Please use the following citation when referencing this work:

McGill, R. J., Conoyer, S. J., & Fefer, S. (2018). Elaborating on the linkage between cognitive and academic weaknesses: Using diagnostic efficiency statistics to inform PSW assessment. *School Psychology Forum, 12,* 118-132.

Elaborating on the Linkage Between Cognitive and Academic Weaknesses: Using Diagnostic Efficiency Statistics to Inform PSW Assessment

Ryan J. McGill

William & Mary

Sarah J. Conoyer

Southern Illinois University Edwardsville

Sarah Fefer

University of Massachusetts Amherst

Author notes

Correspondence concerning this article should be addressed to Ryan J. McGill, William & Mary School of Education, P.O. Box 8795, Williamsburg, VA 23188. E-Mail: rmcgill@wm.edu

Abstract

Within the school psychology literature, it is frequently asserted that deficits in cognitive processing are a defining characteristic of children with academic dysfunction and establishing links between relevant cognitive and academic weaknesses is a core pillar of patterns of strengths and weaknesses assessment models. Accordingly, the present study employed diagnostic utility statistics to determine whether the presence of a significant cognitive weakness accurately distinguishes between participants with and without academic weaknesses ages 7:0 to 18:11 in the linked Kaufman Assessment Battery for Children-Second Edition (KABC-II; Kaufman & Kaufman, 2004a) and Kaufman Test of Educational Achievement-Second Edition (KTEA-II; Kaufman & Kaufman, 2004b) normative sample. Sensitivity and positive predictive values for the CHC-related broad abilities ranged from low to moderate indicating that a cognitive weakness may not be strong a *rule-in* sign for the presence of an academic weakness. Conversely, specificity and negative predictive values were consistently high indicating that, absent a cognitive weakness, it is unlikely that an individual will present with an academic weakness. Potential implications for clinical assessment are discussed.

Keywords: Evidence-Based Assessment, PSW, Diagnostic Validity

Elaborating on the Linkage Between Cognitive and Academic Weaknesses: Using Diagnostic Efficiency Statistics to Inform PSW Assessment

Over the course of the last 20 years, the Cattell-Horn-Carroll theory (CHC; Schneider & McGrew, 2018) has emerged as the consensus model for understanding the structure of intellectual abilities in psychology and education and CHC has been used by assessment researchers to investigate how various cognitive abilities may relate to academic learning. For example, several early studies investigating CHC cognitive-achievement relationships found differential predictive effects for various cognitive abilities and reading, mathematics, and written language achievement across the age span (Evans, Floyd, McGrew, & LeForgee, 2001; Floyd, Evans, & McGrew, 2003; Floyd, McGrew, & Evans, 2008), findings that have also been replicated in more recent work by Cormier and colleagues (Cormier, Bulut, McGrew, & Frison, 2016; Cormier, Bulut, McGrew, & Singh, 2017; Cormier, McGrew, Bulut, & Funamoto, 2017). In addition to being widely disseminated within the school and educational psychology literatures (~597 citations), these studies have been instrumental in attempts to establish links between deficits in cognitive and academic functioning that may aide in the identification of specific learning disability (SLD). For example, McGrew and Wendling (2010) conducted a literature review to identify which CHC abilities may hold promise as early screening markers or "pattern indicators" (p. 652) of a potential learning disorder. After reviewing 19 cognitive-achievement relationship studies, they concluded that a comprehensive assessment of CHC broad and narrow abilities may provide practitioners with the most diagnostic and instructionally relevant information for different academic domains. Even so, the assumption that a deficit in cognitive processing underlies a learning disorder is an idea that predates the advent of CHC. The federal definition of SLD, as presently codified in the regulations that implements Part B of

IDEA, has long implied that the etiology of the condition is a “disorder in one or more of the basic psychological processes” (34 C.F.R. § 300.8(c)(10)) and numerous studies have found that individuals with SLD present with statistically significant cognitive processing differences when compared to typically achieving students (e.g., Johnson et al. 2010; Feifer et al., 2014; Toffalini, Giofre, & Cornoldi, 2017).

Based on these findings, practitioners have been encouraged to search for potential diagnostic meaning in the variability that is observed in an individual’s profiles of cognitive-achievement scores (e.g., Flanagan, Ortiz, & Alfonso, 2013), an approach to clinical assessment referred to as patterns of strengths and weaknesses (PSW). To date, several PSW models have been operationalized including a CHC-inspired approach known as the Dual/Discrepancy Consistency Model (DD/C; Flanagan et al., 2018). Briefly, the DD/C method of SLD identification requires users to (a) identify an academic weakness in one or more composite scores, (b) determine that the academic weakness(es) is not primarily due to exclusionary factors, (c) identify a cognitive weakness in one or more composite scores, and (d) determine whether the resulting strengths and weaknesses reflect a confirmatory PSW pattern. The discrepancies between strengths and weaknesses (i.e., composite scores) must be judged to be statistically and/or clinically significant[1] as well as theoretically consistent. Whereas a software program has been developed to aide practitioners when making these determinations (e.g., Flanagan, Ortiz, & Alfonso, 2017), the DD/C method is not dependent on its use.

As should be evident, the DD/C method (as well as every other PSW approach) assumes that a deficit in cognitive processing is a cardinal characteristic of academic dysfunction and that

[1] In doing so, practitioners may elect to employ intuitive psychometric cut points (e.g., standard score of < 85) or adopt whatever criteria they deem appropriate based on their clinical judgement (Flanagan et al., 2018).

establishing this linkage is clinically useful (Taylor, Miciak, Fletcher, & Francis, 2017). Whereas Fiorello, Flanagan, and Hale (2014) suggest that these methods are “empirically supported” (p. 55), other scholars have questioned whether these assessment procedures should be regarded by practitioners as pseudoscience (e.g., Kranzler et al., 2016b).

