
H0	(31)	versus	H3	(9)	17.90	22	.07
H0	(31)	versus	H4	(5)	82.73	26	<.001
H1	(23)	versus	H2	(15)	.86	8	1
H1	(23)	versus	H3	(9)	17.95	14	.21
H2	(15)	versus	H3	(9)	17.09	6	.008

Note. H0 = Non-hierarchical; H1 = A4→A5; H2 = A3→A4→A5; H3 = A2→A3→A4→A5.

H4 = A1→A2→A3→A4→A5. Numbers in the parenthesis are the number of class parameters. Model in bold is the better model in each pair of comparisons.

^aGiven that the χ^2 value was negative, the meaning of the comparison between H0 and H1 could not be interpreted.

classification accuracy of .88 using the estimator of Johnson and Sinharay (2018) and .87 based on simulated data and classification consistency of .80 and .77, respectively. In contrast, the H3 model had higher levels of classification accuracy (e.g., .90 and .91) and classification consistency (e.g., .86 and .84). Given the better classification accuracy and consistency of H3 compared to H2, H3 (access (A2)→differentiation (A3)→correction (A4)→arrangement (A5)) appears to be the best model for the COPA-SF.

Estimated Item Parameters and Item Statistics in the Five-Skill Model with the Hierarchical Structure

Table 5 presents estimated item parameters and item statistics for the five-skill model with the H3 structure (A2→A3→A4→A5). The average of the guessing parameters was .35. Items 11 and 13 of the differentiation skill block showed the highest guessing parameters. The average of the slipping parameters was .21. Item 28 in the arrangement skill showed the highest slipping parameter. The average RMSEA value was .12. Among the 30 items, 12 items showed that the RMSEA values were below .10. The average of the IDI values was .44. Item 12 and 14 in the

differentiation skill block had IDI values lower than .2. In general, most items were of high quality based on the corresponding IDI values greater than .2. Results suggested that the easiest items were in the differentiation and correction skill blocks. Among 30 items, four out of five items in the differentiation skill block had proportion of correct responses above .80. Two out of eight items in the correction skills had above .80% of correct responses. Item 28 and Item 30 in the arrangement skill block had correct response rates equal or less than .20, suggesting these two items were the most difficult items on the COPA-SF. The average of corrected item-total correlations was .38.

Table 5. Estimated Item Parameters and Item Statistics in the Five-Skill Model With the Hierarchical Structure.

Skill	Item	Labeled Number of an Item in the Long Format	Guessing Parameter (SE)	Slipping Parameter (SE)	RMSEA	IDI	The Proportion of Correct Responses (%)	Corrected Item-Total Correlation
Rapid perception	1	1	.09 (.01)	.18 (.27)	.12	.76	.47	.36
	2	2	.09 (.01)	.10 (.95)	.08	.81	.53	.42
	3	4	.00 (.02)	.25 (.79)	.09	.66	.45	.32
	4	5	.00 (.00)	.30 (.42)	.08	.70	.38	.42
	5	6	.16 (.00)	.15 (.86)	.07	.85	.46	.53
Access	6	11	.31 (.03)	.35 (.11)	.20	.50	.53	.42
	7	12	.28 (.05)	.40 (.13)	.11	.29	.53	.21
	8	14	.16 (.05)	.32 (.11)	.17	.40	.58	.33
	9	15	.00 (.03)	.35 (.11)	.25	.50	.53	.17
	10	17	.70 (.00)	.18 (.06)	.09	.82	.61	.45
Differentiation	11	23	.76 (.07)	.05 (.04)	.16	.25	.84	.40
	12	24	.54 (.06)	.10 (.07)	.13	.13	.84	.28
	13	27	.75 (.07)	.03 (.01)	.18	.43	.78	.38
	14	29	.65 (.06)	.16 (.14)	.12	.08	.80	.26
	15	30	.59 (.07)	.06 (.05)	.13	.28	.81	.42
Correction	16	33	.32 (.07)	.10 (.07)	.07	.31	.76	.36
	17	36	.32 (.05)	.06 (.04)	.11	.62	.65	.49
	18	37	.68 (.05)	.36 (.17)	.12	.32	.49	.36
	19	39	.71 (.07)	.05 (.04)	.09	.28	.82	.40
	20	41	.52 (.07)	.00 (.00)	.10	.29	.86	.54
	21	42	.34 (.06)	.21 (.14)	.17	.27	.66	.33
	22	44	.46 (.05)	.06 (.14)	.06	.61	.66	.47
	23	48	.37 (.06)	.18 (.08)	.17	.36	.65	.41
	24	49	.42 (.06)	.23 (.12)	.08	.40	.58	.39
	25	50	.47 (.06)	.03 (.02)	.11	.55	.72	.53
Arrangement	26	53	.34 (.07)	.13 (.07)	.12	.40	.64	.43
	27	54	.00 (.07)	.28 (.09)	.17	.39	.50	.27
	28	56	.28 (.00)	.65 (.10)	.02	.35	.15	.42
	29	57	.00 (.06)	.52 (.10)	.09	.20	.36	.23
	30	59	.09 (.00)	.52 (.10)	.04	.48	.20	.47
Average			.35 (.04)	.21 (.18)	.12	.44	.47	.38

