Article

Journal of Psychoeducational Assessment
2022, Vol. 40(2) 271–292

DOI: 10.1177/07342829211056396
journals.sagepub.com/home/jpa
SAGE

Evaluation of the Computer-Based Orthographic Processing Assessment: An Application of Cognitive Diagnostic Modeling

Yi-Jui I. Chen[1], **Yi-Hsin Chen**[1], **Jason L. Anthony**[1], and **Noé A. Erazo**[1]

Abstract

The Computer-based Orthographic Processing Assessment (COPA) is a newly developed assessment to measure orthographic processing skills, including rapid perception, access, differentiation, correction, and arrangement. In this study, cognitive diagnostic models were used to test if the dimensionality of the COPA conforms to theoretical expectation, evaluate individual items' quality, and examine the validity and the learning sequence of each skill. Results showed that the COPA captures five distinctive operating attributes, but some items could be revised to increase their item quality. Correlations with external variables confirmed that performances on the COPA are more strongly related to literacy outcomes than to oral language outcomes but that the COPA also demonstrates discriminant validity relative to even proximal measures of word reading and spelling. The mastery probabilities and best-fitting hierarchical model indicate that four of the five attributes follow a learning progression that is consistent with information processing theory and that was assumed by developers of the COPA.

Keywords

orthographic processing, cognitive diagnostic models, assessment validation, learning sequence of orthographic processing

Over four decades of literacy research points to the importance of phonological awareness, phonological memory, rapid automatic naming, alphabetic knowledge, and oral language as

[1]University of South Florida, Tampa, FL, USA

Corresponding Author:
Yi-Jui Iva Chen, Rightpath Research and Innovation Center, University of South Florida, 13301 Bruce B. Downs Blvd., MHC 1704, Tampa, FL 33612, USA.
Email: ichen@ets.org, iva811@berkeley.edu

essential for learning to read and write (National Reading Panel, 2000; National Early Literacy Panel & National Center for Family Literacy, 2008). Orthographic processing, “the ability to form, store, and access orthographic representations” (Stanovich & West, 1989, p. 404), is a set of cognitive skills that have received increasing amounts of attention. Major theories of word recognition and spelling acquisition (Coltheart, 2005; Ehri, 2005; 2014; Harm & Seidenberg, 1999; Perfetti & Hart, 2001; 2002; Seidenberg & McClelland, 1989; Share, 1995; 2004; Stanovich & West, 1989) propose that word learning requires well-specified orthographic representations that are securely linked to semantic and phonological information. These theories suggest that limitations in processing orthographic representations of words lead to reading and spelling difficulties, while strengths in orthographic processing distinguish skilled readers and spellers from unskilled readers and spellers.

Orthographic processing is usually investigated alongside phonological processing—the ability to attend to and make use of the phonological, or sound, structure of oral language (Wagner & Torgesen, 1987). Research shows that phonological processing skills are necessary but not sufficient for learning to read and write in many languages (Adams, 1990; Share, 1995, 1999). Evidence suggests that even after accounting for phonological skills and general abilities (i.e., verbal and non-verbal intelligence), orthographic processing skills contribute substantial amounts of variance to the prediction of word reading and spelling in monolingual (Barker et al., 1992; Chen et al., 2020, 2021; Conrad & Deacon, 2016; Cunningham, 2006; Rahbari et al., 2007; Roman et al., 2009) and bilingual individuals (Chung et al., 2018; Deacon, Chen, et al., 2013; Deacon, Commissaire, et al., 2013; Pasquarella et al., 2014; Sun-Alperin & Wang, 2011). While orthographic processing appears to be an independent and necessary component of word reading and spelling (Apel, 2009; Apel et al., 2011; Berninger et al., 1991, 1994), advancement in this area is hindered by lack of conceptual clarity and measurement problems (Apel, 2011; Burt, 2006; Castles & Nation, 2006; Chen et al., 2020; Hagiliassis et al., 2006; Wagner & Barker, 1994).

The Construct and Measurement of Orthographic Processing

A clear definition of the orthographic processing construct suffers from two major issues. One concern is the distinction between declarative knowledge and procedural operations (Burt, 2006; Castles & Nation, 2006). Declarative knowledge is a static mental database of stored information. An individual’s static, crystalized orthographic knowledge is composed of (a) graphemic representations of written words and parts of words and (b) the rules for permissible letter combinations and positional and contextual constraints on the use of letters in a given language. In contrast, an orthographic procedural operation is characterized by an individual’s ability to perceive, access, store, analyze, and manipulate static orthographic information. Some researchers have conceptualized (and operationalized) orthographic processing along the lines of declarative knowledge (e.g., Georgiou et al., 2009; Kirby et al., 2008), while others have conceptualized it along the lines of procedural operations (e.g., Grainger, 2018; Stanovich & West, 1989).

The second major conceptual challenge revolves around a widespread belief that orthographic skills consist of separate dimensions of *word-specific knowledge* and *general word knowledge* (Apel, 2011; Apel et al., 2019; Compton et al., 2020; Conrad et al., 2013). *General word knowledge* is the knowledge of general orthographic regularities, such as understanding the rules for permissible letter combinations and the positional and contextual constraints on the use of letters. General word knowledge is frequently measured by asking a respondent to choose a permissible letter combination among pseudowords (e.g., which one is most like a word: beff or ffeb?). Word-specific knowledge is the accurate memory of graphemic representations of written words or parts of words and is frequently measured by asking a respondent to choose between two homophones (e.g., which one is a fruit: pear or pair?).

Conceptualization of orthographic processing as declarative knowledge for specific words underlies the design of some orthographic processing measures, which has led to a pervasive measurement problem. Many assessments of orthographic processing employ word recognition tasks or spelling tasks (see Apel, 2011; Burt, 2006; Chen et al., 2020, for review). These approaches to measuring orthographic processing lead to a circular interpretation of findings (Burt, 2006; Castles & Nation, 2006): The predictor of the ability to recognize or spell a given word is the ability to recognize or spell words. In summary, existing measures of orthographic processing contain word recognition and spelling variance that does not pertain to orthographic processing, and these measures fail to include the essential procedural aspects of orthographic processing. Thus, the existing measures are limited because of construct-irrelevant variance and construct underrepresentation (Messick, 1994), and have thwarted progress in understanding the role or roles of various orthographic processing skills in reading and spelling.

An Alternative Operationalization of Orthographic Processing

One means to potentially overcome the noted obstacles is to develop measures that specifically assess the procedural operations involved in orthographic processing. Focusing on the procedural operations offers two advantages. First, by moving away from assessing word-specific spelling and word-specific recognition, the field could examine the relative contributions of phonological processing and orthographic processing in the absence of confounding orthographic processing with crystalized word knowledge. This may improve interpretability of findings through use of measures that demonstrate better discriminant validity. Second, the use of measures focused on procedural operations may permit identification of the roles of various orthographic processing skills as distinct from those of orthographic knowledge.

Two recent studies by Chen et al. (2020, 2021) implemented a measurement approach that emphasized the procedural and operational aspects of orthographic processing. Authors' Computer-based Orthographic Processing Assessment (COPA) focuses on indexing five procedural operations: perception, access, differentiation, correction, and arrangement. To avoid confounding orthographic processing with word-specific recognition and word-specific spelling knowledge, items were developed that ensure knowing the spelling of a word would not necessarily guarantee selection of the correct answer, and conversely, not knowing the spelling of a word would not necessarily guarantee selection of an incorrect answer. That is, when correct answers were real words, correct spellings were not provided as one of the response options. Instead, respondents were provided with multiple orthographic representations that they could then use to form the correct spelling. Chen et al. (2020) demonstrated that orthographic processing, as measured by the COPA, was distinguishable from word recognition and was predictive of spelling ability.

