Please use the following citation when referencing this work:

Benson, N. F., Beaujean, A., McGill, R. J., & Dombrowski, S. C. (2018). Critique of the Core-Selective Evaluation Process. *The DiaLog, 47 (2),* 14-18.

Critique of the Core-Selective Evaluation Process

Nicholas F. Benson

Alexander Beaujean

Baylor University

Ryan J. McGill

William & Mary

Stefan C. Dombrowski

Rider University

Author Note

Correspondence concerning this article should be addressed to Nicholas F. Benson, Department of Educational Psychology, Baylor University, One Bear Place #97301, Waco, TX 76706. E-Mail: Nicholas_Benson@baylor.edu

Abstract

This article briefly describes and offers a critique of the Core-Selective Evaluation Process (C-SEP). The C-SEP is a pattern of strengths and weaknesses (PSW) method for identifying specific learning disabilities (SLD). The authors of the C-SEP propose that the model has several distinguishing characteristics. Defining characteristics appear to be an emphasis on the assessment and consideration of expressive and receptive language abilities, the requirement for using normative data provided from test publishers, and the use of professional judgment. No evidence exists to support the diagnostic utility of the C-SEP or the claims made by its test authors. Instead, careful examination of the C-SEP authors' claims show that the model presently falls short of being a promising model for identifying SLD. Thus, we recommend that diagnosticians, school psychologists, and other assessment specialists eschew the C-SEP method until sufficient evidence emerges to support its use.

Keywords: assessment, specific learning disabilities, diagnostic practices, psychological testing

A Critique of the Core-Selective Evaluation Process

The Core-Selective Evaluation Process (C-SEP) is a new process for identifying a specific learning disability (SLD). Since it is a new process, little has been written about it. What has been written has been authored by the individuals who developed the process (Schrank, Stephens-Pisecco, & Schultz, 2017; Schultz & Stephens, 2015, 2017) and has focused only on potential benefits rather than providing a critical review.

Identification of SLD is a high-stakes decision. Thus, it is important to review any proposed identification process critically before adopting it (VanDerHeyden, 2018). The purpose of this article is to examine the C-SEP critically. First, we briefly outline PSW methods to clarify the conceptual framework of the model. Second, we describe the development of the C-SEP model and the unique features that distinguish it from other PSW models. Third, we critique these features as well as claims about the model proffered by its authors. Finally, we provide recommendations to practitioners based on our review of relevant conceptual issues and empirical research.

C-SEP

The C-SEP is based on the patterns of strengths and weaknesses (PSW) philosophy of operationalizing and identifying SLD. Much has already been written about PSW methods (e.g., McGill, Styck, Palomares, & Hass, 2016; Schultz, Simpson, & Lynch, 2012). We do not rehash the information here except to note that the hallmark of all PSW methods is that SLD is operationalized as pattern of uneven development across intelligence and academic attributes that are manifested in IQ and norm-referenced academic achievement tests.[1] Individuals with a

[1] We use the term *intelligence* throughout this article to refer to the class of attributes within the general domain of cognitive ability, not any particular attribute within that domain. We use the term *IQ test* to refer to any instrument designed to assess intelligence attributes.

SLD are those with strengths in many intelligence attributes, but weaknesses in others that are thought to be directly related to difficulties in academic achievement.

Schultz and Stephens (2017) developed the C-SEP model "after a critical analysis of all published. . .PSW approaches and incorporated the strengths of all SLD identification models and addressed the limitations of current approaches" (p. 3). They wanted a process that could "answer SLD referral questions in a comprehensive, time-efficient, precise, and legally defensible manner"(Schultz & Stephens, 2015, p. 5). The term *core-selective* refers primarily to the ways in which norm-referenced tests are used in SLD identification process. Specifically, norm-referenced scores representing core areas are included in all SLD evaluations. Schultz and Stephens (2017) provided a list of features that distinguish C-SEP from—and address limitations of—other PSW methods. Some of the more pertinent ones are: (a) a comprehensive assessment of all of the components of the federal definition of SLD, including expressive and receptive language abilities; (b) most examiner time is dedicated to interpreting and integrating data instead of test administration because tests are administered only to provide new information; (c) eligibility is primarily based on clinical judgement from using exploratory pattern seeking techniques—statistical analyses are only used to supplement these procedures; (d) academic underachievement is determined using multiple sources, never solely based on norm-referenced scores; and (e) normative data provided from test publishers must be used when conducting comparisons among test scores.

