

# The Influence of a Behavior Rating Scale on School Psychologists' Diagnostic Decision-Making

Lesley A. Higgins

Todd County Schools

205 Airport Road, Elkton, Kentucky 42220, United States

Tel: 1-270-265-2436 E-mail: [lesley.higgins@todd.kyschools.us](mailto:lesley.higgins@todd.kyschools.us)

Carl L. Myers (Corresponding author)

Department of Psychology, Western Kentucky University

1906 College Heights Blvd. #21030, Bowling Green, Kentucky 42101, United States

Tel: 1-270-745-4410 E-mail: [carl.myers@wku.edu](mailto:carl.myers@wku.edu)

Received: March 9, 2014 Accepted: April 3, 2014 Published: April 4, 2014

doi:10.5296/jet.v1i2.5262 URL: <http://dx.doi.org/10.5296/jet.v1i2.5262>

## **Abstract**

Psychologists are increasing their use of behavior rating scales as part of their diagnostic assessments. While behavior rating scales have many advantages, there are numerous cautions to using them, particularly for diagnostic purposes. It is unknown, however, how much influence a behavior rating scale has on a professional's diagnostic impression. Using an experimental design, a national sample of 330 school psychology practitioners from the United States provided diagnostic impressions on one of four scenarios. Two scenarios had evaluation data that supported a diagnosis of Attention-Deficit/Hyperactivity Disorder (ADHD) with one scenario including behavior rating scale standard scores in the average range and one with scores in the clinically significant range. The other two scenarios had evaluation data that did not support a diagnosis of ADHD but also had one with standard scores in the average range and one with scores in the clinically significant range. The findings revealed that behavior rating scale scores do influence school psychologists' diagnostic decisions. However, consistent with recommended best practices, other sources of evaluation data have a stronger influence on school psychologists' diagnostic decisions.

**Keywords:** behavior rating scales, diagnostic decision-making, school psychologists

## 1. Introduction

The assessment and diagnosis of psychological disorders is a common practice of many clinical psychologists and psychiatrists. Likewise, the assessment and diagnosis of developmental and educational disabilities is a common role for school psychologists. Given that the role of assessment and diagnosis is prevalent in applied psychology, it is surprising how little research has been conducted on the influences impacting psychologists' diagnostic decisions. A substantial number of studies have examined the related concept of clinical judgment, examining such factors as years of experience and potential biases of the clinician (e.g., López, 1989; Spengler et al., 2009). However, diagnostic decision-making focuses on the process of diagnosis, including differential diagnosis, and the variables that influence a diagnostic conclusion.

While the topic of diagnostic decision-making is prominent in the medical field (e.g., Sox, Higgins, & Owens, 2013), it is rare to find literature examining how mental health professionals reach diagnostic decisions. Examples of related research in the mental health field include the examination of variables influencing diagnostic decisions in psychiatric emergency rooms (Muroff, Jackson, Mowbray, & Himle, 2007), and whether the method of payment influenced diagnostic decisions by psychologists in the clinical setting (Lowe, Pomerantz, & Pettibone, 2007). Recently, Groenier, Beerthuis, Pieters, Witteman, and Swinkels (2011) evaluated psychologists' interviews during the diagnostic process. While they noted that the interviews were unstructured and the diagnostic processes varied across psychologists, in general, appropriate diagnoses were determined.

Studies evaluating influences on school psychologists' diagnostic decision-making processes are even rarer. An early exception focused on the diagnosis of a learning disability (Kavale & Andreassen, 1984). With a sample of only 22 school psychologists, the authors noted inconsistencies in which pieces of information were weighed most in reaching a diagnostic decision. Gnys, Willis, and Faust (1995) also examined the diagnosis of learning disability by assessing school psychologists' diagnostic decision-making related to inter-subtest scatter on an IQ test. Different scenarios with varying pieces of information were provided and the school psychologists were asked to rate the probability of a learning disability. Their results indicated that the amount of inter-subtest scatter did influence diagnostic impressions. A dissertation by Kretchman (2007) used a survey to examine the responses from a small sample ( $n = 34$ ) of school psychologists regarding the use of different pieces of assessment data to reach diagnostic conclusions and concluded the participants made "sub-optimal" use of the available hypothetical data.

Most recently, Barnard-Brak, Stevens, Robinson, and Holt (2013) examined how hypothetical cases with indicators of different *DSM-IV* disorders and different levels of academic performance (i.e., good or poor) affected the diagnostic decision-making of a sample of 179 school psychologists from Texas. Vignettes were used to vary pieces of information presented to the participants. In general, the school psychologists were more likely to make consistent and correct diagnoses when the academic performance of the students was poor than when it was good. The authors believed the potentially stigmatizing effects of a diagnosis might be

impacting school psychologists' decisions when a student's academic performance was good.