Need for Evaluation of Measurement Methods Focused on Procedural Operations

Although arguably advancing the field in significant ways, the studies by Chen and colleagues (2020, 2021) can be criticized for only minimally evaluating the construct validity of their novel operationalizations of orthographic processing. First, it is necessary to examine if the COPA adequately measures each of the procedural operations involved in orthographic processing that it was intended to measure. Second, although Chen et al. (2020) provided preliminary evidence that their operationalization of orthographic processing was distinguishable from word identification skills, it remains necessary to evaluate if each of five procedural operations (i.e., rapid perception, access, differentiation, correction, and arrangement) were distinguishable from spelling, orthographic knowledge, and phonological processing.

The Present Study

In the current study, we evaluate the construct validity and psychometric properties of the COPA and its methods of operationalizing orthographic processing. We aim to (a) test if the dimensionality of the COPA conforms to theoretical expectation, (b) evaluate individual items' psychometric properties, (c) examine convergent and discriminant validity of each procedural operation subtest, and (d) explore the learning sequence of each procedural operation.

Cognitive diagnostic models (CDMs) are a family of relatively new statistical models that are appropriate for addressing our four research goals. One, CDMs offer a means to evaluate the extent to which the COPA reflects the five hypothesized procedural operations of rapid perception, access, differentiation, correction, and arrangement. Two, CDMs can provide guessing, slipping, and discrimination parameters for each test item included in the COPA. Third, CDMs permit us to correlate the probability of mastery of each procedural operation, or latent attribute, with scores obtained on external measures of orthographic knowledge, word recognition, phonological processing, and spelling ability. Fourth, because CDMs estimate the probability of mastery of each attribute for each student (de la Torre, 2011; Lee et al., 2012), individuals' *cognitive profiles*[1] of mastery and non-mastery status of each attribute can be examined for conformity with theoretical expectations.

Method

Participants

Participants included 74 second-grade students (39 boys and 35 girls) who attended two schools (one urban and one suburban) in northern California. The parental consent was obtained before the data collection. The mean age of participants at the time of data collection was 8.03 years (SD = 0.33 years). The majority of participants were White, non-Hispanic (68%). The remaining participants were Asian (15%), African American (6%), Hispanic/Latino (4%), multiracial (3%), or unreported (4%).

Measures

Orthographic processing. The COPA (Chen et al., 2020) was used to assess participants' orthographic processing ability. The computer-based assessment included 60 test items and 10 practice items. The first fifty-one items required a fixed response, while the last nine items required an open-ended response.

Items 1–10 were to measure the ability to count the number of letters among a group of symbols within 5 seconds (e.g., *g$cykh*); items 11–20, the ability to identify a letter combination among four options after seeing the target letter combination (e.g., *tplkmn*) for 3 seconds and then followed by 4 second break (e.g., *tplkmn*, *tpikmn*, *tplhmn*, *tplkwn*); items 21–31, the ability to identify the permissible combination among three other impermissible combinations (e.g., *tiught*, *cuught*, *tought*, *tought*, *cyught*); items 32–51, the ability to correct a spelling by either adding or deleting a letter (e.g., *detal*: l, *h*, i, *o*); and items 52–60, the ability to rearrange letters to create a real word. A detailed description of item design can be found in Supplemental Appendix A. Table 1 summarizes how these items link to the attributes (i.e., procedural operations) and provided corresponding definitions, behaviors, and related item features.

Table 1. Summary of the Orthographic Processing Assessment used in Chen et al. (2020, 2021).

Items	Attribute	Definition	Behavior	Number of Items	Response Type	Time limitation	Example
1–10	Rapid perception	The ability to identify letters within a very short time	Identify the letters among a variety of symbols within 5 seconds	10 items 4 trials	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	Choose the letter string that is viewed a few seconds ago	10 items 2 trials	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	Identify a letter combination that looks real	11 items 1 trial	Fixed	Unlimited time	tiught cuught tought cyught
32–41	Correction: Addition	The ability to add a letter to make spelling correct	Correct the spelling by adding a letter	10 items 1 trial	Fixed	Unlimited time	detal: l h i o
42–51	Correction: Removal	The ability to delete a letter to make a spelling correct	Correct the spelling by deleting a letter	10 items 1 trial	Fixed	Unlimited time	lauyout: y u o a
51–60	Arrangement	The ability to arrange letter to make a real word	Use all of the letters to make a real word	9 items 1 trial	Open	Unlimited time	a, c, e, e, r, t _ _ _ _ _

Oral receptive vocabulary. The Peabody Picture Vocabulary Test-Fourth Edition (PPVT; Dunn & Dunn, 2012) was used to assess participants' receptive oral vocabulary. It includes 228 test items. The PPVT asks children to point to a picture that best illustrates a vocabulary word spoken by an examiner. The test developers report test–retest reliabilities ranging from .92 to .96 and split-half reliabilities ranging from .89 to .97 (Dunn & Dunn, 2012).

Word recognition. The Letter Word Identification subtest of the Woodcock–Johnson III (WJ-III; Woodcock et al., 2007) was used to measure participants' word recognition ability. This subtest requires children to say the name of a printed letter or to read a printed word. The 76 items assess 13 letters (e.g., *P*, *E*), 22 single-syllable words (e.g., *must*), and 31 polysyllabic words (e.g., *domesticated*). The median reliability of this test is .94 (McGrew et al., 2007).

Word attack. This was assessed with the Word Attack subset of the Woodcock–Johnson III Test of Achievement (WJ-III; Woodcock et al., 2007). This subtest is designed to assess children's ability to phonologically decode pseudowords, such as *gusy* or *sluke*. It has 32 items. For respondents aged 7–10 years, the split-half reliability ranged from .88 to .92 (McGrewet al., 2007).

Spelling. The spelling subtest of the WJ-III was used to measure participants' spelling ability. This subtest asks children to spell individual letters or words presented orally in a sentence by an examiner. The 59 items assess 4 hand-writing skills, 10 uppercase and lowercase letters (e.g., *Z*, *e*), 16 single-syllable words (e.g., *rain*, *six*), and 29 polysyllabic words (e.g., *difference*, *congenital*). The median test reliability of this test, reported by McCrew and colleagues, is .90 (McGrew et al., 2007).

Phonological awareness. The Elision and Blending subtests of the Comprehensive Test of Phonological Processing (Wagner et al., 1999) were used to assess participants' phonological awareness. The elision task requires that children omit a sound from a word that they hear. For example, children hear, "Say '*winter*' without the/t/sound," where the correct response is *winner*. Children do not see the spelling of the stimulus in this task. The blending task requires that children blend together phonemes that are presented from an audio recording. For example, the isolated sounds/s//t//æ//m//p/form the word *stamp* when blended together. Both subtests have 20 items. Cronbach's alphas for these measures were reported in Wagner et al. (1999). The Elision reliabilities ranged from .89 to .91, and Blending reliabilities ranged from .79 to .87.

Orthographic knowledge. The non-normed Orthographic Choice Task, developed by Olson et al. (1985), was used to measure participants' word-specific orthographic knowledge. This 23-item test asked children to pick the correct spelling from among a field of two homophones (e.g., *rain* vs. *rane*). The sample-specific Cronbach's alpha is .83.

Statistical Analyses

Overview. The log-linear cognitive diagnostic model (LCDM; Henson et al., 2009), the general diagnostic model (GDM; von Davier, 2008), and the generalized deterministic inputs, noisy "and" gate model (GDINA; de la Torre, 2011) are three types of CDMs. These three general models are saturated, and each can be further specified through the addition of parameter constraints that test specific hypotheses about relations among the attributes (de la Torre & Minchen, 2019).

The DINA model, one of the simplest CDMs, is employed in this study because the COPA is comprised of items that theoretically only measure one single attribute. This is referred to as the "simple structure" of items. Different from the GDINA model, the DINA model assumes the equal probability of success for all attribute vectors in the same group (de la Torre, 2011).