Evidence-Based Practice and the Validation of Clinical Procedures

Consonant with the broader evidence-based practice (EBP) movement, it is important to use science to guide assessment. A scientific approach to psychological assessment requires that claims about the merits of any assessment measure, software tool, or set of interpretive procedures be framed in such a way that they can be subjected to rigorous empirical investigation by proponents and critics of these approaches. In this way, school psychologists must be mindful of the principle of *falsifiability* (i.e., a claim or belief is stated in such a way that it can be contradicted by disconfirming evidence [Popper, 1962]) and adopt the attitude of a professional skeptic wherein the burden of proof rests with the person making the claim. Put simply, if a claim cannot be falsified, it cannot be proven right or wrong and is therefore by definition pseudoscience.

It is also important to keep in mind that not all evidence is equal and practitioners have been encouraged to base practice decisions primarily on the best available research evidence (e.g., American Psychological Association Presidential Task Force on Evidence-Based Practice, 2006). However, determining whether an approach to clinical assessment is *evidence-based* is complicated. As noted by Youngstrom, Choukas-Bradley, Calhoun, and Jensen-Doss (2015), thousands of articles, blogs, and advertisements compete for our attention, with only a small minority of these resources providing information that is both scientifically valid and clinically relevant.

Whereas these questions are ultimately empirical, we are mindful of the fact that few practitioners have the time to skim hundreds of articles to find the one or two gems that may be helpful for adjudicating these matters. As a potential remedy, Youngstrom et al. (2015) have advanced a framework for clinical assessment inspired by the evidence-based medicine movement which they term evidence-based assessment (EBA). When evaluating the literature, the EBA model encourages practitioners to focus primarily on empirical studies that provide insight about the probability that our assessment procedures yield beneficial information for diagnostic and treatment applications. As many assessment-related decisions in school psychology are rendered in the presence of uncertainty (i.e., measurement error), it is important to ascertain the degree to which our assessment procedures help improve our chances of being right. That is, it is one thing to *suggest* that a cognitive weakness accurately predicts an achievement weakness and thus should be a focal point of SLD assessment, and another to provide the appropriate empirical evidence (e.g., diagnostic efficiency statistics) needed to support this claim.

Establishing Relevant Links Between Cognitive and Academic Weaknesses

Proponents of these PSW assessment frequently cite studies illustrating the relationships between cognitive abilities and achievement and the different ways that children suspected of having a learning disorder have discrepant cognitive-achievement profiles (e.g., Cormier, Bulut, McGrew, & Frison, 2016; Feifer et al., 2014; McGrew & Wendling, 2010), and these findings are now well established. However, statistically significant group differences are singularly insufficient for diagnostic use due to the significant overlap that is frequently observed across the distributions for these groups. As a result, decisions about individuals are more appropriately addressed with a diagnostic utility approach and the statistics that are generated from these analyses (Beaujean, 2017). That is, it is important for practitioners who may be utilizing PSW methods to know how well a cognitive weakness *actually* predicts an academic weakness in order to weight this information appropriately (Youngstrom & Van Meter, 2016). Unfortunately, these data have yet to be reported in available PSW assessment materials (e.g., Flanagan, Ortiz, & Alfonso, 2017; Ventura County SELPA, 2017).

In the only investigation that could be located, Kranzler et al. (2016a), reported diagnostic efficiency statistics associated with the application of DD/C model procedures to assessment data for 300 participants in the Woodcock-Johnson III normative sample. Specifically, they investigated the base rate (i.e., portion of participants identified as having SLD), positive predictive value (PPV; proportion of cases where individuals with the condition have a positive test result), and negative predictive value (NPV; proportion of cases where individuals without the condition have a negative test result). Mean specificity (i.e., ability of a test to correctly identify individuals without SLD) and NPVs were 92% and 89% across CHC cognitive abilities and achievement domains indicating that the absence of a cognitive weakness was very accurate in detecting what they deemed to be "true negatives" (individuals without an achievement weakness). On the other hand, sensitivity (i.e., ability of a test to correctly identify individuals with a condition) and PPV estimates (21% and 34% respectively) were quite low, indicating that the presence of a cognitive weakness may not be very useful at accurately identifying what they classified as "true positives" (individuals with an achievement weakness).

In spite of these findings, the procedures that were employed have been criticized. In a response to Kranzler and colleagues, Flanagan and Schneider (2016) noted that the use of single measures (subtests) was not consistent with the DD/C method which emphasizes the use of composite scores. Thus, better outcomes may have been obtained if these indices had been used instead of the subtests. They also took issue with the nomenclature employed in the study, in particular, conflating cross-battery assessment with DD/C. Given the potential implications of these findings for those currently engaged in PSW assessment or those considering whether these procedures should be adopted in clinical practice, it would be valuable to examine whether these results generalize to other measures and scores to determine if the presence of a cognitive weakness functions as a useful sign for the presence of academic dysfunction at the level of the individual.

Purpose of the Present Investigation

Although a popular practice in various jurisdictions, there has been a debate whether the fundamental assumptions that undergird PSW assessment are tenable (e.g., Beaujean, Benson, McGill, & Dombrowski, 2018; McGill, Styck, Palomres, & Hass, 2016). Accordingly, the present study employed diagnostic utility statistics to determine whether the presence of a significant cognitive weakness accurately distinguishes between participants with and without academic weaknesses ages 7:0 to 18:11 in the linked Kaufman Assessment Battery for Children-Second Edition (KABC-II; Kaufman & Kaufman, 2004a) and Kaufman Test of Educational Achievement-Second Edition (KTEA-II; Kaufman & Kaufman, 2004c) normative sample. That is, what is the probability that a randomly selected student with a deficit in cognitive processing also has a deficit in academic functioning? As Flanagan & Schneider (2016) suggest that the presence of a cognitive weakness increases the *probability* of having an academic weakness, it stands to reason that the indices associated with the prediction of this condition should exceed chance levels. It is believed that the present investigation will provide school psychologists with important information about the potential utility of PSW assessment methods.