C-SEP Process

C-SEP has four major steps: Review, Plan, Assess, and Decide. The Review and Plan steps are similar to most other SLD assessment methods. The Review step involves reviewing the extant information available about a student, including exclusionary factors and response-to-

intervention (RtI) data, to establish the student fails to meet state standards in one of the SLD areas.

The Plan step involves creating a plan for what instruments to administer. Most of these involve intelligence, academic achievement, and language tests that provide norm-referenced scores in the "core" areas, although behavior observations and "informal" tests (e.g., curriculum-based, language demands assessment) can also be included if such information is not already part of a student's record. The Intelligence Core is a set of scores that provide information on "the broadest range of abilities recommended by publisher" (Schultz & Stephens, 2017, p. 10). The Language Core is a set of scores "that broadly measures the CHC factors (G*c*, G*a*, G*wm*, G*ltr* [*sic*], G*s*) related to oral expression and listening comprehension" (p. 11). The Achievement Core are a set of scores that assess the six core academic areas: Basic Reading, Reading Comprehension, Reading Fluency, Math Calculation, Math Problem Solving, and Written Expression. If a school district does not require information in all six achievement areas, then clinicians may choose to administer only tests that assess the areas of concern.

The Assess step involves collecting data outlines in the Plan step. The Decide step involves an integration of data and application of PSW policy. The data integration process is qualitative and exploratory because it is "pattern seeking," so requires "a high degree of expertise and professional judgment" (Schrank et al., 2017, p. 37). It involves (a) examining the trustworthiness (validity) of the data; (b) triangulating the data; (c) examining exclusionary factors; and (d) analyzing the data for PSW, which is done by the sorting data (usually based on CHC broad abilities) and visually displaying the results.

Some Critiques of the C-SEP

Required language assessment. The requirement that expressive and receptive language abilities be assessed in all SLD evaluations distinguishes the C-SEP from other SLD identification methods. This requirement comes from federal language used to define SLD (34 CFR § 300.8.c.10.1), although the meaning of the term "basic psychological processes involved in understanding or in using language, spoken or written" is somewhat unclear. For example, when asked for clarification of the SLD definition the U.S. Department of Education (2006) did not add any details to the definition, opting instead to note that the definition is consistent with the procedures for identifying students suspected of having a SLD (e.g., 34 CFR §§ 300.307–300.309). Nowhere in the SLD identification procedures is a language assessment specified as being required for all SLD evaluations.

The inconsistency between SLD definition and evaluation procedures has existed since SLD was first codified in federal guidelines (Hallahan & Mercer, 2002). Thus, the federal definition of SLD likely just stems from history rather than based on scientific literature about the need to assess language for every manifestation of SLD. Nonetheless, this assumption that language is important to assess for all SLD evaluations can be experimentally examined by examining the incremental contributions of scores representing the core language attributes to diagnostic accuracy and instructional effectiveness (Hunsley & Meyer, 2003).

Criteria for PSW. Since C-SEP is a PSW method, a question that arises is: how are PSW determined? Schultz and Stephens (2017) wrote that patterns are "evident by significant variance. . . variation must be important and meaningful (practical) when scores differ by ~1 SD [standard deviation] when considering confidence intervals of norm referenced tests" (p. 4). This makes it appear that PSW exist when the lower bound of a confidence interval (CI) for one test score is "~1 SD" away from the upper bound of a CI for another test score (e.g., 77–83 and 98–

104). Yet, none of the C-SEP case studies provided (e.g., Schrank et al., 2017; Schultz & Stephens 2017) used CIs in the Decision step; instead they use differences in obtained scores (e.g., 80 and 95). Whether obtained scores or confidence intervals are used, however, it is clear that PSW are established in C-SEP using simple difference methods of comparing norm-referenced test scores.