Note. IDI means item discrimination index.

Mastery Probabilities and Classification Accuracy and Consistency

The five-skill model with the H3 hierarchy model was used to obtain all diagnostic information. Based on Bayesian maximum a posteriori estimation, the proportions of mastery of the five COPA-SF skills (i.e., rapid perception, access, differentiation, correction, and arrangement) were .54, .74, .56, .53, and .42, respectively, which showed a decreasing pattern. This result also supported the H3 hierarchical model (A2→A3→A4→A5). As mentioned earlier, with two estimators (i.e., Johnson & Sinharay, 2018 and simulated data), classification accuracy rates were above .90 and classification consistency rates were around .85. For the five individual skills, classification accuracy rates were .95 or higher, and classification consistency rates were .93 or higher. Regarding classification consistency between the DINA and non-parametric DINA models, there were 56% of the profile agreement rate and 97% of the attribute agreement rate.

Correlations with External Measures

Table 6 showed the correlations between the mastery probabilities of the five skills in the H3 hierarchical model and the six external measures (i.e., word recognition, pseudoword reading, elision, blending, spelling ability, orthographic knowledge, and oral language; see the lower triangular). Overall, the magnitude of correlations with external measures were similar to the COPA-LF (see the upper triangular in Table 6). Like the COPA-LF, the correlation between the rapid perception skill and oral language was the weakest ($r = .05$) in the COPA-SF. Word recognition, pseudoword reading, and spelling had the strongest correlations with all skills in the COPA-SF (r ranging from .53 to .65). The rapid perception skill had the strongest correlation with blending ($r = .36$). The access skill had the strongest correlation with word recognition ($r = .45$). Word recognition correlated more strongly with three skills in the COPA-SF (i.e., differentiation ($r = .65$), correction ($r = .64$), and arrangement ($r = .63$)) than external measures.

Table 6. Correlations Between the Five Skills of the COPA-SF With External Measures in the Lower Triangular and Correlations Between the Five Skills of the COPA-LF With External Measures in the Upper Triangular.

	1	2	3	4	5	6	7	8	9	10	11	12
1. Rapid perception	-	.29	.21	.28	.51	.22	.26	.31	.38	.23	.13	.06
2. Access	.22	-	.50	.56	.89	.51	.47	.32	.34	.51	.38	.33
3. Differentiation	.38	.70	-	.93	.49	.49	.56	.40	.31	.45	.49	.28
4. Correction	.33	.66	.97	-	.54	.51	.53	.45	.36	.46	.51	.32
5. Arrangement	.55	.55	.82	.85	-	.53	.56	.41	.38	.54	.32	.34
6. Word recognition	.23	.45	.65	.64	.63	-	.73	.51	.41	.74	.56	.46
7. Pseudoword reading	.26	.32	.47	.44	.53	.73	-	.55	.46	.61	.46	.36
8. Elision	.32	.16	.37	.38	.42	.51	.55	-	.40	.48	.25	.31
9. Blending	.36	.42	.45	.44	.38	.41	.46	.40	-	.34	.01	.39
10. Spelling	.26	.37	.53	.55	.57	.74	.61	.48	.34	-	.47	.34
11. Orthographic knowledge	.13	.43	.36	.34	.29	.56	.46	.25	.01	.47	-	.12
12. Oral language	.05	.39	.42	.44	.39	.46	.36	.31	.39	.34	.12	-