Method

Participants were children and adolescents ages 7:0 to 18:11 ($N = 2{,}025$) from the KABC-II/KTEA-II norming sample. The standardization was nationally stratified across age, sex, race/ethnicity, parent educational level (as a proxy for socioeconomic status), and region based on 2001 U.S. census estimates. Demographic characteristics for the present sample are reported in Table 1. Inspection of the values reported in Table 1 as well as those associated with the broader normative sample in the KABC-II manual (Kaufman & Kaufman, 2004b) reveal close correspondence to reference parameters.

Measurement Instruments

What follows is a brief description of the measurement instruments used in the present study. A general description of each test's theoretical structure is provided as well as a brief summary of relevant psychometric.

Kaufman Assessment Battery for Children-Second Edition (KABC-II). The KABC-II is an individually administered test of cognitive abilities for children and adolescents ages 3:0 to 18:11 years. It is composed of 10 core subtests which combine to form five group-specific composite scores that are aligned with CHC theory: Crystallized Ability (Gc), Fluid Reasoning (Gf), Visual Processing (Gv), Long-Term Memory (Glr), and Short-Term Memory (Gsm). To estimate general intelligence, a full scale Fluid-Crystallized (FCI) composite score is also provided though that score was not the focus of the current investigation. All composite scores are expressed as standard scores with a mean of 100 and a standard deviation of 15. Mean internal consistency estimates for the included ages in this study ranged from .88 to .93 for the composite scores. Extensive normative and psychometric data can be found in the KABC-II manual.

Kaufman Test of Educational Achievement-Second Edition (KTEA-II). The KTEA-II is an individually administered test of educational achievement designed to measure four academic domains: Reading, Mathematics, Written Language, and Oral Language. It is comprised of 14 subtests that combine to form four domain composite scores and a full scale total achievement score. All scores are expressed as standard scores with a mean of 100 and a standard deviation of 15. Mean internal consistency estimates for the included ages in this study ranged from .90 to .97 for the composite scores that were assessed. Extensive technical information can be found in the KTEA-II manual.

Data Analyses

Data analyses proceeded in several steps. First, the normative data were inspected to identify the prevalence of cognitive and academic weaknesses across study participants. A cognitive weakness was defined as a standard score < 90 or < 85 on any of the five CHC-related broad ability composites from the KABC-II. An achievement weakness was defined as a standard score < 90 or < 85 on any of four achievement composites (Reading, Mathematics, Written Language, and Oral Language) from the KTEA-II. We examined results across these two particular cut points because they are prominently featured in the PSW literature (e.g., Flanagan, Ortiz, & Alfonso, 2013). Binary coding of these data was completed using SPSS version 23. Results of these analyses were then entered into a 2x2 contingency table to facilitate the calculation of diagnostic efficiency statistics.

Several diagnostic efficiency statistics were calculated across achievement domains and cutoff criteria using the *Diagnostic Utility* program by Watkins (2002). These indices included sensitivity (true positive [i.e., cognitive and academic weaknesses present]) and specificity (true negative [i.e., no cognitive or academic weaknesses]). Sensitivity and specificity estimates can be combined into a single number called the likelihood ratio (LR). Positive and negative likelihood ratios (LR) indicate whether a test is useful as a *rule-in* or *rule-out* indicator for a condition. According to Streiner (2003), LR^+ values exceeding 10 are indicative of a quality rule-in test and LR^- values approaching zero are preferred for a strong rule-out test. Alternatively, positive and negative predictive values refer to the proportion of individuals who are correctly classified as having or not having an academic weakness based on the presence or absence of a cognitive weakness respectively. Additionally, area under the curve (AUC) estimates from receiver operating characteristic (ROC) curve analyses were also calculated. In a ROC curve, the true positive and false positive rates are plotted on a graph and the AUC represents the distance between the curve and a diagonal line representing chance classification. Higher AUC values reflect a more accurate test. The following guidelines for interpreting AUC have been recommended by Youngstrom (2014): .50-.70 (low accuracy), .70-.90 (medium accuracy), .90-1.00 (high accuracy).

Results

The means and standard deviations of KABC-II/KTEA-II scores for sample participants based on the conditional status of academic functioning are reported in Table 2. Mean composite scores were statistically different and lower for all CHC broad abilities for participants with weaknesses in Reading (d = -0.84 to -1.47), Mathematics (d = -0.86 to -1.33), Written Language (d = -0.74 to -1.17), and Oral Language (d = -0.81 to -2.95) when compared to the scores for participants without weaknesses in the same domains. All of these values represent large effect sizes using the conventional guidelines furnished by Cohen (1988).

Table 3

Diagnostic Efficiency Statistics for Suggested Patterns of Strengths and Weaknesses Cognitive Standard Score Thresholds Associated with a Weakness in Reading

Broad Ability	AUC	Sensitivity	Specificity	LR+	LR-	PPV	NPV
Crystallized Ability							
< 90	.745	.592	.898	5.82	0.45	.667	.864
< 85	.727	.540	.915	6.36	0.50	.579	.901
Fluid Reasoning							
< 90	.682	.496	.868	3.76	0.58	.564	.833
< 85	.684	.462	.907	4.99	0.59	.519	.866
Visual Processing							
< 90	.647	.455	.839	2.82	0.64	.493	.817
< 85	.631	.378	.884	3.26	0.70	.414	.867
Short-Term Memory							
< 90	.648	.459	.838	2.84	0.64	.494	.818
< 85	.653	.439	.868	3.33	0.64	.419	.877
Long-Term Memory							
< 90	.694	.536	.852	3.63	0.54	.556	.842
< 85	.652	.406	.899	4.03	0.65	.466	.874

Note. AUC = area under curve, LR+ = likelihood ratio for positive test results LR- = likelihood ratio for negative test results, PPV = positive predictive value, NPV= negative predictive value.