### *1.1 Purpose of Current Research*

The purpose of the current study was to further evaluate the importance school psychologists place on pieces of assessment information. In particular, the influence of scores from a behavior rating scale was experimentally manipulated. Using a similar methodology as previous studies, vignettes were developed for this study that varied scores from a behavior rating scale, while keeping other assessment data consistent, to see how much influence scores on a behavior rating scale have on school psychologists' diagnostic impressions. Table 1 provides a description of the four vignettes and the congruency of data sources.

Table 1. Characteristics of vignettes and congruency of assessment data

Various sources of data	Behavior Rating Scale	Congruency of data
1. Supports diagnosis	Supports diagnosis	Congruent
2. Supports diagnosis	Does not support diagnosis	Incongruent
3. Does not support diagnosis	Supports diagnosis	Incongruent
4. Does not support diagnosis	Does not support diagnosis	Congruent

### *1.2 Description of Behavior Rating Scales*

Behavior rating scales are standardized instruments that contain a list of behavioral descriptors and a rater (e.g., parent or teacher) indicates the extent to which the behavior is present. Some versions of behavior rating scales also require self-ratings by the person being evaluated. When developing the instruments, clusters of individual items are determined through factor analyses or principal component analyses and are named to reflect a relevant behavioral or psychological construct. A wide variety of constructs, such as depression, withdrawal, anxiety, hyperactivity, and attention problems, are measured on behavior rating scales. Norms are developed to provide standard scores on each construct. A mental health professional is able to use the standard scores obtained from a behavior rating scale to compare behaviors of a referred child to others of the same age and even same gender. Such information can be used to assist practitioners with diagnostic decision-making.

Behavior rating scales have been developed for all ages of students and even for children as young as 18 months of age. Since the mid-1980s, the use of behavior rating scales by school psychologists has grown tremendously in the assessment of social and emotional behaviors in children (Merrell, 2008; Shapiro & Heick, 2004). Shapiro and Heick attribute the rise in the use of behavior rating scales as due to the expanding research base and the increased technical adequacy of the instruments. In addition, the increased use of behavior rating scales can likely be attributed to their many advantages. The instruments are easy to administer, relatively objective, time efficient, provide data on behaviors not easily observed by an external observer, and provide useful data for screenings and evaluations (Angello et al.,

2003; Chafouleas, Riley-Tillman, & Sugai, 2007; Elliott, Busse, & Gresham, 1993).

Although behavior rating scales provide quantitative data on a broad range of children's problems and competencies, they also have limitations. The rating scales do not identify the etiology of an individual's problems and most rating scales only assess current functioning over a two to six month time frame (McConaughy & Ritter, 2008). Behavior rating scales do not provide complete information about the individual's personal or environmental factors, information relevant to the function of a behavior problem, or an explicit description of the behaviors of concern (Angello et al., 2003; McConaughy & Ritter, 2008). Furthermore, behavior rating scales really only measure perceptions of specified behaviors rather than provide any direct measurement of behaviors. Behavior ratings can be impaired or influenced by the rater's memory, values, attitudes, and motivations, as well as situational factors (McConaughy & Ritter, 2008).

School psychologists commonly use behavior rating scales as part of the screening and identification process for children referred for special education services (Chafouleas et al., 2007; Elliott et al., 1993; Merrell, 2008; Shapiro & Heick, 2004). The use of the scales for screening purposes is widely accepted (Carter, Briggs-Gowan, & Davis, 2004). However, using behavior rating scales for diagnostic purposes is more controversial with numerous authors stating the scales are not sufficient for determining a diagnosis (Carter et al., 2004; Chafouleas et al., 2007; Reid & Maag, 1994; Sattler, 2002). Best practices in diagnostic assessment would dictate that behavior rating scales should only be used in conjunction with other methods of assessment, such as interviews with the parents and teachers, direct observations, review of school records, and achievement assessments (Angello et al., 2003).

Although a multi-method assessment approach is recommended for diagnostic assessments, it is unknown how results from behavior rating scales inform and influence school psychologists' diagnostic decision-making processes. Specifically, it is unknown if any school psychologists use the results of a behavior rating scale to directly diagnose specific disorders (i.e., a school psychologist concludes a child has a specific diagnosis simply because of a high score on a scale measuring that construct). That direct connection would be difficult to determine accurately using school psychologists' self-ratings, given the general consensus about the inappropriateness of such a practice. A broader question is how much influence does a behavior rating scale have on a school psychologist's diagnostic decision-making process?