Problems with using simple difference methods in SLD evaluations are well known (Reynolds, 1984-1985; Crawford & Garthwaite, 2002). One of the most problematic concerns is that simple score differences do not incorporate any information about the base rate of score differences. It is not unusual, by any definition of that term, for individuals without any known cognitive impairments to have score differences greater than 1 SD (e.g., Binder, Iverson, & Brooks, 2009). Consequently, examining simple differences is likely to indicate there are PSW when none exists.

Another problem with using simple differences is the assumption that because scores are put on the same scale (e.g., IQ scale with a mean of 100 and SD of 15) they are directly comparable. This assumption is generally false. It is true that if two standard scores on different instruments are both, for example, 85, it means that the raw scores are both 1 SD below the average score of the instruments' norming samples. It is not true that they both represent an equal amount of the attributes being assessed. Thus, an 85 IQ on instrument one could represent skills that are developmentally typical, while the 85 IQ on instrument two represent skills that are developmentally lagged. This issue is somewhat ameliorated when all the instruments used have the same normative group. Currently, the only instruments that have scores in all the C-SEP core areas that uses the same normative group is the Woodcock-Johnson IV (WJ IV; Schrank, McGrew, Mather, & Woodcock, 2014). So, unless schools purchase the WJ IV battery there is

no reason to believe that score from one instrument are directly comparable to scores from another instrument across all core areas.

Pattern seeking analysis. Unlike other PSW methods, the data analyses required by C-SEP are notably abductive. By its nature, pattern seeking analyses are an *exploratory* endeavor; used to develop hypotheses or explain phenomena, not (dis)confirm them. Thus, the promotion of pattern-seeking techniques in the C-SEP method raises the question regarding the purpose of an SLD evaluation: Should the intended purpose of an SLD evaluation be exploratory? Undoubtedly, exploratory analyses are important (Behrens & Yu, 2003), but they are incomplete. Adding professional judgment to exploratory analysis is not a panacea either since reliance on professional judgment tends to increase the problems associated with cognitive biases (Grove & Meehl, 1996; Wilcox & Schroeder, 2015). It is difficult to assign appropriate weights to these sources of information and filter out information that is erroneous or irrelevant, and the belief that clinical intuition can be used to transform fallible sources of information into clinically useful information is a misguided (Lilienfeld, Wood, & Garb, 2006).

Granted, SLD decisions in school settings are not made unilaterally; they must be made by a multidisciplinary team. However, team approaches to decision making are subject to numerous imperfections (e.g., Bohns, Mahdi, & Xu, 2016; Germar, Schlemmer, Krug, Voss, & Mojzisch, 2014; Milgram, 1963). Moreover, research suggests that multidisciplinary teams often ignore relevant sources of information. For example, multidisciplinary teams often override evidence that does not confirm the presence of SLD with clinical judgments supporting need for services (Bocian, Beebe, MacMillan, & Gresham, 1999) while they rarely override child performance data when environmental, cultural, or economic disadvantage are likely to be the primary cause of SLD (Fletcher & Naverrete, 2003).

Critique of Claims Made by the Authors of the C-SEP

The authors have made a variety of claims about the C-SEP, seemingly to justify its development and promote its use. Some claims seem reasonable, while others appear problematic. We have provided a selective critique of what we contend are the most problematic claims. Space considerations preclude an exhaustive review.