Results of the diagnostic efficiency analyses are shown by academic area in Tables 3-6. Inspection of these tables reveal relatively consistent results across the academic domains. Overall correct classification (a.k.a., “hit rate”) ranged from 72.6% to 86.9% with a mean of 78.9% due mainly to the high rate of true negatives.

Results for the diagnostic efficiency analyses for Reading are shown in Table 3. AUC values, which improve as sensitivity increases and the false positive rate (1—specificity) decreases, ranged from low to moderate (.631 to .745) indicating marginal discriminant validity across the CHC domains. Sensitivity (proportion of true positives) was moderate with a mean of .476. A sensitivity value of 47.6% indicates that, on average, less than one in two students with an academic weakness in Reading will present with a concomitant weakness in one or more

CHC broad abilities as measured by the KABC-II. As a consequence, PPVs, which represent the accuracy of a positive test result, mostly hovered around chance levels and odds ratio estimates were low to moderate indicating that the presence of a cognitive weakness is not a robust rule-in indicator for the presence of a Reading weakness. On the contrary, mean specificity (85.5%) and NPV estimates (85.9%) were high indicating that absent a cognitive weakness it is very unlikely that an individual will have present with a weakness in Reading as measured by the KTEA-II.

Table 4

Diagnostic Efficiency Statistics for Suggested Patterns of Strengths and Weaknesses Cognitive Standard Score Thresholds Associated with a Weakness in Mathematics

Broad Ability	AUC	Sensitivity	Specificity	LR+	LR-	PPV	NPV
Crystallized Ability							
< 90	.697	.524	.870	4.04	0.54	.573	.846
< 85	.691	.487	.895	4.66	0.57	.471	.901
Fluid Reasoning							
< 90	.700	.526	.874	4.18	0.54	.582	.847
< 85	.681	.462	.899	4.61	0.59	.468	.897
Visual Processing							
< 90	.677	.503	.852	3.39	0.58	.530	.836
< 85	.663	.437	.889	3.95	0.63	.430	.892
Short-Term Memory							
< 90	.656	.473	.838	2.93	0.62	.493	.827
< 85	.655	.448	.863	3.27	0.63	.385	.891
Long-Term Memory							
< 90	.656	.481	.830	2.84	0.62	.486	.828
< 85	.634	.381	.888	3.40	0.69	.394	.882

Note. AUC = area under curve, LR+ = likelihood ratio for positive test results LR- = likelihood ratio for negative test results, PPV = positive predictive value, NPV= negative predictive value.

Results for the diagnostic efficiency analyses for Mathematics are reported in Table 4. AUC values ranged from low to moderate (.634 to .700) indicating marginal discriminant validity across the CHC domains. Sensitivity was low to moderate with a mean of .472. As a consequence, PPVs and odds ratio estimates were also moderate indicating that the presence of a cognitive weakness is not a robust rule-in indicator for the presence of a Mathematics weakness. On the contrary, mean specificity (86.9%) and NPV estimates (86.4%) were high indicating that absent a cognitive weakness it is very unlikely that an individual will have present with a weakness in Mathematics as measured by the KTEA-II.

Table 5

Diagnostic Efficiency Statistics for Suggested Patterns of Strengths and Weaknesses Cognitive Standard Score Thresholds Associated with a Weakness in Written Language

Broad Ability	AUC	Sensitivity	Specificity	LR+	LR-	PPV	NPV
Crystallized Ability							
< 90	.700	.523	.874	4.18	0.54	.587	.845
< 85	.684	.474	.894	4.49	0.58	.468	.896
Fluid Reasoning							
< 90	.678	.492	.865	3.65	0.58	.553	.833
< 85	.682	.463	.901	4.70	0.59	.480	.895
Visual Processing							
< 90	.628	.427	.828	2.49	0.69	.458	.810
< 85	.623	.369	.877	3.02	0.71	.372	.876
Short-Term Memory							
< 90	.644	.453	.835	2.74	0.59	.482	.818
< 85	.645	.430	.861	3.09	0.66	.378	.885
Long-Term Memory							
< 90	.666	.496	.837	3.04	0.60	.508	.830
< 85	.639	.389	.890	3.56	0.68	.411	.881

Note. AUC = area under curve, LR+ = likelihood ratio for positive test results LR- = likelihood ratio for negative test results, PPV = positive predictive value, NPV= negative predictive value.

Results for the diagnostic efficiency analyses for Written Language are reported in Table 5. AUC values ranged from low to moderate (.623 to .700) indicating marginal discriminant validity across the CHC domains. Sensitivity was low to moderate with a mean of .451. As a consequence, PPVs and odds ratio estimates were also moderate indicating that the presence of a cognitive weakness is not a robust rule-in indicator for the presence of a Written Language weakness. On the contrary, mean specificity (86.6%) and NPV estimates (85.6%) were high indicating that absent a cognitive weakness it is very unlikely that an individual will have present with a weakness in Written Language as measured by the KTEA-II.