The DINA model possesses several advantages, such as straightforward interpretation of results, ready application of parsimony principle during model selection, appropriateness with small sample sizes, and better classification rates with small sample sizes (de la Torre et al., 2018; Rojas et al., 2012). The DINA model involves only two parameters for each item: the guessing (g_j) and slipping (s_j) parameters, where j represents an item. A latent response to item j for student i 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 item j, and K is the total number of attributes. The latent response (η_{ij}) assumes a value of 0 or 1, where $\eta_{ij} = 1$ indicates that student i possesses all the attributes required for item j and $\eta_{ij} = 0$ indicates that student i lacks at least one of these attributes (de la Torre, 2009). Thus, the guessing parameter is defined by the probability of a correct response for examinees who may not possess all the attributes or lack one of the necessary attributes required for an item, denoted by $g_j = P(X_{ij} = 1 \mid \eta_{ij} = 0)$. The slipping parameter is defined by the probability of an incorrect response for those who possess all the necessary attributes, denoted by $s_j = P\,(X_{ij} = 0 \mid \eta_{ij} = 1)$. In addition to these two item parameters, de la Torre (2008) suggested 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 0.

In the same way that item and ability parameters are estimated in IRT models, maximum likelihood estimation with Expectation-Maximization (E-M) algorithm is implemented to estimate item parameters and latent attribute patterns (i.e., cognitive profiles) in the DINA model. The likelihood for each possible attribute pattern is estimated, and the attribute pattern with the highest likelihood is assigned to an examinee. Taking hierarchical relations among attributes into consideration reduces the number of possible attribute patterns (de la Torre, 2009; Leighton et al., 2004). For mathematical equations and detailed estimation procedures, interested readers can refer to Jimmy de la Torre (2009).

Data analysis strategy. First, we used the SAS macro named EZLID (Hsu & Tsai, 2007) to generate the most commonly used local dependence indices, Pearson's chi-square, χ^2, and the likelihood ratio, G^2 (Chen & Thissen, 1997) to evaluate local item dependence (LID) among items. These LID indices, χ^2 and G^2, were computed based on the parameter estimates of the best-fitting IRT model using the PROC IRT procedure in Statistical Analysis Software (SAS Institute, Inc., 2020). Because the critical values of 3.84 and 6.63, with corresponding alpha levels of .05 and .01, respectively, have been found overly conservative for identifying dependent item pairs (Chen & Thissen, 1997), we used a more liberal criterion of 10 proposed by Pommerich and Segall (2008) to identify pairs of items needing further investigation of item dependence.

Second, the dimensionality of the COPA was evaluated with one-attribute, five-attribute, and six-attribute DINA models that employed marginal maximum likelihood (MML) estimation. The one-attribute model tested the unidimensionality of the COPA. The Q-matrix for the one-attribute model specified that all items measure the same general orthographic processing attribute. The five-attribute model tested if performances on the COPA reflect five latent attributes (i.e., procedural operations) as proposed by Chen et al. (2020). Specifically, the Q-matrix specified single loadings on rapid perception, access, differentiation, correction, and arrangement attributes. Because the COPA includes two different types of correction items, addition and removal, the six-attribute model tested if performances on the correction items are better explained by two latent attributes: addition and removal. That is, items that loaded on the correction attribute of the five-attribute model were specified in the six-attribute model as loading on an addition attribute (items 32–41) or a removal attribute (items 42–51). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) fit indices, as well as likelihood ratio (LR) tests, were used to identify the best fitting of the three models. The model with the lower values of AIC or BIC presents a better fit. A statistically significant result with the LR test indicates that a more complex model fits the data better, whereas a non-significant result supports a more parsimony model.

Third, hierarchical models were used to examine the learning sequences among attributes. The COPA was designed according to information processing theory. Information processing theory posits that, like a computer, the human brain has limited capacity to solve problems involving uncertainty. Information processing theory also assumes that humans process multiple, simultaneous inputs of information through several mental processes. Examples of mental processes include encoding, temporarily storing, retrieving, transforming, and permanently storing information (Çeliköz et al., 2019). Depending on the task that an individual is trying to accomplish, there will be more or less uncertainty on how to solve it. It is typically believed that tasks that involved more mental processes have greater uncertainty (Simon, 1978). For example, the arrangement task (anagram), which involved six mental processes (e.g., perceive, store, retrieve, compare, transform, and manipulate stimuli) has greater uncertainty than the rapid perception that only involves a single mental process (e.g., perceive stimulus; Chen et al., 2020). The COPA was developed under the assumption that the mastery of orthographic processing skills follows a consecutive sequence of rapid perception, access, differentiation, correction, and finally arrangement. To evaluate this assumption, we performed a backward method to build hierarchical models. Four hierarchical models were tested in this study. Hierarchical model 1 (H1) only assumed Correction needs to be mastered before Arrangement. Hierarchical model 2 (H2) added the assumption to H1 that Differentiation needs to be mastered before Correction. Hierarchical model 3 (H3) added the assumption to H2 that Access needs to be mastered before Differentiation. Hierarchical model 4 (H4) added the assumption to H3 that Rapid Perception must be mastered before Access, and followed the strictest stage-based assumption. Comparison of AIC and BIC fit indices and LR tests were used to identify the best fitting hierarchical model.

Four, item properties based on the best fitting hierarchical DINA model were evaluated next. Specifically, guessing, slipping, and discrimination indices were computed via the “CDM” R package (Robitzsch et al., 2020). Relatively small guessing and slipping parameters and relatively large IDI statistics indicate relatively high-quality items.

Lastly, the best fitting hierarchical DINA model was used to conduct a cognitive diagnostic analysis, using Bayesian maximum a posteriori (MAP) estimation. Marginal mastery probabilities for attributes and proportions of diagnostic cognitive profiles were checked. The proportion-correct differences (PDIFF) between masters and non-masters for individual items were computed to check internal validity (Roussos et al., 2007). Large PDIFFs for individual items and the whole assessment indicate the high quality of test items, the Q-matrix, and diagnostic profiles (Chen et al., 2019; Roussos et al., 2007). A PDIFF of .4 has been recommended as a cutoff value (Chen et al., 2019; Jang, 2009; Roussos et al., 2007). Lastly, correlations of individuals’ mastery probabilities of each attribute with the external measures were computed to examine convergent and discriminant validity.

Results

Dimensionality of Computer-based Orthographic Processing Assessment

Table 2 reports model fits for the one-attribute, five-attribute, and six-attribute models. Akaike information criterion and BIC fit indices favor the five-attribute model, given that it yielded the lowest values for AIC and BIC. An LR test found the five-attribute model fits significantly better than the one-attribute model, χ^2 (30) = 222.49, p < .001 (Table 3). Another LR test found the six-attribute model did not make a statistically significant improvement than the five-attribute model, χ^2 (32) = 19.25, p = .84. Based on the parsimony rule, we selected the five-attribute model as the superior model.

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

Models	Max (χ^2)	AIC	BIC	MADcor	# class par
One-attribute	32.76, p < .001	4965	5244	0.09	1
Five-attribute	23.73, p = .001	4802	5150	0.09	31
Six-attribute	18.69, p = .027	4847	5269	0.09	63

Note. Max(χ^2) = the maximum of all χ^2 statistics accompanied with a p-value obtained by the Holm procedure; AIC = Akaike information criterion; BIC = Bayesian information criterion; MADcor = mean of absolute deviations in observed and expected correlations (Di-Bello et al., 2007); # class par = number of class parameters.

Table 3. Likelihood Ratio Tests Between One-Attribute, Five Attribute, and Six-Attribute Non-Hierarchical Models.