Attention-Deficit/Hyperactivity Disorder (ADHD) was the disorder picked for this study because it is one of the most common childhood disorders (American Psychiatric Association, 2013). School psychologists were asked to rate the likelihood of ADHD after reading one of four vignettes that either contained information consistent or inconsistent with an ADHD diagnosis, or a vignette containing a mixture of supporting and non-supporting evaluation data. Behavior rating scale scores systematically varied in those scenarios (i.e., average range or clinically significant). Logically, results should show that when all information consistently supports a diagnosis of ADHD, school psychologists are more likely to rate the student as having ADHD. Conversely, results should show that when all information does not

support a diagnosis of ADHD, school psychologists are unlikely to rate the student as having ADHD. The primary focus of this study, however, was how scores on a behavior rating scale influence a diagnostic impression. Thus, when a variety of assessment information is held constant, and only the behavior rating scale scores change, how does that affect a diagnostic impression? Of particular interest to this study is what happens to school psychologists' diagnostic impressions when the scores on a behavior rating scale suggest a strong likelihood of ADHD but other assessment information does not support such a diagnosis? Likewise, what happens to diagnostic impressions when most information indicates a strong likelihood of ADHD, but the scores on a behavior rating scale do not support such a diagnosis?

## 2. Method

### 2.1 Participants

The national sample of potential participants included school psychologists that were members of the National Association of School Psychologists (NASP) in the United States. NASP provided a randomly chosen sample of 800 school psychologists that had identified themselves as practitioners. The sample of 800 practitioners was then randomly divided into four groups of 200 school psychologists. Usable responses were received from 330 participants from 49 states for a return rate of 41.3%.

Demographic information regarding participants' years of experience, gender, and highest degree are presented in Table 2. Respondents in all four groups were fairly similar across all the assessed demographic variables of gender, highest degree, and years of experience. A chi-square distribution test was used to examine if the four groups differed on the variables of gender and highest degree, but there were no significant differences. For gender,  $\chi^2(3, 330) = 2.81, p > .05$ , and for the highest degree,  $\chi^2(3, 330) = 6.02, p > .05$ . A one-way ANOVA was used to look at years of experience but again there was no significant difference in years of experience among the four groups,  $F(3, 316) = .018, p = .997$ .

Table 2. Participant demographics

	Scenario 1 (n = 83)	Scenario 2 (n = 71)	Scenario 3 (n = 87)	Scenario 4 (n = 89)	Total (n = 330)
Gender					
Males	24.1%	15.5%	23.0%	16.9%	20.0%
Females	75.9%	84.5%	77.0%	83.1%	80.0%
Degree					
Masters	24.1%	16.9%	18.4%	27.0%	21.8%
Specialist	48.2%	57.7%	47.1%	50.6%	50.6%
Doctorate	27.7%	25.4%	34.5%	22.5%	27.6%
Mean Years Experience	14.8 (SD = 10.7)	14.7 (SD = 11.1)	15.1 (SD = 10.6)	14.9 (SD = 11.2)	14.9 (SD = 10.9)

The respondents were predominately female (80.0%), and that percentage is similar to demographic statistics from the 2004-05 membership year for NASP (Curtis et al., 2008), which indicated that 77% of school psychology practitioners are female. Similarly, 27.6% of this sample had doctoral degrees and Curtis et al. reported 24.4% of practitioners had doctoral degrees. This sample did have a higher percentage of specialist degrees than reported by Curtis et al. (50.6% vs. 39.9%). However, the trend in the field has been increasing numbers of female, specialist-level practitioners (Curtis, Grier, & Hunley, 2004). The mean years of experience for this sample (14.9) was also very comparable to the 14.0 mean presented by Curtis et al. (2008). Therefore, because this sample's demographics are very similar to NASP's membership statistics, it is probable that a representative sample of NASP members was obtained and the results are more likely to generalize to the field as a whole.

## 2.2 Instrument

Four scenarios and a response form were developed for this study. In addition to basic demographic information, the participants were asked to indicate how often (i.e., never, rarely, sometimes, fairly frequently, often) they used the *Behavior Assessment System for Children-2* (*BASC-2*, Reynolds & Kamphaus, 2004) as a measure of their familiarity with the instrument discussed in the scenarios. The *BASC-2* was chosen because it is a commonly used behavior rating scale in schools, is a well-developed instrument, and is recommended as an appropriate instrument for assessing ADHD (Merrell, 2008; Tobin, Schneider, Reck, & Landau, 2008). The *BASC-2* scales of Hyperactivity and Attention Problems were selected because those two scales assess the core aspects of ADHD (Reynolds & Kamphaus, 2004). Participants were also asked to indicate how often (i.e., very rarely/never, 1-3 times a year, 4-6 times a year, 7-9 times a year, 10 or more times a year) they provide assessment information for ADHD as a measure of school psychologists' role in this specific activity.