Claim 1: Recent technological advances have improved the measurement of cognitive abilities

Schultz and Stephens (2017) state that "…tests have been updated and improved by being more cognitively complex allowing for more than one ability to be assessed" (p. 17). This statement conflates cognitive complexity and multidimensionality. The former is often inferred from a test's loading on a general factor of intelligence (i.e., *g*), and refers to the extent to which a test measures the ability to handle information that is complex, which often involves thinking abstractly and solving novel problems (Gottfredson, 1997). The latter refers to the number of attributes a test measures. Most cognitively complex tests are unidimensional (i.e., they are primarily measures of *g*). In contrast, multidimensional tests are not necessarily cognitively complex. In fact, the strategy of designing tests to measure *g* as well as more specific cognitive abilities tends to produce tests that are good measures of *g* but poor measures of the more specific abilities because most of the common variance among the specific abilities is also in common with *g* (Beaujean & Benson, 2018).

It could be argued that the alignment of cognitive tests with the Cattell-Horn-Carroll taxonomy (CHC; Schneider & McGrew, 2018) is a technological advance. However, multiple structural validity studies have shown that popular tests of intelligence are good measures of *g* but inadequate measures of many broad and narrow CHC abilities (Dombrowski, Canivez,

Watkins, & Beaujean, 2015; Dombrowski, McGill, & Canivez, 2017). Thus, recent research does not support the claim that technological advances have improved the measurement of CHC abilities.

Claim 2: C-SEP allows practitioners to avoid dichotomous identification decisions

The basis for this conclusion is that PSW models utilize a normative-developmental perspective that "…allows diagnosticians to go beyond classifying and sorting students by offering explanations more precisely and comprehensively" (Schultz & Stephens, 2015, p. 9). This claim is problematic for two reasons. First, it incorrectly assumes that practitioners using other models do not utilize a normative-developmental perspective. Second, use of a normative-developmental perspective in no way obviates the need to render dichotomous identification decisions—a student must either be deemed eligible or deemed ineligible for special education and related services.

Claim 3: The structure of cognitive tests allows inferences about neurological functioning

Schultz and Stephens (2017) claim that cognitive test scores provide insights about neurological functioning, and that no specialized training in neuropsychology is needed to make inferences about neurological functioning. To support this claim, the authors argue the following:

> Major cognitive batteries were developed and informed by the field of neuropsychology. By using the test publisher's organizational structure and applying it to the referral concerns, tests are already organized. In other words, the "neuropsychology has been done for you (Schultz & Stephens, 2017, p. 16).

Most contemporary cognitive, neuropsychological, and academic tests provide a bevy of measured scores to appeal to as many potential test users as possible. Many test scores are not grounded in the attribute's theory or developed from sound measurement or psychometric theory

(Beaujean & Benson, 2018). Thus, there is no reason to believe that the "neuropsychology has been done for you". Practitioners should practice within their boundaries of competence (American Psychological Association, 2017). As noted by Meyer and colleague (2001), "Tests do not think for themselves, nor do they directly communicate with patients. Like a stethoscope, a blood pressure gauge, or an MRI scan, a psychological test is a dumb tool…" (p. 153).

Claim 4: C-SEP ensures CHC abilities are measured reliably

Schultz and Stephens (2017) claim that the use of integrated data analysis ensures that abilities are measured reliably. Although it may be good practice, such integration in no way ensures reliability. Combining a highly reliable score with one or more scores with lower reliability will tend to produce a composite score that is less reliable than simply interpreting the highly reliable score.

Claim 5: C-SEP increases diagnostic precision

Schultz and Stephens (2017) note that "C-SEP procedures help examiners collect diagnostically precise data by going from a broad testing perspective (core) to an increasingly narrow scope (selective) of testing" (p. 17). While this approach may increase efficiency relative to simply administering all possible tests in a battery, there is no evidence to support the claim that this approach increases diagnostic precision. Additionally, the C-SEP model allows practitioners to interpret individual subtests as proxies for broad abilities. Reviews of the subtest interpretation literature have revealed that subtest scatter/profiles are not useful for discriminating diagnostic groups, including individuals with SLD (Watkins, 2000; Watkins, 2003). Moreover, a recent simulation study revealed that using multiple indicators of broad abilities does little to increase the diagnostic accuracy of Hale and Fiorello's (2004)

concordance/discordance model (Miciak et al., 2018)—it is unlikely that different outcomes would be found for the C-SEP model.