Models		Deviance		Difference		
A	B	A	B	A–B	df_{diff}	p
One attribute	Five attributes	4723	4500	223	30	<.001
One attribute	Six attributes	4723	4481	242	62	<.001
Five attributes	Six attributes	4500	4481	19	32	.96

Note. df_{diff} = difference in the degree of freedom between two models.

Table 4. Fit Statistics for Five-Attribute Hierarchical Models Compared to a Five Non-Hierarchical Model.

Models	Max (χ^2)	AIC	BIC	MADcor	# class par
H0: Non-hierarchical	23.73 p = .001	4802	5150	0.09	31
H1: A4→A5	24.58 p =.001	4784	5114	0.09	23
H2: A3→A4→A5	24.59 p =.001	4771	5082	0.09	15
H3: A2→A3→A4→A5	21.91 p =.005	4762	5059	0.09	9
H4: A1→A2→A3→A4→A5	23.12 p =.003	4928	5216	0.09	5

Note. Max(χ2) = The maximum of all χ^2 statistics accompanied with a p-value obtained by the Holm procedure; AIC = Akaike information criterion; BIC = Bayesian information criterion; MADcor = mean of absolute deviations in observed and expected correlations (Di-Bello et al., 2007); # class par = number of skill class parameters; H1–H4 = Hierarchical models 1–4. H1 = Correction (A4) is mastered before arrangement (A5); H2 = differentiation (A3) is master before correction (A4), and then arrangement (A5); H3 = access (A2) is mastered before differentiation (A3), correction (A4), and then arrangement (A5); H4 = perception (A1) is mastered before access (A2), differentiation (A3), correction (A4), and then arrangement (A5).

Learning Sequence among Computer-based Orthographic Processing Assessment Subtests

As shown in Table 4, the hierarchical models 1, 2, and 3 (H1, H2, and H3) describe the data better than the non-hierarchical model, according to relative values of AIC and BIC. H3 (AIC = 4762, BIC = 5059) showed the best fit with the smallest values of AIC and BIC, and surprisingly, H4 (AIC = 4928, BIC = 5216) had a worst with the largest values of AIC and BIC. Consistent with the AIC and BIC, LR tests (Table 5) demonstrated that H3 characterized the data better than H1, H2, H4, and non-hierarchical models. Specifically, more saturated models, including H0, H1, H2, did not make a significant improvement over H3. The LR tests are χ^2 (22) = 3.82, p = .99, for the H0 and H3 comparison; χ^2 (14) = 6.02, p = .97 for the H1 and H3 comparison, and χ^2 (6) = 2.76, p =.83 for H2 and H3 comparison. Moreover, H3 made a statistically significant improvement than the most parsimony model, H4, χ^2 (4) = 173.81, p < .001. Therefore, H3 was selected for further analysis.

Evaluation of Computer-based Orthographic Processing Assessment Items

The internal consistency reliability coefficients, Cronbach alphas, for the whole test and five subtests were .88, .85, .62, .69, .81, and .53, respectively. Table 6 presents item statistics based on the best-fitting DINA model, Hierarchical model 3 (H3). The percentage of correct responses ranged from 5% to 91%. Overall, the arrangement items were more difficult than the other items.

Table 5. Likelihood Ratio Tests Between Five Attribute Hierarchical and Non-Hierarchical Models.

Model	Deviance	(#of class parameter)		Model	Deviance	(# of class parameter)	χ^2	*df*	*p*
H0	4500	(31)	versus	H1	4498	(23)	−2.20	8	1
H0	4500	(31)	versus	H2	4501	(15)	1.06	16	1
H0	4500	(31)	versus	H3	4504	(9)	3.82	22	.99
H0	4500	(31)	versus	H4	4678	(5)	177.63	26	<.001
H1	4498	(23)	versus	H2	4501	(15)	3.26	8	.92
H1	4498	(23)	versus	H3	4504	(9)	6.02	14	.97
H1	4498	(23)	versus	H4	4678	(5)	179.83	18	<.001
H2	4501	(15)	versus	H3	4504	(9)	2.76	6	.83
H2	4501	(15)	versus	H4	4678	(5)	176.57	10	<.001
H3	4504	(15)	versus	H4	4678	(5)	173.81	4	<.001

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 is bold is the better model in each pair of comparison. H2 is the best fitting model among four hierarchical models.

As shown in Table 6, items 13, 22, 25, 26, 31, 32, 51, 55, 58, and 60 had IDI values less than .20, indicating the low quality of these test items. These items also had low PDIFFs. PDIFFs are the difference in the percentage of respondents who answered a given item correctly between the mastery and the non-mastery groups. Taking item 13 as an example, 50% of respondents in the non-mastery group and 47% of respondents in the mastery group answered item 13 correctly, and thus PDIFF for item 13 is −3%, indicating that non-masters performed slightly better on item 13 than masters. There were 26 (43%) items where the PDIFFs were greater than .40 (i.e., high quality items), 23 (38%) items where the PDIFFs were between .20 and .40 (i.e., moderate quality items), and 11 (18%) items where the PDIFFs were less than .20 (low-quality items). Items 13, 22, 25, 26, 31, 32, 51, 55, 58, and 60 were identified as having PDIFF values less than .20, which signifies low-quality items.

Our evaluation of possible item dependencies among items involved examination of χ^2 and G^2 indices (Table 7). The average χ^2 and G^2 values for the entire COPA were 1.25 and 1.29, respectively. The average χ^2 and G^2 values per scale ranged from 0.93 (Addition) to 2.22 (Removal) and from 0.93 (Differentiation and Addition) to 2.40 (Removal), respectively. Thus, the items, in general, were largely independent, conditional on the factor structure of the COPA. Nonetheless, two item pairs demonstrated appreciable item dependence when employing a liberal criterion of LID indices (i.e., greater than 10; Pommerich & Segall, 2008). Specifically, item 9 and item 10 from the Rapid Perception and item 43 and item 47 from Removal evinced possible dependencies with χ^2 and G^2 values of 11.06 and 11.96 and 9.98 and 12.69, respectively.

Mastery Probabilities, Diagnostic Cognitive Profiles, and Classification Accuracy and Consistency

Proportions of mastery for the five skills based on the best-fitting hierarchical model, H3, were .55, .87, .77, .77, and .51, for Rapid Perception, Access, Differentiation, Correction, and Arrangement, respectively. The most common cognitive profiles among students were 01110 (18%) and 01111 (11%). The cognitive profiles that included at least 5% of students were 00000 (9 %), 11110 (7%), and 01000 (7%). The remaining cognitive profiles were infrequent, 11111 (4%), 10000 (4%), and 11000 (3%).

Based on Bayesian maximum a posteriori (MAP) estimation, classification accuracy rates of H3 were .91 using the estimator by Johnson and Sinharay (2018) and .97 based on simulated data that was randomly drawn from 5000 observations from the sample. Classification consistency rates were

Table 6. Item Parameters.