Table 6

Diagnostic Efficiency Statistics for Suggested Patterns of Strengths and Weaknesses Cognitive Standard Score Thresholds Associated with a Weakness in Oral Language

Broad Ability	AUC	Sensitivity	Specificity	LR+	LR-	PPV	NPV
Crystallized Ability							
< 90	.752	.614	.890	5.60	0.43	.632	.882
< 85	.722	.516	.929	7.29	0.52	.563	.915
Fluid Reasoning							
< 90	.688	.513	.862	3.73	0.56	.533	.852
< 85	.689	.454	.922	5.89	0.59	.478	.915
Visual Processing							
< 90	.638	.448	.828	2.61	0.66	.444	.830
< 85	.639	.391	.887	3.48	0.68	.352	.903
Short-Term Memory							
< 90	.650	.469	.831	2.78	0.63	.460	.836
< 85	.663	.424	.901	4.30	0.63	.401	.909
Long-Term Memory							
< 90	.677	.517	.836	3.16	0.57	.492	.849
< 85	.678	.461	.894	4.37	0.60	.405	.914

Note. AUC = area under curve, LR+ = likelihood ratio for positive test results LR- = likelihood ratio for negative test results, PPV = positive predictive value, NPV= negative predictive value.

Results for the diagnostic efficiency analyses for Oral Language are reported in Table 6. AUC values ranged from low to moderate (.639 to .752) indicating marginal discriminant validity across the CHC domains. Sensitivity was moderate with a mean of .481. As a consequence, PPVs and odds ratio estimates were also moderate indicating that the presence of a cognitive weakness is not a robust rule-in indicator for the presence of an Oral Language weakness. On the contrary, mean specificity (87.8%) and NPV estimates (88.1%) were high indicating that absent a cognitive weakness it is very unlikely that an individual will have present with a weakness in Written Language as measured by the KTEA-II.

Discussion

The current study employed diagnostic efficiency statistics to determine whether cognitive weaknesses accurately distinguish between individuals with academic deficits in the KABC-II/KTEA-II normative sample. As establishing this linkage is a critical step in PSW assessment, the present results provide clinicians with important information that can be used to guide clinical assessment. This is one of the first investigations to examine whether the diagnostic utility statistics reported by Kranzler et al. (2016a) generalize to other measurement instruments such as the KABC-II. Since school psychologists who engage in PSW analyses utilize a wide variety of instrumentation, it is believed that the present results will be a useful addition to the practice-based literature.

Across all academic domains, the discriminant validity of weaknesses in various CHC-related broad abilities was moderate, mostly hovering around chance levels. Whereas these results suggest that the presence of a cognitive weakness does not function as strong *rule-in* test for the presence of an academic deficit, the absence of a cognitive weakness appears to be a very robust *rule-out* indicator for this condition. As practitioners employ assessments primarily for the purposes of validating relevant clinical concerns, these sensitivity and positive predictive indices indicate that the presence of a cognitive weakness may only moderately improves the chances of accurately identifying children and adolescents with academic weaknesses [2]. Nevertheless, it should be noted that our sensitivity and PPVs were noticeably higher than those reported by Kranzler and colleagues (2016a). We suggest this is likely the result of the fact that Kranzler et

---

[2] The results of the present study support the use of a more liberal threshold (< 90) as it produced sensitivity and PPV estimates that were consistently < 85. Interestingly, the imposition of a different cut point impacted the ordering of the cognitive predictors in the Mathematics and Written Language models.

al. imposed additional parameters on the data consistent with DD/C procedures which may lower prevalence rates and thus degrade sensitivity and PPVs (Coombs, Dawes, & Tversky, 1970). As a result, the estimates furnished in the present study likely represent upper-bound estimates for these effects.

In terms of the relative importance of CHC abilities, weaknesses in Crystallized Ability and Fluid Reasoning were consistently the strongest predictors of academic dysfunction which is not surprising given their relationship to general intelligence. Yet, practitioners have been warned that a primary weakness in Crystallized Ability may not reflect a confirmatory PSW pattern given the conceptual overlap of that construct with virtually every area of achievement (Flanagan & Alfonso, 2015). Interestingly, the patterning of the diagnostic efficiency statistics reflecting the individual CHC abilities that were most associated with a weakness in various areas of achievement did not cohere with the interpretive recommendations in some PSW guidebooks. For example, in the Ventura County PSW manual (Ventura County SELPA, 2017), users are provided with a matrix of processing-achievement relationships to help facilitate the identification of confirmatory PSW patterns in assessment data[3]. The strength of these relationships are evaluated on a five-point scale after appraising the "research base of processing-achievement relations" (p. B8). In that document, it is suggested that Long-Term Storage and Retrieval is one of the strongest predictors (4; highest rating) for both of the math abilities measured by the Mathematics composite in the present study. However, the present results indicate that a weakness in Long-Term Storage and Retrieval was the worst predictor of a math weaknesses among the five CHC abilities that were examined indicating that the use of these and

[3] As noted by Miciak et al. (2014), although all PSW models indicate that cognitive-achievement deficits should be theoretically related, it remains unclear which academic and cognitive skills are linked by theory.

other related anecdotal resources should be used with caution to make high stakes decisions about individuals.

Clinical Efficiency

When a test significantly outperforms the prediction of an external criterion using *a priori* base rates, the test is said to be "clinically efficient" and several indices for calculating clinical efficiency are available. Of these, the Dawes condition (1962) is the most straightforward and states that it is best to use tests over the base rate only when PPV exceeds .50 (i.e., chance). According to Bokhari and Hubert (2015), if PPV is less than .50, if one obtains a positive test result, "there is a higher probability that the person does not have "it" than they do" (p. 746). Applying this criterion to the present study, only 35% of the PPV estimates exceeded this criterion. Although these assessments only represent one pillar within the broader PSW approach, practitioners are encouraged to carefully consider that when information that is diagnostically useful is infused with information that is of marginal value for predicting an outcome, a *dilution effect* can occur wherein the overall utility of a decision-making model is degraded regardless of when the relatively weaker piece of evidence is considered in the evaluative process (Nisbett, Zukier, & Lemley, 1981). From an EBA perspective, the present results illustrate well that the presence of a cognitive weakness is singularly insufficient for making inferences about the presence of an academic disorder. Even so, it remains unclear what types of additional information are needed to increase the posterior probability of correctly diagnosing SLD within the PSW assessment framework.