Claim 6: C-SEP is useful for differential diagnosis

Schultz and Stephens (2017) refer to a table on page 18 of the C-SEP manual and suggest that it can be used to differentiate students with SLD from those with intellectual disabilities (ID). In the table it is noted students with ID display little variation in their cognitive profiles, with all or nearly all scores ≤ 70, with the possible exception of one or more strengths in scores derived from tests with low *g* loadings. Bergeron and Floyd (2013) examined IQ test profiles in students with ID. Although the group profile for ID students was flat, very few students were found to have uniformly low profiles. In fact, most had one or more part scores (i.e., index, cluster, or factor score) above 70, and many even displayed at least one part score above 80. Based on the results, Bergeron and Floyd concluded that more than a third of students with ID would not meet criteria for ID if elevated part scores were used to disconfirm the diagnosis/eligibility determination.

Summary

The C-SEP is a PSW model for identifying SLD. It is largely a collection of recommended practices rather than a model per se. The most distinctive characteristics of the C-SEP model are arguably the emphasis on the assessment and consideration of expressive and receptive language abilities, the requirement for using normative data provided from the publishers, and the primacy of professional judgment. No evidence exists to support the diagnostic utility of the C-SEP.

Although the authors make numerous claims that may invoke enthusiasm, careful examination of the evidentiary basis of these claims show that the model presently falls short of

being a promising model for identifying SLD. The preponderance of available empirical evidence suggests that searching for PSW may not be worth the time and effort it requires. Accordingly, we recommend that diagnosticians, school psychologists, and other assessment specialists eschew the C-SEP model until such time as sufficient evidence emerges to support its use.

References

34 CFR §§ 300.307–300.309 (2004).

34 CFR § 300.8.c.10.1 (2004).

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct* (2002, Amended June 1, 2010 and January 1, 2017). Retrieved from http://www.apa.org/ethics/code/index.aspx

Beaujean, A. A., & Benson, N. F. (2018). Theoretically-consistent cognitive ability test development and score interpretation. Contemporary School Psychology, doi:10.1007/s40688-018-0182-1

Behrens, J. T., & Yu, C.-h. (2003). Exploratory data analysis. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology, vol 2: Research methods in psychology* (pp. 33--64). Hoboken, NJ: Wiley.

Bergeron, R., & Floyd, R. G. (2013). Individual part score profiles of children with intellectual disability: A descriptive analysis across three intelligence tests. *School Psychology Review, 42,* 22-38.

Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: "Abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*. doi:10.1093/arclin/acn001

Bocian, K. M., Beebe, M. E., MacMillan, D. L., & Gresham, F. M. (1999). Competing paradigms in learning disabilities classification by schools and the variations in the meaning of discrepant achievement. *Learning Disabilities Research & Practice, 14*, 1-14.

Bohns, V. K., Roghanizad, M. M., & Xu, A. Z. (2014). Underestimating our influence over others' unethical behavior and decisions. *Personality and Social Psychology Bulletin, 40,* 348-362

Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia, 40,* 1196–1208. doi:10.1016/S0028-3932(01)00224-X

Dombrowski, S. C., Canivez, G. L., Watkins, M. W., & Beaujean, A. A. (2015). Exploratory bifactor analysis of the Wechsler Intelligence Scale for Children-Fifth Edition with the 16 primary and secondary subtests. *Intelligence, 53,* 194–201.

Dombrowski, S. C., McGill, R. J. & Canivez, G. L. (2017). Exploratory and hierarchical factor analysis of the WJ IV Full Test battery. *School Psychology Quarterly*. Advance online publication. doi: 0.1037/spq0000221

Fletcher, T., & Navarrete, L. (2003). Learning disabilities or difference: A critical look at the issues associated with the misidentification and placement of Hispanic students in special education programs. *Rural Special Education Quarterly, 22,* 37-4.