Items	Guess	Slip	IDI	RMSEA	% of correct response: Whole, %	Non-master, %	Master, %	PDIFF
Item 1	0.06	0.19	0.75	0.11	47	6	83	77%
Item 2	0.07	0.10	0.83	0.06	53	8	90	82%
Item 3	0.03	0.31	0.66	0.11	39	3	70	67%
Item 4	0.08	0.25	0.68	0.13	45	9	75	66%
Item 5	0.00	0.31	0.69	0.14	38	0	70	70%
Item 6	0.00	0.16	0.84	0.17	46	0	85	85%
Item 7	0.55	0.25	0.20	0.22	66	56	75	19%
Item 8	0.49	0.22	0.29	0.18	65	50	78	28%
Item 9	0.49	0.30	0.21	0.27	61	50	70	20%
Item 10	0.49	0.22	0.29	0.15	65	50	78	28%
Item 11	0.00	0.39	0.61	0.16	53	0	61	61%
Item 12	0.20	0.42	0.37	0.13	53	20	58	38%
Item 13	0.49	0.53	−0.02	0.22	47	50	47	−3%
Item 14	0.29	0.37	0.34	0.14	58	30	63	33%
Item 15	0.20	0.42	0.37	0.24	53	20	58	38%
Item 16	0.00	0.53	0.47	0.19	41	0	47	47%
Item 17	0.00	0.30	0.70	0.15	61	0	70	70%
Item 18	0.08	0.42	0.50	0.22	51	10	58	48%
Item 19	0.00	0.39	0.61	0.23	53	0	61	61%
Item 20	0.29	0.22	0.49	0.20	72	30	78	48%
Item 21	0.12	0.42	0.47	0.20	47	12	58	46%
Item 22	0.77	0.11	0.13	0.19	86	76	89	13%
Item 23	0.59	0.09	0.32	0.16	84	59	91	32%
Item 24	0.60	0.09	0.32	0.13	84	59	91	32%
Item 25	0.77	0.07	0.16	0.12	89	76	93	17%
Item 26	0.59	0.26	0.15	0.12	70	59	74	15%
Item 27	0.41	0.10	0.49	0.27	78	41	89	48%
Item 28	0.35	0.35	0.30	0.05	58	35	65	30%
Item 29	0.54	0.12	0.34	0.11	80	53	88	35%
Item 30	0.48	0.09	0.43	0.20	81	47	91	44%
Item 31	0.48	0.33	0.19	0.10	62	47	67	20%
Item 32	0.77	0.09	0.14	0.18	88	76	91	15%
Item 33	0.48	0.16	0.37	0.15	76	47	84	37%
Item 34	0.52	0.24	0.24	0.26	70	53	75	23%
Item 35	0.71	0.04	0.26	0.15	91	71	96	26%
Item 36	0.18	0.21	0.62	0.13	65	18	79	61%
Item 37	0.19	0.42	0.39	0.16	49	18	58	40%
Item 38	0.30	0.40	0.29	0.24	53	29	60	30%
Item 39	0.59	0.11	0.30	0.16	82	59	89	31%
Item 40	0.46	0.24	0.29	0.26	69	47	75	28%
Item 41	0.48	0.02	0.50	0.15	86	47	98	51%
Item 42	0.29	0.23	0.48	0.19	66	29	77	48%
Item 43	0.60	0.00	0.41	0.18	91	59	100	41%
Item 44	0.25	0.21	0.54	0.17	66	24	79	55%

(continued)

Table 6. (continued)

Items	Guess	Slip	IDI	RMSEA	% of correct response Whole, %	Non-master, %	Master, %	PDIFF
Item 45	0.59	0.04	0.37	0.14	88	59	96	38%
Item 46	0.59	0.19	0.21	0.16	76	59	81	22%
Item 47	0.19	0.05	0.76	0.12	77	18	95	77%
Item 48	0.24	0.23	0.54	0.10	65	24	77	54%
Item 49	0.23	0.31	0.45	0.16	58	24	68	45%
Item 50	0.29	0.16	0.55	0.13	72	29	84	55%
Item 51	0.42	0.42	0.16	0.19	54	41	58	17%
Item 52	0.67	0.11	0.23	0.20	78	71	81	10%
Item 53	0.33	0.08	0.59	0.12	64	33	92	59%
Item 54	0.30	0.31	0.39	0.13	50	31	68	38%
Item 55	0.12	0.69	0.19	0.15	22	11	32	20%
Item 56	0.00	0.71	0.29	0.02	15	0	29	29%
Item 57	0.22	0.50	0.27	0.15	36	22	50	28%
Item 58	0.00	0.90	0.11	0.08	5	0	11	11%
Item 59	0.00	0.61	0.39	0.01	20	0	39	39%
Item 60	0.00	0.90	0.11	0.01	5	0	11	11%

Note. IDI = item discrimination index; RMSEA = root mean square error of approximation; PDIFF= proportion-correct differences between masters and non-masters.

Table 7. Examination of Local Item Dependence.

	LID index	*N*	Minimum	Maximum	Mean	Standard deviation
Rapid perception	χ^2	45	0.06	11.06	1.41	2.14
	G^2	45	0.06	11.96	1.44	2.27
Access	χ^2	45	0.001	4.90	1.03	1.44
	G^2	45	0.001	4.98	1.04	1.46
Differentiation	χ^2	55	0.001	3.99	0.95	1.00
	G^2	55	0.001	3.80	0.93	0.96
Addition	χ^2	45	0.001	4.67	0.93	1.09
	G^2	45	0.001	4.61	0.93	1.09
Removal	χ^2	45	0.001	9.98	2.22	2.64
	G^2	45	0.001	12.69	2.40	3.01
Arrangement	χ^2	36	0.02	6.15	0.96	1.35
	G^2	36	0.02	5.13	1.00	1.29
Mean	χ^2	226	0.01	6.79	1.25	1.61
	G^2	226	0.01	7.20	1.29	1.68

Note. LID = local item dependence; *N* = the number of item pairs.

.90 using the estimator by Johnson and Sinharay (2018) and .91 based on simulated data. For five individual skills, accuracy rates were .93 or higher, and classification consistency rates were .91 or higher.

Correlation with External Measures

The correlations between the five attributes and six external measures (i.e., Word recognition, Word attack, Elision, Blending, Spelling, Orthographic knowledge, and Oral language) ranged from .06 to .56 (Table 8). The correlation between Rapid Perception and Oral language was the weakest, r = .06, whereas the correlations between Differentiation and Word attack, and Arrangement and Word attack were the largest, rs = .56. From among the five attributes, Perception generally had the lowest correlations with external measures (rs =.13 to .38). However, Perception's correlation with Blending (r = .38) was the highest correlation of Blending with any the COPA attribute. Access, Differentiation, Correction, and Arrangement had moderate correlations with all external measures (rs = .31 to .56). From among the seven external measures, Word Recognition, Word Attack, and Spelling yielded the strongest correlations with attributes from the COPA (mean rs= .45, .48, and .44, respectively); whereas Oral language and Blending yielded the lowest correlations with attributes from the COPA (mean rs = .27 and .35, respectively).

Discussion

The Dimensionality of the Computer-based Orthographic Processing Assessment and Item Quality

Examining the results of the DINA models shows that the COPA indexes five distinct attributes, including rapid perception, access, differentiation, correction, and arrangement. These five attributes represent the skills of identifying letters within a very short time (i.e., rapid perception), storing letter combinations and then retrieving them from short-term memory (i.e., access), differentiating permissible and impermissible letter combinations (i.e., differentiation), adding or deleting a letter to make spelling correct (i.e., correction), and arranging letter to make a real word (i.e., arrangement). The finding of the dimensionality reflects that addition and removal items are well characterized by one attribute, correction. This finding aligns with Chen and colleagues (2020) assumption that the removal and addition items can be combined into one attribute.

Items 13, 22, 25, 26, 31, 32, 51, 52, 55, 58, and 60 need to be scrutinized. These items either have a relatively low IDI or low PDIFFs. Most low-quality items (item 51, 52, 55, 58, and 60) are the arrangement items and differentiation items (items 22, 25, 26, and 31). The arrangement attribute is the most difficult attribute in the COPA because the respondents need to produce an answer rather than select an answer. The low item quality indices of these four items are likely caused by the difficulty of the arrangement items, and most of the participants in this study did not have the ability to answer those items correctly. Thus, the arrangement items could not effectively discriminate the participants' orthographic processing ability in our study. Future studies with more advanced readers are needed to appropriately evaluate the arrangement items.