Note. The magnitude of correlations large than .23 is significant at the 5% level.

Discussion

The results of the current study found that the five-skill model with the H3 hierarchy (access → differentiation → correction → arrangement) had the best model fit for the COPA-SF. This finding is similar to the results found in [Chen et al. \(2022\)](#) study with the 60-item COPA-LF. Consistent with COPA-LF, COPA-SF reflected five dimensions, including rapid perception, access, differentiation, correction, and arrangement.

The mastery probabilities and learning sequence found in COPA-SF are consistent with those found in COPA-LF. The mastery probabilities of the five attributes in the COPA-SF, are .54, .74, .56, .53, and .42. These estimates are consistent with the mastery probabilities of the five attributes in COPA-LF, which are .55, .87, .77, .77, and .51. Both COPA-SF and COPA-LF reflect a learning progression among four attributes, that is, access → differentiation → correction → arrangement. The learning sequence is slightly different from the sequence proposed by [Chen et al. \(2020\)](#), that is, rapid perception → access → differentiation → correction → arrangement. The discrepancy between the proposed learning sequence in [Chen et al. \(2020, 2022\)](#) and this study could be due to imposing a five-second time-out rule on the rapid perception items. The speeded nature of the perception items made the rapid perception items more difficult for second-grade students than expected. Nevertheless, the COPA-SF successfully demonstrated the identical learning progression discovered in the COPA-LF ([Chen et al., 2022](#)).

As we purposely selected the items with the highest IDI values for each subtest to form the COPA-SF, most items in the COPA-SF have above .20 IDI values with an average of .44, suggesting all items have high quality. Only Items 12 and 14 (Items 24 and 29 in the original COPA-LF) and item 14 (Item 29 in the COPA-LF) showed the IDI values were below .20. However, the IDI values for these two items were .32 and .34 in the COPA-LF, respectively ([Chen et al., 2022](#)). The lower IDI values of Items 12 and 14 in the COPA-SF might be associated with the substantial reduction of the number of items.

In [Chen et al. \(2022\)](#) study, the COPA-LF has classification accuracy rates of .91 using the estimator by [Johnson and Sinharay \(2018\)](#) and at .97 using simulated data randomly drawn from 5000 observations from the sample. The COPA-LF has classification consistency rates of .90 using the estimator by [Johnson and Sinharay \(2018\)](#) and of .91 using simulated data. For five individual skills, accuracy rates were .93 or higher, and classification consistency rates were .91 or higher. With half of the items, the COPA-SF has the capacity to demonstrate comparable accuracy and consistency as the COPA-LF. This finding signifies the utility of the COPA-SF since it requires half of the time of the COPA-LF while maintaining consistency and accuracy when determining second-grade students' mastery status of orthographic processing skills.

Previous research demonstrated that the COPA-LF has small to moderate correlations with word recognition, pseudoword reading, phonological processing, spelling ability, orthographic knowledge, and oral language ([Chen et al., 2020, 2022](#)). In the COPA-LF, [Chen et al. \(2022\)](#) found that the correlations between the COPA-LF and these external variables ranged from .06 to .56. The present study showed that the COPA-SF has a similar degree of correlation with these external variables, ranging from .05 to .65. Moreover, because orthographic processing is a critical skill for word recognition ([Cunningham, 2006](#); [Stanovich & West, 1989](#)) and spelling ([Chen et al., 2021](#); [Zarić et al., 2021](#)), the COPA-LF was expected to more strongly correlated with external measures involving processing letters and print ([Chen et al., 2020](#)), which it did. Similarly, an essential piece of evidence for validating the COPA-SF was that it should have a stronger correlation with spelling, word recognition, pseudoword reading, and orthographic knowledge than oral language, which it did, providing another piece of concurrent validity evidence.