Four scenarios were created to assess the influence of behavior rating scales on school psychologists' diagnostic decision making (see Appendix). The scenarios include a variety of pieces of assessment information, based on Tobin et al.'s (2008) description of ADHD assessment best practices. All scenarios included bulleted lists of information from teacher and parent interviews, as well as systematic classroom observation data and standard scores (i.e., T scores) on the Hyperactivity and Attention Problems scales from the *BASC-2*. *BASC-2* T scores of 57 (not supportive of an ADHD diagnosis) and 73 (supportive of an ADHD diagnosis) were chosen as the behavior rating scale scores to be included in the scenarios. T scores have a mean of 50 and a standard deviation of 10. Thus, a T score of 57 is slightly elevated but in the upper part of the average range. A T score of 73 is considered clinically significant. Each piece of information was listed in the same sequence on each scenario.

In Vignette 1, all of the assessment information, including the behavior rating scale scores, supported an ADHD diagnosis. Vignette 2 included parent and teacher interview data and systematic observation data that supported an ADHD diagnosis; however, the scores from the *BASC-2* did not support an ADHD diagnosis. In Vignette 3, the parent and teacher interview information and systematic observation information did not support a diagnosis of ADHD, but the *BASC-2* Hyperactivity and Attention Problems scales were significantly high. In

Vignette 4, none of the information provided supported an ADHD diagnosis. Participants were asked to rate the likelihood the student has ADHD on a six point Likert scale where 1 = “Not at all” and 6 = “Definitely.”

After the scenarios were created, they were field-tested by sending them to 40 school psychologists from the local region and three nearby states. The pilot study group was asked to provide feedback on the scenarios and survey questions. Based on their feedback, two changes were made. The wording of one survey item was changed to ask how often practitioners provided assessment information for an ADHD evaluation rather than how often they “diagnosed” ADHD because of numerous comments questioning the appropriateness of school psychologists diagnosing ADHD. In addition, many respondents commented on the student’s grades listed in the scenarios and raised questions about the implications of a significant “adverse effect” on the student’s educational performance. To be eligible for special education services in the United States, it must be documented that the disability has an adverse effect on academic performance. To avoid diagnostic impressions based on adverse effect, the student’s grades were eliminated from the scenarios.

### 2.3 Procedure

Addresses from 800 randomly selected members of NASP were obtained from the NASP headquarters. The focus of this study was how behavior rating scales might influence practitioners’ diagnostic impressions and, thus, it was requested of NASP to only provide addresses for school psychology practitioners, and exclude others such as trainers or students. After field-testing the instrument, the national sample of potential participants was mailed a cover letter explaining the purpose of the study and informed consent information, one scenario, and a pre-stamped postcard to indicate and mail back their responses. The returned postcards contained no identifying information, other than the scenario number. The anonymity was intended to encourage or allow candid responses. Unfortunately, the anonymity also meant no follow-up mailings could be conducted for non-respondents.

### 3. Results

The use of the *BASC-2* is apparently common, as most participants were familiar with the specific instrument picked for inclusion in the scenario. The vast majority of school psychologists (85.5%) indicated they used the *BASC-2* “fairly frequently,” “often,” or “sometimes.” Deleting those that indicated “sometimes” still left more than two-thirds of the sample (68.9%) indicating “fairly frequently” and “often.” Only 7.7% responded that they never used the *BASC-2*.

The results also indicated most school psychologists were involved in providing assessment information for an ADHD evaluation. Almost half of the sample (42.3%) provided ADHD assessment information at least ten times per school year. When categories of responses are combined, most of the participants (78.1%) indicated they provided ADHD assessment information at least four times a year. Another 13% of the respondents provided ADHD assessment information one to three times a year, while only 8.9% indicated they “very rarely/never” provide such information. Thus, this sample of school psychologists was

generally familiar with collecting assessment information for an ADHD diagnosis.