The relatively low quality for the differentiation items is unexpected. We suspect that the low item quality might due to these items lacking the sensitivity to detect the difference in second-grade students' ability to differentiate permissible and impermissible spelling patterns, especially when these students have similar educational backgrounds and English print exposure experience. This assumption seems to align with Chen and colleagues (2020) findings. Chen and colleagues (2020) reported that these differentiation items were consistently more difficult for Taiwanese participants than for the U.S. second-grade participants. They further indicated that these differentiation items were able to detect the difference in the ability in differentiating permissible and impermissible patterns between the U.S. and Taiwanese participants who had very distinct educational backgrounds and English print exposure experience from each other. However, it seems that the differentiation items in the COPA might not be suitable to detect the difference in the U.S. second-grade students' ability to differentiate permissible and impermissible patterns given their similar learning and English print exposure experience. When assessing second-grade students in the United States who presumably see English print on a daily basis, these targets (permissible letter combinations) and foils (impermissible letter combinations) might need to increase the

Table 8. Correlation with External Measures.

	1	2	3	4	5	6	7	8	9	10	11	12
1. Perception		.29	.21	.28	.51	.22	.26	.31	.38	.23	.13	.06
2. Access			.50	.56	.89	.51	.47	.32	.34	.51	.38	.33
3. Differentiation				.93	.49	.49	.56	.40	.31	.45	.49	.28
4. Correction					.54	.51	.53	.45	.36	.46	.51	.32
5. Arrangement						.53	.56	.41	.38	.54	.32	.34
6. Word recognition							.73	.51	.41	.74	.56	.46
7. Word attack								.55	.46	.61	.46	.36
8. Elision									.40	.48	.25	.31
9. Blending										.34	.01	.39
10. Spelling											.47	.34
11. Orthographic knowledge												.12
12. Oral language												

Note. N = 74. Correlations larger than .23 are significant at p < .05.

subtlety. In short, perhaps expanding the item content of the differentiation attribute to include more complex and sophisticated patterns may improve the measurement of this theoretically important orthographic processing procedural operation.

Consistent with Chen and colleagues (2020), our study also indicated that item 13 was the item with the lowest quality. Item 13 is an access item: respondents identify a letter combination among four options after seeing the target letter combination for 3 seconds and then followed by a 4 second break. One possible explanation for this low item quality is that item 13 (the target, *tplkmn* with the corresponding foils, *tpiknm*, *tplhmn*, and *tplkwn*) has several visually similar letters, such as *h* versus *n*, *i* versus *l*, *m* versus *w*, and these stimuli are not pronounceable. Therefore, within 4 seconds of seeing the target, most second-grade respondents might not be able to precisely store these letters, which in turn reduces item 13's functionality to detect second-grade students' ability to store and retrieve letter combinations. Item 13 can be improved by using more distinguishable foils.

Learning Sequence and Mastery Probabilities

The best-fitting hierarchical model and mastery probabilities of the five attributes imply that there is a learning progression among four of the orthographic processing skills, i.e., access → differentiation → correction → arrangement. The learning sequence evidenced in this study comes very close to the sequence proposed by Chen and colleagues (2020), i.e., rapid perception → access → differentiation → correction → arrangement. That mastery of rapid perception did not precede mastery of access in the present study could be due to imposing a five-second time-out rule on rapid perception items. In other words, perhaps the speeded nature of the perception items made it more difficult than expected, and maybe relaxing the time-out criterion would yield the expected learning sequence. An alternative explanation is that perhaps these rapid perception items involve more complex processing than they were intended to measure. When encountering a rapid perception item on the COPA, respondents are required to distinguish letters from within an array of symbols, count the number of letters in the array, and press the number key to report the number of letters, all within a time limitation. To comprehensively test the learning sequence of these five attributes, the COPA needs to expand its design by incorporating both time-limited and time-unlimited items for each attribute. Moreover, future development of the rapid perception items could use a purer design to remove the effect of numeracy. One could design a tablet-based assessment and ask respondents to touch the letters among an array of symbols rather than counting the number of English letters. Alternatively, one could use a decision task to ask respondents to decide whether or not a presented stimulus is an English letter.

In line with Chen and colleagues (2020) proposition, the best-fitting hierarchical model and the mastery probabilities show that access is acquired before differentiation, correction, and arrangement. This finding confirms access is a foundational skill and serves as a prerequisite for multiprocessing skills, such as differentiation, correction, and arrangement.

Unlike rapid perception and access items which have a time limit in answering an item or seeing a stimulus, differentiation, correction, and arrangement items are not time-limited, and thus they can be directly compared with each other. Research has shown that children develop the sensitivity to differentiate permissible and impermissible letter combinations at a very early stage (Cassar & Treiman, 1997; Rothe et al., 2014; Treiman, 1993). The development of such sensitivity reflects that young children acquire knowledge of the statistical regularities of spelling patterns before they have fully developed their reading abilities (Apel et al., 2013; Rothe et al., 2014). Consistent with previous research, our best-fitting hierarchical model indicates that the differentiation skill seems to be developed before the ability to correct spelling and arrange letters to form spellings.

Although the best-fitting hierarchical model, H3, shows that the differentiation skill is developed before the correction skill, the estimated mastery probabilities of differentiation and correction attributes are identical. One explanation for this is that although differentiation might begin to develop before correction, it is likely that the acquisition time frame for these two attributes might be partially overlapping or very close to each other. Thus, the estimated mastery probabilities based on a group of second-grade students are not sensitive enough to capture the difference between them. An alternative explanation is that the correction items provide specific instruction for respondents to either add or delete a letter, so the respondents only need to focus on one type of correction procedure when answering an item (i.e., only think about adding a letter for an addition item). This reduces the item difficulty and results in a higher mastery probability. If no specific instructions were provided and addition, removal, and transposition needed to be simultaneously considered while answering a correction item, then the item difficulty of correction items would increase and the estimated mastery probability of the correction attribute would decrease. The correction items in the COPA could make such revisions and examine the mastery probability of the differentiation and correction attributes.

Relation with Word Recognition, Spelling, Orthographic Knowledge, and Phonological Processing

One main challenge in designing measures for orthographic processing is that orthographic processing assessments are frequently confounded with the measures of word recognition, spelling ability, orthographic knowledge, and phonological processing (Burt, 2006; Castles & Nation, 2008; Chalmers & Burt, 2008; Chen et al., 2020; Hagiliassis et al., 2006). Chen and colleagues (2020) demonstrated that the COPA provides indices of orthographic processing skills that are only moderately correlated with and are therefore distinguishable from word identification skills. The present study picked up where Chen and colleagues left off by demonstrating that each of five orthographic processing skills measured by the COPA is related to but distinguishable from word-specific orthographic knowledge, phonological awareness, word identification, and spelling abilities.

Consistent with Chen and colleagues' (2020) findings, we found that five of the COPA attributes have weaker correlations with oral vocabulary (ranging from .06 to .34) than those with word identification (range from .22 to .53) and spelling (range from .23 to .54). Orthographic processing has been widely considered as a critical skill for word recognition (Barker et al., 1992;

Chung et al., 2018; Cunningham et al., 2001; Stanovich & West, 1989) and spelling ability (Chen et al., 2021; Chung et al., 2018). As oral vocabulary does not necessarily involve orthographic representations (i.e., graphemes), oral vocabulary should demonstrate weaker correlations with the five COPA attributes, compared to word identification and spelling. Therefore, the stronger correlations with word identification and spelling than those with oral vocabulary are another piece of evidence for the validity of the COPA.

Limitations and Future Study

The small and relatively homogeneous sample included in the present study reduces the confidence in our findings and certainly limits their generalizability. In partial defense of the study's design, we employed a statistical approach that addressed our research questions and that has been shown reliable with small samples (Paulsen & Valdivia, 2021; Sen & Cohen, 2021). Nonetheless, future studies concerned with learning trajectories of orthographic processing skills would benefit from larger samples of children who vary more in their literacy abilities.