Limitations

Whereas the present study overcame previously noted limitations by using composite scores from a different ability measure (KABC-II), this study is not without limitations that should be considered when interpreting the results. First, even though we used standard procedures and terminology associated with the diagnostic efficiency approach, there are areas in which their literal application is problematic. For example, if a participant has a cognitive weakness but not an academic weakness, they would technically be regarded as a "false-positive" in this context. As noted by Flanagan and Schneider (2016), there are many potential reasons why cognitive deficits may not lead to academic deficits for some individuals as "cognitive abilities are causally related to academic abilities, but the causal relationship is of moderate size, and only probabilistic, not deterministic" (p. 141). Accordingly, we have focused most of our discussion on the indices associated with the probability of this outcome to avoid confusion with a terminal diagnostic outcome (i.e., SLD). Again, we reiterate that the present investigation focuses only on examining the linkage between cognitive weaknesses and academic weaknesses which is one step, albeit an important one in the broader PSW assessment process. To be clear, these results are not instructive for determining whether any particular PSW model is useful at accurately diagnosing SLD. Nevertheless, it is important to note that specific PSWs confirming or disconfirming the presence of SLD have yet to be established (Mather & Schneider, 2015). Until these matters are clarified, we believe that the dissemination and use of these methods will continue to engender controversy. Future research examining whether these results are invariant across relevant demographic variables (e.g., gender, race/ethnicity) would also be instructive.

Implications for Practice

Establishing links between cognitive and academic weaknesses is an important component of PSW assessment and users are invited to speculate about the mechanisms that may underlie these phenomena. The present results illustrate well that the many of these associations may not be as strong as they are commonly portrayed in our field. As a result of this "false-positive paradox" it is difficult to envision a scenario in which clinicians can confidently ascribe meaning to these observations when they occur. As a result, practitioners are encouraged to consider forgoing adoption of these methods until compelling empirical evidence is furnished to implicate their use for high stakes diagnostic and treatment decisions. Given the best available research, we believe this is a defensible position from an EBA perspective.

Additional Resources

A number of helpful resources relating to evidence-based assessment and clinical decision-making have been developed by Dr. Eric Youngstrom. These include an assessment care package and links to several journal articles describing the broader EBA framework. Please see http://ericyoungstrom.web.unc.edu/

Additionally, the Texas Center for Learning disabilities at the University of Houston also has many helpful resources regarding learning disability assessment and identification plus links to numerous empirical studies that have evaluated the technical adequacy of PSW and other-related cognitive discrepancy models: https://www.texasldcenter.org/.

References

American Psychological Association Presidential Task Force on Evidence-based Practice (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271-285. doi: 10.1037/0003-066X.61.4.271

Beaujean, A. A. (2017). Commentary on "Strengths and Weaknesses in the Intellectual Profile of Different Subtypes of Specific Learning Disorder" (Toffalini et al., 2017). *Clinical Psychological Science, 5,* 874-877. doi: 10.1177/2167702617708629

Beaujean, A. A., Benson, N. F., McGill, R. J., & Dombrowski, S. C. (2018). A misuse of IQ scores: Using the Dual Discrepancy/Consistency Model for identifying specific learning disabilities. *Journal of Intelligence, 6 (36),* 1: 1-25. doi: 10.3390/jintelligence6030036

Bokhari, E., & Hubert, L. (2015). A new condition for assessing the clinical efficiency of a diagnostic test. *Psychological Assessment, 27,* 745-754. doi: 10.1037/pas0000093

Cohen , J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Psychology Press.

Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: An elementary introduction*. Englewood Cliffs, NJ: Prentice-Hall.

Cormier, D. C., Bulut, O., McGrew, K. S., & Frison, J. (2016). The role of Cattell-Horn-Carroll (CHC) cognitive abilities in predicting writing achievement during the school-age years. *Psychology in the Schools*, *53*, 787-803. doi: 10.1002/pits.21945

Cormier, D. C., Bulut, O., McGrew, K. S., & Singh, D. (2017). Exploring the relations between Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement. *Applied Cognitive Psychology, 31,* 530-538. doi: 10.1002/acp.3350

Cormier, D. C., McGrew, K. S., Bulut, O., & Funamoto, A. (2017). Revisiting the relations between the WJ-IV measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during the school-age years. *Journal of Psychoeducational Assessment, 35,* 731-754. doi: 10.1177/0734282916659208

Dawes, R. M. (1962). A note on base rates and psychometric efficiency. *Journal of Consulting Psychology, 26,* 422-424. doi: 10.1037/h0044612

Evans, J. J., Floyd, R. G., McGrew, K. S., & LeForgee, M. H. (2001). The relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review*, *31*, 246-262. Retrieved from http://www.nasponline.org

Feifer, S. G., Nader, R. G., Flanagan, D. P., Fitzer, K. R., & Hicks, K. (2014). Identifying specific reading subtypes for effective educational remediation. *Learning Disabilities: A Multidisciplinary Journal, 20,* 18-30.

Flanagan, D. P., & Alfonso, V. C. (2015, February). *The SLD pattern of strengths and weaknesses and the WJ-IV*. Mini-skills workshop presented at the annual meeting of the National Association of School Psychologists, Orlando, FL.

Flanagan, D. P., Alfonso, V. C., Sy, M. C., Mascolo, J. T., McDonough, E. M., & Ortiz, S. O. (2018). Dual discrepancy/consistency operational definition of SLD: Integrating multiple data sources and multiple data gathering methods. In V. C. Alfonso & D. P. Flanagan (Eds.), *Essentials of specific learning disability identification* (2nd ed.; pp. 431-476). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2017). *Cross-Battery Assessment Software System 2.0 (X-BASS)*. Hoboken, NJ: Wiley.