The average correlation between the COPA-SF and orthographic knowledge (average $r = .31$) was unexpected. As the assessment of the non-normed Orthographic Choice Task (Olson et al., 1985) measures students' ability to choose the correct spelling among two options (e.g., *rain* vs. *rane*), we expected the correlation between the COPA-SF and orthographic knowledge would be close to the correlation between the COPA-SF and spelling (average $r = .42$), but it was noticeably lower. One possible explanation for this unexpected finding is that Olson et al.'s (1985) orthographic choice task might not tap the spelling recognition ability for which it was originally designed. This potential explanation is further supported by the difference in the degree of correlation between spelling and word recognition ($r = .74$) and between spelling and orthographic knowledge ($r = .47$).

The correlations among the five skills ranged from moderate to high. Specifically, the correlation between the differentiation and correction subscales was exceptionally high ($r = .97$). This pattern was also observed in the COPA-LF, where the correlation was .93 (as shown in Table 6). In response to these results, we conducted additional analyses by fitting the data to a four-skill model and comparing it with the original five-skill model. The results indicated that the four-skill model fit the data better than the five-skill model (see Table S2 in Supplemental Materials). However, from a pedagogical perspective, differentiation and correction skills should be considered as separate skills, as they measure different aspects of ability. The differentiation skill assesses the ability to distinguish between permissible and impermissible combinations, while the correction skill evaluates the ability to produce correct spelling. These two skills provide distinct diagnostic information to researchers and practitioners, especially when teaching young children to read and spell (Chen et al., 2020), so we opted to keep both subscales in the measure.

However, this high correlation prompted us to formulate a new hypothesis about the development of differentiation and correction skills. The strong correlation implies that young children likely develop these two skills simultaneously. Although these skills are distinct in their pedagogical application, analyzing only the correctness of the answers may not reveal their differences. To test this hypothesis, future studies should employ methods like eye-tracking, response time, or brain imaging to examine the differences and similarities between these skills.

In this study, we used the DINA model with parametric (i.e., maximum likelihood (ML) estimation) and nonparametric approaches to examine the dimensionality and the hierarchical structure of the COPA-SF. The RMSEA has been widely used in parametric approaches with the ML estimation to assess item fit and select the best-fitting model, particularly for larger samples ($N > 200$) (Kunina-Habenicht et al., 2009; Liu et al., 2016). However, our analysis revealed that most RMSEA values exceeded .10 when examining the dimensionality and hierarchical models for our data. This outcome may be attributed to the small sample size in our study. To our knowledge, no research has comprehensively evaluated the performance of fit indices for assessing model and item fit in small samples. Given that sample sizes below 100 are common in educational settings, future research should investigate the effectiveness of fit indices under small sample conditions.

In the nonparametric cognitive diagnosis approach, the average loss value index clearly demonstrated that the H4 hierarchical model (Perception (A1) → Access (A2) → Differentiation (A3) → Correction (A4) → Arrangement (A5)) had the highest average loss value, suggesting this model did not fit the data well. However, the average loss values did not show large differences for the rest of the models. The findings were in line with previous studies that found that model-fit indices (e.g., RMSEA under the ML estimation and the posterior predictive probability under the Bayesian estimation) showed slight differences in different attribute hierarchical models using ML and Bayesian estimations (Hsu et al., 2023; Liu, 2018). As model-fit indices have not been systematically examined for selecting the best-fitting attribute hierarchical model, more studies are warranted.