We encourage future studies to include think-aloud tasks and invite participants to share what they are thinking during literacy and orthographic processing assessments because this may yield fruitful hypotheses about why some orthographic processing skills perform better than others and the potentially unique roles of orthographic processing in literacy.

Because we took the simple structure approach to design the COPA items, our evidence of the dimensionality is limited by item design. Two approaches can be taken to overcome this limitation. First, future studies could create multiple tasks for each orthographic processing skill, similar to our design for the correction skill, in which we include the addition task and the deletion task. Second, future studies could use a complex structure approach to design the COPA items. That is, one could create items that measure two or more attributes (e.g., items that involve rapid perception and access) and then conduct CDM analysis to see if the COPA still reflects five attributes, rapid perception, access, differentiation, correction, and arrangement.

Finally, the COPA is unique in that it permits researchers to measure the procedural aspects of orthographic processing. However, a limitation of the COPA is that it does not independently measure the declarative aspects of orthographic processing. Therefore, some educators and researchers may wish to use the COPA in combination with simplistic declarative measures to yield a comprehensive understanding of students' orthographic processing skills.

Conclusion

In addition to phonological processing, orthographic processing is another critical factor for pinpointing beginning readers and spellers' challenge in learning to read and spell (Chen et al., 2021). The use of this assessment in literacy research will help educators and researchers to understand another source of challenges when learning to read and spell, which in turn will provide more targeted and suitable instructional support. Specifically, by analyzing participants' responses to the COPA, we can pinpoint potential problems that hinder word reading and spelling acquisition. For example, if students struggle with differentiation items, it suggests they have not yet developed adequate sensitivity for English spelling patterns. Teaching letter clusters or common vowel patterns will be beneficial to this group of students. If students have difficulty in answering the access items, this might reflect that they encounter challenges in quickly storing and accessing new letters. This group of students may benefit from a working memory intervention with a focus on the alphabet. Moreover, the COPA allows researchers to measure the procedural aspect of orthographic processing. Therefore, utilizing this assessment in combination with other declarative measures of orthographic processing will provide a more comprehensive understanding of orthographic processing.

In sum, the COPA stands alone in its unique ability to assess orthographic processing. This separates the COPA from most measures of orthographic processing, which primarily assess orthographic knowledge and are heavily confounded by word reading and spelling ability. Despite a few low-quality items, the COPA appears to be a promising measure. In this initial analysis, the COPA appears to reflect five distinctive operating skills: rapid perception, access, differentiation, correction, and arrangement. Correlations with external variables confirmed that performances on the COPA are more strongly related to literacy outcomes than to oral language outcomes but that the COPA also demonstrates discriminant validity relative to even proximal measures of reading and spelling. The mastery probabilities and best-fitting hierarchical model indicate that four of the five skills follow a learning progression that is consistent with information processing theory and that was assumed by developers of the COPA. Finally, the study revealed ways in which minor revisions to the COPA may bring the tool even closer to the developers' goals of creating a comprehensive assessment of orthographic processing skills that will help advance literacy research, theory, and practice.

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 Iva Chen https://orcid.org/0000-0001-9845-5764
Yi-Hsin Chen https://orcid.org/0000-0002-6102-7544
Noé Erazo https://orcid.org/0000-0002-1253-5956

Supplemental Material

Supplemental material for this article is available online.

Note

1. Cognitive profiles are often represented by a series of binary codes (i.e., 0 and 1), where 1 and 0 represent mastery and non-mastery, respectively, of a given attribute. The number of codes included in a cognitive profile is determined by the number of attributes measured by the assessment. For example, with the five-attribute model of the orthographic processing assessment, a string of five digits would represent each individual's orthographic processing profile. In this case, the attributes are in the order in which we assume they will be acquired: perception, access, differentiation, addition, removal, and arrangement. A profile of "10000" would convey that this student has only mastered the perception attribute.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. MIT Press.

Apel, K. (2009). The acquisition of mental orthographic representations for reading and spelling development. *Communication Disorders Quarterly*, *31*(1), 42–52. https://doi.org/10.1177%2F1525740108325553.

Apel, K. (2011). What is orthographic knowledge? *Language, Speech, and Hearing Services in Schools*, *42*(4), 592–603. https://doi.org/10.1044/0161-1461(2011/10-0085).

Apel, K., Brimo, D., Wilson-Fowler, E. B., Vorstius, C., & Radach, R. (2013). Children develop initial orthographic knowledge during storybook reading. *Scientific Studies of Reading*, *17*(4), 286–302. https://doi.org/10.1080/10888438.2012.692742.

Apel, K., Henbest, V. S., & Masterson, J. (2019). Orthographic knowledge: Clarifications, challenges, and future directions. *Reading and Writing*, *32*(4), 873–889. https://doi.org/10.1007/s11145-018-9895-9.

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.

Barker, T. A., Torgesen, J. K., & Wagner, R. K. (1992). The role of orthographic processing skills on five different reading tasks. *Reading Research Quarterly*, *27*(4), 334–345. https://doi.org/10.2307/747673.

Berninger, V. W., Cartwright, A. C., Yates, C. M., Swanson, H. L., & Abbott, R. D. (1994). Developmental skills related to writing and reading acquisition in the intermediate grades. *Reading and Writing*, *6*(2), 161–196. https://doi.org/10.1007/bf01026911.

Berninger, V. W., Yates, C., & Lester, K. (1991). Multiple orthographic codes in reading and writing acquisition. *Reading and Writing*, *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*. (pp. 151–179). Psychology Press. https://doi.org/10.4324/9780203841211.

Castles, A., & Nation, K. (2008). Learning to be a good orthographic reader. *Journal of Research in Reading*, *31*(1), 1–7. https://doi.org/10.1111/j.1467-9817.2007.00367.x.

Çeliköz, N., Erişen, Y., & Şahin, M. (2019). Cognitive learning theories with emphasis on latent learning, gestalt and information processing theories. *Journal of Educational and Instructional Studies in the Word*, *9*(3), 18–33. https://files.eric.ed.gov/fulltext/ED598366.pdf.

Chalmers, K. A., & Burt, J. S. (2008). Phonological and semantic information in adults' orthographic learning. *Acta Psychologica*, *128*(1), 162–175. https://doi.org/10.1016/j.actpsy.2007.12.003.

Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265–289. https://doi.org/10.3102/10769986022003265.

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, Wilson, M., Irey, R. C., & Requa, M. K. (2020). Aninnovative measure of orthographic processing: Development and initial validation. *Language Testing*, *37*(3), 435–452. https://doi.org/10.1177/0265532220909310.

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.

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.

Coltheart, M. (2005). Modeling reading: The dual-route approach. In M. J. Snowling, & C. Hulme (Eds.), *The science of reading: A handbook*. (pp. 6–23). Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch1.

Compton, D. L., Gilbert, J. K., Kearns, D. M., & Olson, R. K. (2020). Using an item-specific predictor to test the dimensionality of the orthographic choice task. *Annals of Dyslexia*, *70*(2), 243–258. https://doi.org/10.1007/s11881-020-00202-0.

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.

Conrad, N. J., Harris, N., & Williams, J. (2013). Individual differences in children's literacy development: The contribution of orthographic knowledge. *Reading and Writing*, *26*(8), 1223–1239. https://doi.org/10.1007/s11145-012-9415-2.

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.