Flanagan, D. P., & Schneider, W. J. (2016). Cross-battery assessment? XBA PSW? A case of mistaken identity: A commentary on Kranzler and colleagues' "Classification agreement analysis of cross-battery assessment in the identification of specific learning disorders in children and youth." *International Journal of School and Educational Psychology, 4,* 137-145. doi: 10.1080/21683603.2016.1192852

Fiorello, C. A., Flanagan, D. P., & Hale, J. B. (2014). Response to the special issue: The utility of the pattern of strengths and weaknesses approach. *Learning Disabilities: A Multidisciplinary Journal, 20,* 87-91.

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools*, *40*, 155-171. doi: 10.1002/pits.10083

Floyd, R. G., McGrew, K. S., & Evans, J. J. (2008). The relative contributions of the Cattell-Horn-Carroll cognitive abilities in explaining writing achievement during childhood and adolescence. *Psychology in the Schools*, *45*, 132-144. doi: 10.1002/pits.20284

Individuals with Disabilities Education Act regulations, 34 C.F.R. § 300.8(c)(10) *et seq*.

Johnson, E. S., Humphrey, M., Mellard, D. F., Woods, K., & Swanson, H. L. (2010). Cognitive processing deficits and students with specific learning disabilities: A selective meta-analysis of the literature. *Learning Disability Quarterly, 33,* 3-18. doi: 10.1177/073194871003300101

Kaufman, A. S., & Kaufman, N. L. (2004a). *Kaufman Assessment Battery for Children-Second Edition*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Kaufman Assessment Battery for Children-Second Edition manual*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004c). *Kaufman Test of Educational Achievement-Second Edition*. Circle Pines, MN: American Guidance Service.

Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016a). Classification agreement analysis of cross-battery assessment in the identification of specific learning disorders in children and youth. *International Journal of School & Educational Psychology, 4,* 124-136. doi: 10.1080/21683603.2016.1155515

Kranzler, J. H., Floyd, R. G., Benson, N. Zaboski, & Thibodaux, L. (2016b). Cross-battery assessment pattern of strengths and weaknesses approach to the identification of specific learning disorders: Evidence-based practice or pseudoscience? *International Journal of School and Educational Psychology, 4,* 146-157. doi: 10.1080/21683603.2016.1192855

Mather, N., & Schneider, D. (2015). The use of intelligence tests in the diagnosis of specific reading disability. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.), *Handbook of Intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 415-434). New York: Springer.

McGill, R. J., Styck, K. S., Palomares, R. S., & Hass, M. R. (2016). Critical issues in specific learning disability identification: What we need to know about the PSW model. *Learning Disability Quarterly, 39,* 159-170. doi: 10.1177/0731948715618504

McGrew, K. S., & Wendling, B. J. (2010). Cattell-Horn-Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools*, *47*, 651-675. doi: 10.1002/pits.20497

Miciak, J., Fletcher, J. M., Stuebing, K. K., Vaughn, S., & Tolar, T. D. (2014). Patterns of cognitive strengths and weaknesses: Identification rates, agreement, and validity for learning disabilities identification. *School Psychology Quarterly, 29,* 21-37.

Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology, 13,* 248-277. doi: 10.1016/0010-0285(81)90010-4

Popper, K. R. (1962). *Conjectures and refutations: The growth of scientific knowledge.* London: Basic Books.

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73-165). New York: Guilford Press.

Streiner, D. L. (2003). Diagnosing tests: Using and misusing diagnostic and screening tests. *Journal of Personality Assessment, 81,* 209-219. doi: 10.1207/S15327752JPA8103_03

Taylor, W. P., Miciak, J., Fletcher, J. M., & Francis, D. J. (2017). Cognitive discrepancy models for specific learning disabilities identification: Simulations of psychometric limitations. *Psychological Assessment, 29,* 446-457. doi: 10.1037/pas0000356

Toffalini, E., Giofre, D., & Cornoldi, C. (2017). Strengths and weaknesses in the intellectual profile of different subtypes of specific learning disorder: A study on 1,049 diagnosed children. *Clinical Psychological Science, 5,* 402-409. doi: 10.1177/2167702616672038

Ventura County SELPA. (2017). *The Ventura County SELPA pattern of strengths and weaknesses model for specific learning disability eligibility procedural manual*. Retrieved from http://www. venturacountyselpa.com.

Watkins, M. W. (2002). *Diagnostic Utility Statistics* [Computer software]. Phoenix, AZ: Ed & Psych Associates.

Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology, 39,* 204-221. doi: 0.1093/jpepsy/jst062

Youngstrom, E. A., Choukas-Bradley, S., Calhoun, C. D., & Jensen-Doss, A. (2015). Clinical guide to the evidence-based assessment approach to diagnosis and treatment. *Cognitive and Behavioral Practice, 22,* 20-35. doi: 10.1016/j.cbpra.2013.12.005

Youngstrom, E. A., & Van Meter, A. (2016). Empirically supported assessment of children and adolescents. *Clinical Psychology: Science and Practice, 23,* 327-347. doi: 10.1111/cpsp.12172

Table 1

Demographic Information for the KABC-II Standardization Sample Administered the KTEA-II Ages 7-18 (N = 2,025)

Variable	*n*	Percent of Sample	Percent of U.S. Population[a]
Sex			
Female	1,019	50.3	50.9
Male	1,006	49.7	49.1
Ethnic Group			
White	1,257	63.0	61.7
Hispanic	352	17.4	18.7
African American	286	14.1	15.3
Other	112	5.5	5.1
Census Region			
South	695	34.3	35.3
North Central	526	26.0	26.9
West	527	26.0	23.8
Northeast	277	13.7	19.2
Mother's Education			
11th Grade or Less	301	14.9	14.3
High School Graduate	657	32.4	31.9
1-3 Years College	603	29.8	30.3
4 Year Degree or Higher	464	22.9	23.6
Exceptionality Status			
Diagnosed or Classified	429	21.2	22.4
No Status	1,596	78.8	77.6

Note. Demographic labels correspond to those reported in the KABC-II technical manual (Kaufman & Kaufman, 2004b).
[a]2001 *Current Population Survey* values.