Limitation and Future Study

There are several limitations associated with the present study that may lead to future research. First, we only focused on second-grade students, which might limit the generalizability of the findings of the current study. Future studies should seek to validate the COPA-SF with participants in additional grade levels. Second, this study did not examine whether the COPA-SF has biased items. For example, we did not analyze if there was an impact of home language or cultural differences on performance. Thus, future research could focus on conducting measurement invariance and differential item functioning analysis on the COPA-SF. Third, this study thoughtfully selected the CDM framework to validate the COPA-SF, leveraging its unique strengths in assessing fine-grained cognitive skills. However, future research could complement our CDM-based findings by incorporating IRT and CFA approaches (e.g., [Chen et al., 2023](#)), potentially enhancing the comprehensive understanding of the COPA-SF's psychometric properties and further expanding its utility in various practical contexts. Lastly, a computerized adaptive version of the COPA-SF might have the potential to reduce response time and cover a broader range of difficulty levels in a shorter amount of time than the shortened version presented here. Thus, developing a computerized adaptive test of the COPA-SF could be a beneficial avenue for future research.

Conclusion

The COPA-SF is an innovative, time-efficient, multidimensional measure of orthographic processing. The COPA-SF was found to have similarly promising psychometric properties as the COPA-LF. Both the COPA-LF and the COPA-SF measure five skills and reflect a linear progression among four skills (i.e., access → differentiation → correction → arrangement). The COPA-SF has equivalent consistency and accuracy when determining participants' mastery of orthographic processing skills. The COPA-SF has better quality items than the COPA-LF. Thus, the COPA-SF is a reliable and effective tool for that can be used by researchers and practitioners to measure students' orthographic processing skills in classroom settings and clinical practices.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Yi-Jui I. Chen  <https://orcid.org/0000-0001-9845-5764>

Yi-Jhen Wu  <https://orcid.org/0000-0002-0038-9234>

Yi-Hsin Chen  <https://orcid.org/0000-0002-6102-7544>

Supplemental Material

Supplemental material for this article is available online.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52(3), 317–332. <https://doi.org/10.1007/BF02294359>