Cunningham, A. E., Perry, K. E., & Stanovich, K. E. (2001). Converging evidence for the concept of orthographic processing. *Reading and Writing: An Interdisciplinary Journal*, *14*(5/6), 549–568. https://doi.org/10.1023/A:1011100226798.

de la Torre, J. (2008). An empirically based method of Q-Matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, *45*(4), 343–362. https://doi.org/10.1111/j.1745-3984.2008.00069.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. (2011). The generalized DINA model framework. *Psychometrika*, *76*(2), 179–199. https://doi.org/10.1007/s11336-011-9207-7.

de la Torre, J., & Minchen, N. D. (2019). The G-DINA model framework. In M. von Davier, & Y.-S. Lee (Eds.), *Handbook of diagnostic classification models: Methodology of educational measurement and assessment* (pp. 155–169). Springer. https://doi.org/10.1007/978-3-030-05584-4_7.

de la Torre, J., van der Ark, L. A., & Rossi, G. (2018). Analysis of clinical data from cognitive diagnosis modeling framework. *Measurement and Evaluation in Counseling and Development*, *51*(4), 281–296. https://doi.org/10.1080/07481756.2017.1327286.

Deacon, S. H., Chen, X., Luo, Y., & Ramirez, G. (2013a). 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.

Deacon, S. H., Commissaire, E., Chen, X., & Pasquarella, A. (2013b). Learning about print: The development of orthographic processing and its relationship to word reading in first grade children in French immersion. *Reading and Writing*, *26*(7), 1087–1109. https://doi.org/10.1007/s11145-012-9407-2.

DiBello, L. V., Roussos, L. A., & Stout, W. F. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. R. Rao, S. Sinharay (Eds.), *Handbook of statistics* (vol. *26*, Psychometrics, pp. 979–1030). Amsterdam, The Netherlands: Elsevier.

Dunn, L. M., & Dunn, D. M. (2012). *Peabody picture vocabulary test (PPVT-4)*. Pearson Education Inc.

Ehri, L. C. (2005). Development of sight word reading: Phases and findings. In M. Snowling, & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 135–154). Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch8.

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.

Georgiou, G. K., Parrila, R., & Kirby, J. R. (2009). RAN components and reading development from Grade 3 to Grade 5: What underlies their relationship? *Scientific Studies of Reading*, *13*(6), 508–534. https://doi.org/10.1080/10888430903034796.

Grainger, J. (2018). Orthographic processing: A ‘mid-level’ vision of reading: The 44th Sir Frederic Bartlett lecture. *Quarterly Journal of Experimental Psychology*, *71*(2), 335–359. https://doi.org/10.1080/17470218.2017.1314515.

Hagiliassis, N., Pratt, C., & Johnston, M. (2006). Orthographic and phonological processes in reading. *Reading and Writing*, *19*(3), 235–263. https://doi.org/10.1007/s11145-005-4123-9.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*(3), 491–528. https://doi.org/10.1037/0033-295x.106.3.491.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191–210. https://doi.org/10.1007/s11336-008-9089-5.

Hsu, Y., & Tsai, T. (2007). *EZLID: A SAS® macro for local item dependence assessment*. NorthEast SAS Users Group.

Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for fusion model application to LanguEdge assessment. *Language Testing*, *26*(1), 31–73. https://doi.org/10.1177/0265532208097336.

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.

Kirby, J. R., Desrochers, A., Roth, L., & Lai, S. S. V. (2008). Longitudinal predictors of word reading development. *Canadian Psychology*, *49*(2), 103–110. https://doi.org/10.1037/0708-5591.49.2.103.

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 model: An approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement*, *41*(3), 205–237. https://doi.org/10.1111/j.1745-3984.2004.tb01163.x.

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Woodcock Johnson III normative update technical manual rolling meadows*. Riverside Publishing.

Messick, S. (1994). Stand-based score interpretaton: Establishing valid grounds for valid inferences. *ETS Research Report Series*, *1994*(2), 291–305. https://doi.org/10.1002/j.2333-8504.1994.tb01630.x.

National Early Literacy Panel & National Center for Family Literacy (2008). *Developing early literacy: Report of the national early literacy panel*. National Institute for Literacy. https://www.nichd.nih.gov/sites/default/files/publications/pubs/documents/NELPReport09.pdf.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. (NIH Publication No. 00–4769). U. S. Government Printing Office.

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* (Vol. 4, pp. 1–64). Academic Press.

Pasquarella, A., Deacon, H., Chen, B. X., Commissaire, E., & Au-Yeung, K. (2014). Acquiring orthographic processing through word reading: Evidence from children learning to read French and English. *International Journal of Disability*, *61*(3), 240–257. https://doi.org/10.1080/1034912X.2014.932579.

Paulsen, J., & Valdivia, D. S. (2021). Examining cognitive diagnostic modeling in classroom assessment conditions. *The Journal of Experimental Education*, 1–18. (online first) https://doi.org/10.1080/00220973.2021.1891008.

Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67–86). American Psychological Association. https://doi.org/10.1037/10459-000.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). John Benjamins. https://doi.org/10.1075/swll.11.14per.

Pommerich, M., & Segall, D. O. (2008). Local dependence in an operational CAT: Diagnosis and implications. *Journal of Educational Measurement*, *45*(3), 201–223. https://doi.org/10.1111/j.1745-3984.2008.00061.x.

Rahbari, N., Sénéchal, M., & Arab-Moghaddam, N. (2007). The role of orthographic and phonological processing skills in the reading and spelling of monolingual Persian children. *Reading and Writing*, *20*(5), 511–533. https://doi.org/10.1007/s11145-006-9042-x.

Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2020). *Package 'CDM'*. CDM. https://cran.r-project.org/web/packages/CDM/CDM.pdf.

Rojas, G., de la Torre, J., & Olea, J. (2012, April). *Choosing between general and specific cognitive diagnosis models when the sample size is small*. Paper presented at the meeting of the National Council on Measurement in Education, Vancouver, Canada. Pearson

Roman, A. A., Kirby, J. R., Parrila, R. K., Wade-Woolley, L., & Deacon, S. H. (2009). Toward a comprehensive view of the skills involved in word reading in Grades 4, 6, and 8. *Journal of Experimental Child Psychology*, *102*(1), 96–113. https://doi.org/10.1016/j.jecp.2008.01.004.

Rothe, J., Schulte-Körne, G., & Ise, E. (2014). Does sensitivity to orthographic regularities influence reading and spelling acquisition? A 1-year prospective study. *Reading and Writing*, *27*(7), 1141–1161. https://doi.org/10.1007/s11145-013-9479-7.

Roussos, L. A., DiBello, L. V., Stout, W., Hartz, S. M., Henson, R. A., & Templin, J. L. (2007). The fusion model skills diagnosis system. In J. P. Leighton, & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications*. (pp. 275–318). Cambridge University Press. https://doi.org/10.1017/CBO9780511611186.

SAS Institute, Inc. (2020). *SAS/STAT 15.2® user's guide: The IRT procedure*.

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.

Sen, S., & Cohen, A. S. (2021). Sample size requirements for applying diagnostic classification models. *Frontiers in Psychology*, *11*, 621251. https://doi.org/10.3389/fpsyg.2020.621251.

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.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*(2), 95–129. https://doi.org/10.1006/jecp.1998.2481.

Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology*, *87*(4), 267–298. https://doi.org/10.1016/j.jecp.2004.01.001.

Simon, H. A. (1978). Information-processing theory of human problem solving. In W. K. Estes (Ed.), *Handbook of learning & cognitive processes: V. Human information*. (p. 271–295). Lawrence Erlbaum.

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.

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.

Treiman, R. (1993). *Beginning to spell: A study of first grade children*. Oxford University Press. https://doi.org/10.1093/oso/9780195062199.001.0001.

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*(2), 287–307. https://doi.org/10.1348/000711007x193957.

Wagner, R. K., & Barker, T. A. (1994). The development of orthographic processing ability. In V. W. Berninger (Ed.), *The varieties of orthographic knowledge I: Theoretical and developmental issues*. (pp. 243–276). Kluwer. https://doi.org/10.1007/978-94-017-3492-9_8.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*(2), 192–212. https://doi.org/10.1037/0033-2909.101.2.192.

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.