Germar, M., Schlemmer, A., Krug, K., Voss, A., & Mojzisch, A. (2014). Social influence and perceptual decision making: A diffusion model analysis. *Personality and Social Psychology Bulletin*, *40,* 217-231.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24,* 79-132.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, And Law*, *2*, 293-323.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook.* New York: Guilford Press.

Hallahan, D. P., & Mercer, C. D. (2002). Learning disabilities: Historical perspectives. In R. Bradley, L. Danielson & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 1--68). Mahwah, NJ: Lawrence Erlbaum.

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: conceptual, methodological, and statistical issues. *Psychological Assessment, 15,* 446–455.

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2008). What's wrong with this picture? In S. O. Lilienfeld, J. Ruscio, S. J. Lynn, S. O. Lilienfeld, J. Ruscio, S. J. Lynn (Eds.), *Navigating the mindfield: A user's guide to distinguishing science from pseudoscience in mental health* (pp. 133-144). Amherst, NY, US: Prometheus Books.

McGill, R. J., Styck, K. M., 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.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., & ... Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56,* 128-165.

Miciak, J., Taylor, W. P., Stuebing, K. K., & Fletcher, J. M. (2018). Simulation of LD identification accuracy using a pattern of processing strengths and weaknesses method with multiple measures. *Journal of Psychoeducational Assessment, 36,* 21-33.

Milgram, S. (1963). Behavioral Study of obedience. *The Journal of Abnormal and Social Psychology, 67,* 371-378.

Reynolds, C. R. (1984-1985). Critical measurement issues in learning disabilities. *The Journal of Special Education, 18,* 451-476.

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. 73–162). New York, NY: Guilford.

Schrank, F.A., McGrew, K.S., Mather, N., & Woodcock, R.W. (2014). *Woodcock-Johnson IV.* Rolling Meadows, IL: Riverside Publishing.

Schrank, F. A., Stephens-Pisecco, T. L., & Schultz, E. K. (2017). *The WJ IV Core-Selective Evaluation Process Applied to Identification of a Specific Learning Disability* (Woodcock-Johnson IV Assessment Service Bulletin No. 8). Itasca, IL: Houghton Mifflin Harcourt.

Schultz, E. K., Simpson, C. G., & Lynch, S. (2012). Specific learning disability identification: What constitutes a pattern of strengths and weaknesses. *Learning Disabilities: A Multidisciplinary Journal, 18,* 87-97.

Schultz, E. K., & Stephens, T. L. (2015). Core-Selective Evaluation Process: An efficient and comprehensive approach to identify students with SLD using the WJ IV. *The Dialog, 44* (2), 5-12.

Schultz, E. K., & Stephens, T. L. (2017). *Core-Selective Evaluation Process (C-SEP): Overview and procedures, Texas edition.* Authors.

U.S. Department of Education. (2006). Assistance to states for the education of children with disabilities and preschool grants for children with disabilities; final rule. *Federal Register* 71(156):46540–46845.

VanDerHeyden, A. M. (2018). Why do school psychologists cling to ineffective practices? Let's do what works. *School Psychology Forum, Research in Practice, 12*, 44-52.

Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly*, *15*, 465-479.

Watkins, M. W. (2003). IQ Subtest Analysis: Clinical Acumen or Clinical Illusion? *The Scientific Review of Mental Health Practice: Objective Investigations of Controversial and Unorthodox Claims in Clinical Psychology, Psychiatry, and Social Work, 2*(2), 118-141.

Wilcox, G., & Schroeder, M. (2015). What comes before report writing? Attending to clinical reasoning and thinking errors in school psychology. *Journal of Psychoeducational Assessment, 33*, 652-661. doi:10.1177/0734282914562212