Table 2

Means and Standard Deviations for KABC-II/KTEA-II Normative Sample Participants Ages 7-18 (N = 2,025) According to Conditional Status of Academic Functioning

		No Weakness (n = 1,503)		Weakness (n = 518)	
Model	Score	*M*	*SD*	*M*	*SD*
Reading < 90	Crystallized Ability	104.75	12.73	86.54	11.91
	Fluid Reasoning	103.96	13.61	88.95	13.02
	Visual Process	103.17	14.07	91.27	13.85
	Short-Term Memory	103.23	13.76	91.15	14.79
	Long-Term Memory	104.23	13.95	89.20	12.36
	Reading Composite	106.12	10.97	80.56	7.07
		No Weakness (n = 1,687)		Weakness (n = 334)	
Model	Score	*M*	*SD*	*M*	*SD*
Reading < 85	Crystallized Ability	103.24	13.23	84.13	11.96
	Fluid Reasoning	102.86	13.75	86.23	13.05
	Visual Process	102.23	14.14	89.46	14.35
	Short-Term Memory	102.41	13.99	88.61	14.57
	Long-Term Memory	103.06	14.06	86.84	12.42
	Reading Composite	104.02	11.98	77.11	6.54
		No Weakness (n = 1,520)		Weakness (n = 505)	
Model	Score	*M*	*SD*	*M*	*SD*
Mathematics < 90	Crystallized Ability	104.02	13.33	88.15	12.56
	Fluid Reasoning	104.06	13.41	88.15	12.93
	Visual Process	103.57	13.70	89.76	13.68
	Short-Term Memory	103.19	13.80	90.86	14.61
	Long-Term Memory	103.75	14.18	90.24	12.89
	Mathematics Composite	106.08	10.82	81.32	7.16

Model	Score	No Weakness (n = 1,732)		Weakness (n = 293)	
		M	*SD*	*M*	*SD*
Mathematics < 85	Crystallized Ability	102.59	13.73	85.15	12.06
	Fluid Reasoning	102.54	13.88	85.62	12.83
	Visual Process	102.33	13.88	87.10	14.36
	Short-Term Memory	102.04	14.13	88.75	14.89
	Long-Term Memory	102.48	14.38	97.97	12.75
	Mathematics Composite	103.77	11.91	77.07	6.63

Model	Score	No Weakness (n = 1,509)		Weakness (n = 512)	
		M	*SD*	*M*	*SD*
Written Language < 90	Crystallized Ability	103.96	13.17	88.65	13.51
	Fluid Reasoning	103.61	13.63	89.80	13.92
	Visual Process	102.82	14.09	92.17	14.57
	Short-Term Memory	103.19	13.89	91.11	14.49
	Long-Term Memory	103.84	14.11	90.17	13.01
	Written Language Composite	106.34	10.94	80.91	7.08

Model	Score	No Weakness (n = 1,709)		Weakness (n = 312)	
		M	*SD*	*M*	*SD*
Written Language < 85	Crystallized Ability	102.70	13.64	85.75	12.93
	Fluid Reasoning	102.59	13.85	86.52	13.53
	Visual Process	101.98	14.33	89.92	14.13
	Short-Term Memory	102.16	14.14	89.01	14.63
	Long-Term Memory	102.77	14.24	87.29	12.52
	Written Language Composite	104.09	12.01	76.95	6.39

Model	Score	No Weakness (n = 1,550)		Weakness (n = 475)	
		M	*SD*	*M*	*SD*
Oral Language < 90	Crystallized Ability	104.41	12.82	85.89	11.81
	Fluid Reasoning	103.43	13.53	89.19	14.22
	Visual Process	102.79	13.93	91.44	14.82
	Short-Term Memory	102.94	13.63	90.89	15.05

	Long-Term Memory	103.57	13.98	89.98	13.64
	Oral Language Composite	106.53	11.13	81.29	7.19

		No Weakness (n = 1,752)		Weakness (n = 273)	
Model	Score	*M*	*SD*	*M*	*SD*
Oral Language < 85	Crystallized Ability	102.89	13.12	81.95	11.33
	Fluid Reasoning	102.42	13.82	85.16	13.37
	Visual Process	101.82	14.14	89.24	15.36
	Short-Term Memory	102.01	13.90	87.97	15.98
	Long-Term Memory	102.48	14.06	86.92	14.19
	Oral Language Composite	104.30	12.15	76.89	6.56

Note. M = mean; SD = standard deviation. Differences significant at $p < .01$ for all six score comparisons across academic weakness groups.

Author Bios

Ryan J. McGill, Ph.D., NCSP, BCBA-D, is assistant professor and director of the school psychology program at the William & Mary School of Education. His scholarly practice interests include psychological assessment and measurement, judgement and decision-making in applied psychology, and assessment and identification of specific learning disability.

Sarah J. Conoyer, Ph.D., NCSP, is assistant professor in the department of psychology at Southern Illinois University Edwardsville. Her research interests are in developing formative assessments to assist middle and high schools in identifying students that need extra supports, evaluating methods to monitor students at-risk for academic difficulties in content areas such as social studies and science, as well as examining how school psychologists and teachers use data to make instructional decisions.

Sarah Fefer, Ph.D., BCBA, is assistant professor in the department of student development at University of Massachusetts Amherst. Her research and clinical interests include disruptive behavior in children and adolescents, assessment and intervention related to academic and behavioral competence, and working with families and communities to support student success.