- Allen, V., Rahman, N., Weissman, A., MacCann, C., Lewis, C., & Roberts, R. D. (2015). The Situational Test of Emotional Management – Brief (STEM-B): Development and validation using item response theory and latent class analysis. *Personality and Individual Differences, 81*, 195–200. <https://doi.org/10.1016/j.paid.2015.01.053>
- Apel, K., Thomas-Tate, S., Wilson-Fowler, E. B., & Brimo, D. (2011). Acquisition of initial mental graphemic representations by children at risk for literacy development. *Applied PsychoLinguistics, 33*(2), 365–391. <https://doi.org/10.1017/S0142716411000403>
- Austin, E., Saklofske, D., & Smith, M. (2018). Development and validation of two short forms of the Managing the Emotions of Others (MEOS) Scale. *Frontiers in Psychology, 9*, 974. <https://doi.org/10.3389/fpsyg.2018.00974>
- Berninger, V. W., Yates, C., & Lester, K. (1991). Multiple orthographic codes in reading and writing acquisition. *Reading and Writing: An Interdisciplinary Journal, 3*(2), 115–149. <https://doi.org/10.1007/BF00420030>
- Burt, J. S. (2006). What is orthographic processing skill and how does it relate to word identification in reading? *Journal of Research in Reading, 29*(4), 400–417. <https://doi.org/10.1111/j.1467-9817.2006.00315.x>
- Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology, 89*(4), 631–644. <https://doi.org/10.1037/0022-0663.89.4.631>
- Castles, A., & Nation, K. (2006). How does orthographic learning happen? In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (1st ed., pp. 151–179). Psychology Press. <https://doi.org/10.4324/9780203841211>
- Chen, Y.-H., Hsu, C.-L., Wu, Y.-J., Yi, Z., Wang, Y., & Thompson, D. R. (2023). *Exploring attribute hierarchies of the van Hiele theory using diagnostic classification modeling and structural equation modeling*. The University of Chicago School Mathematics Project. <https://ucscmp.uchicago.edu/resources/van-hiele/>
- Chen, Y.-H., Senk, S. L., Thompson, D. R., & Voogt, K. (2019). Examining psychometric properties and level classification of the van Hiele geometry test using CTT and CDM framework. *Journal of Educational Measurement, 54*(4), 733–756. <https://doi.org/10.1111/jedm.12235>
- Chen, Y.-J. I., Chen, Y.-H., Anthony, J. L., & Erazo, N. A. (2022). Evaluation of the computer-based orthographic processing assessment: An application of cognitive diagnostic modeling. *Journal of Psychoeducational Assessment, 40*(2), 271–292. <https://doi.org/10.1177/07342829211056396>
- Chen, Y.-J. I., Cunningham, A. E., Rabe-Hesketh, S., Hinshaw, S. P., & Irey, R. C. (2021). The effect of orthographic neighbors on second-grade students' spelling acquisition. *Reading Research Quarterly, 56*(1), 119–141. <https://doi.org/10.1002/rrq.291>
- Chen, Y.-J. I., Wilson, M., Irey, R. C., & Requa, M. K. (2020). An innovative measure of orthographic processing: Development and initial validation. *Language Testing, 37*(3), 435–452. <https://doi.org/10.1177/0265532220909310>
- Chiu, C.-Y., & Douglas, J. A. (2013). A nonparametric approach to cognitive diagnosis by proximity to ideal response patterns. *Journal of Classification, 30*(2), 225–250. <https://doi.org/10.1007/s00357-013-9132-9>
- Chiu, C.-Y., Sun, Y., & Bian, Y. (2018). Cognitive diagnosis for small educational programs: The general nonparametric classification method. *Psychometrika, 83*(2), 355–375. <https://doi.org/10.1007/s11336-017-9595-4>
- Chung, S. C., Chen, X., & Deacon, S. H. (2018). The relation between orthographic processing and spelling in grade 1 French immersion children. *Journal of Research in Reading, 41*(2), 290–311. <https://doi.org/10.1111/1467-9817.12104>
- Conrad, N. J., & Deacon, S. H. (2016). Children's orthographic knowledge and their word reading skill: Testing bidirectional relations. *Scientific Studies of Reading, 20*(4), 339–347. <https://doi.org/10.1080/10888438.2016.1183128>

- Cummings, D. W. (1988). *American English spelling*. The Johns Hopkins Press.
- Cunningham, A. E. (2006). Accounting for children's orthographic learning while reading text: Do children self-teach? *Journal of Experimental Child Psychology*, 95(1), 56–77. <https://doi.org/10.1016/j.jecp.2006.03.008>
- Deacon, S. H., Chen, X., Luo, Y., & Ramirez, G. (2013). Beyond language borders: Orthographic processing and word reading in Spanish–English bilinguals. *Journal of Research in Reading*, 36(1), 58–74. <https://doi.org/10.1111/j.1467-9817.2011.01490.x>
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34(1), 115–130. <https://doi.org/10.3102/1076998607309474>
- de la Torre, J., van der Ark, L. A., & Rossi, G. (2015). Analysis of clinical data from cognitive diagnosis modeling framework. *Measurement and Evaluation in Counseling and Development*, 51(4), 074817561556911. <https://doi.org/10.1177/0748175615569110>
- Dunn, L. M., & Dunn, D. M. (2012). *Peabody Picture Vocabulary Test (PPVT-4)*. Pearson Education Inc.
- Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, 18(1), 5–21. <https://doi.org/10.1080/10888438.2013.819356>
- Hsu, C.-L., Chen, Y.-H., & Wu, Y.-J. (2023). Using a Bayesian estimation to examine attribute hierarchies of the 2007 TIMSS mathematics test: A demonstration using R packages. *Practical Assessment, Research and Evaluation*, 28(1), 11. <https://doi.org/10.7275/pare.1265>
- Hu, J., Miller, M. D., Huggins-Manley, A. C., & Chen, Y.-H. (2016). Evaluation of model fit in cognitive diagnosis models. *International Journal of Testing*, 16(2), 119–141. <https://doi.org/10.1080/15305058.2015.1133627>
- Johnson, M. S., & Sinharay, S. (2018). Measures of agreement to assess attribute-level classification accuracy and consistency for cognitive diagnostic assessments. *Journal of Educational Measurement*, 55(4), 635–664. <https://doi.org/10.1111/jedm.12196>
- Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2009). A practical illustration of multidimensional diagnostic skills profiling: Comparing results from confirmatory factor analysis and diagnostic classification models. *Studies In Educational Evaluation*, 35(2-3), 64–70. <https://doi.org/10.1016/j.stueduc.2009.10.003>
- Lee, Y.-S., de la Torre, J., & Park, Y. S. (2012). Relationships between cognitive diagnosis, CTT, and IRT indices: An empirical investigation. *Asia Pacific Education Review*, 13(2), 333–345. <https://doi.org/10.1007/s12564-011-9196-3>
- Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoaka's rule-space approach. *Journal of Educational Measurement*, 41(3), 205–237. <https://doi.org/10.1111/j.1745-3984.2004.tb01163.x>
- Liu, R. (2018). Misspecification of attribute structure in diagnostic measurement. *Educational and Psychological Measurement*, 78(4), 605–634. <https://doi.org/10.1177/0013164417702458>
- Liu, Y., Tian, W., & Xin, T. (2016). An application of M2 statistic to evaluate the fit of cognitive diagnostic models. *Journal of Educational and Behavioral Statistics*, 41(1), 3–26. <https://doi.org/10.3102/1076998615621293>
- Lowe, P. A. (2021). The test anxiety measure for college students-short form: Development and examination of its psychometric properties. *Journal of Psychoeducational Assessment*, 39(2), 139–152. <https://doi.org/10.1177/0734282920962947>
- McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Woodcock Johnson III normative update technical manual*. Riverside Publishing.
- Olson, R. K., Kliegl, R., Davidson, B. J., & Foltz, G. (1985). Individual and developmental differences in reading disability. In T. Waller (Ed.), *Reading research: Advances in theory and practice* (4, pp. 1–64). Academic Press.

- Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2020). *CDM: Cognitive Diagnosis Modeling in R. (Version 8.2-6)*. <https://cran.r-project.org/web/packages/CDM>
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523–568. <https://doi.org/10.1037/0033-295x.96.4.523>
- Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, 55(2), 151–226. [https://doi.org/10.1016/0010-0277\(94\)00645-2](https://doi.org/10.1016/0010-0277(94)00645-2)
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 24(4), 402–433. <https://doi.org/10.2307/747605>
- Sturm, A., Kuhfeld, M., Kasari, C., & Mcracken, J. T. (2017). Development and validation of an item response theory-based Social Responsiveness Scale short form. *Journal of Child Psychology and Psychiatry*, 58(9), 1053–1061. <https://doi.org/10.1111/jcpp.12731>
- Sun-Alperin, M. K., & Wang, M. (2011). Cross-language transfer of phonological and orthographic processing skills from Spanish L1 to English L2. *Reading and Writing: An Interdisciplinary Journal*, 24(5), 591–614. <https://doi.org/10.1007/s11145-009-9221-7>
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing (CTOPP)*. Pro-Ed.
- Woodcock, R. W., Mather, N., & McGrew, K. S. (2007). *Woodcock-Johnson III tests of achievement*. Riverside Publishing.
- Zarić, J., Hasselhorn, M., & Nagler, T. (2021). Orthographic knowledge predicts reading and spelling skills over and above general intelligence and phonological awareness. *European Journal of Psychology of Education*, 36(1), 21–43. <https://doi.org/10.1007/s10212-020-00464-7>
- Zheng, Y., & Chiu, C. Y. (2019). *NPCD: Nonparametric Methods for Cognitive Diagnosis in R. (version 1.0-11)*. <https://CRAN.R-project.org/package=NPCD>