After reading a scenario, the school psychologists indicated the likelihood of the student in the scenario having ADHD on a six-point Likert scale. Mean ratings for each of the scenarios are presented in Table 3. The results were analyzed using a one-way ANOVA, resulting in significant differences across the four groups,  $F(3, 326) = 116.45$ ,  $p < .001$ ,  $\eta^2 = .53$ . Tukey post-hoc comparisons of the four groups indicated each group was significantly different from every other group ( $p < .001$  for all comparisons). The effect size ( $\eta^2 = .53$ ) was found to exceed Cohen's (1988) convention for a medium effect. One implication of the results is that the scenarios were valid for the purpose of this study. That is, the obtained mean score for Vignette 1 (where all assessment information supported ADHD) should be, and was, significantly higher than all other vignettes. Likewise, the mean score for Vignette 4 (where none of the assessment information supported ADHD) should be, and was, significantly lower than all other vignettes. The primary purpose of the study was to evaluate school psychologists' diagnostic impressions when the assessment data are incongruent. The results indicate that school psychologists do not use a behavior rating score as their primary influence to make a diagnostic decision. In fact, the opposite result occurred as the mean score for Vignette 2 (all information except the behavior rating scale score supported ADHD) was significantly higher than the mean score for Vignette 3 (only the behavior rating scale score supported ADHD). In other words, the various pieces of assessment data (i.e., interview information and observations) outweighed the behavior rating scale score when making a diagnostic decision.

Table 3. Mean ratings on vignettes supporting/not supporting diagnosis of ADHD

Assessment data	Behavior Rating Scale	Mean	<i>SD</i>
1. Support	Support	3.68	0.77
2. Support	No support	3.15	0.84
3. No support	Support	2.31	0.68
4. No support	No support	1.70	0.61

*Note.* The ratings were on a six point scale where 1 = Not at all, 2 = Slight possibility, 3 = Moderate possibility, 4 = Likely, 5 = Very likely, and 6 = Definitely.

#### 4. Discussion

The direct influence of behavior rating scale results on school psychologists' diagnostic impressions has not been previously studied. Therefore, this research makes a unique contribution to the education and training literature. Until now, it was unknown how a behavior rating scale might influence a school psychologist's diagnostic impression about the presence or absence of a disorder. While the reasoning behind diagnostic decision-making is still unknown, this study clearly indicates the scores on a behavior rating scale have some



influence on school psychologists' diagnostic impressions. When all other assessment information was held constant, the different behavior rating scale scores resulted in a statistically significant difference in ratings of the likelihood of ADHD. The medium effect size indicated the influence of behavior rating scale scores is fairly substantial. However, the results also indicate school psychologists appropriately consider multiple pieces of information when confronted with a diagnostic decision. In fact, other pieces of evaluation data suggesting ADHD seemed to outweigh behavior rating scale scores that did not support the diagnosis of ADHD. Thus, results from the current study imply that school psychologists are engaging in best practices when using behavior rating scale results as advocated by numerous authors (Angello et al., 2003; Carter et al., 2004; Chafouleas et al., 2007; Reid & Maag, 1994; Sattler, 2002).

As expected, when none of the information supported an ADHD diagnosis, school psychologists were unlikely to indicate the presence of ADHD and when all of the information supported an ADHD diagnosis, school psychologists were much more likely to indicate the presence of ADHD. Somewhat surprising, however, was that even when all information supported a diagnosis of ADHD, school psychologists' mean rating of the likelihood of ADHD was only 3.68 on a six-point scale. Based on the Likert scale descriptors, such a score is between "moderate possibility" and "likely." However, obtaining only a moderate rating for that scenario was probably very appropriate. Even though all the information in the scenario supported the diagnosis of ADHD, it still contained a limited amount of information. School psychologists in this sample appear to have been appropriately cautious in their diagnostic impressions.

A limitation of this study is that information collected was based on self-ratings. Self-ratings are a limitation because individuals may be hesitant to accurately report on their professional actions, especially if such actions (i.e., relying too heavily only on behavior rating scale results) are recognized as not being best practice. Another limitation of this study is that there was only a 41.3% response rate. While such a response rate is fairly good for a survey type of methodology, the generalizability of the results to all school psychologists, of course, must be made with caution. However, despite the relatively low response rate, the demographics of the participants seemed to reflect the population of NASP school psychology practitioners as a whole. The mean number of years experience reported by this sample of school psychologists matched the demographics of the field, but was not evaluated as an independent variable. Thus, it is unknown whether years of experience impacts diagnostic decision-making.

## **5. Conclusion**

Behavior rating scales are a prominent part of a school psychologist's assessment procedures (Shapiro & Heick, 2004). An implication of this study is the knowledge and assurance that school psychologists are using multiple pieces of assessment information to form diagnostic impressions about students. Future research might examine in more detail what specific pieces of assessment information are more likely to influence a diagnostic impression. As examples, it is possible school psychologists might put more weight on teacher report than

parent report, or they might put more weight on systematic observation data than teacher report. Years of experience is another variable that could be examined further, as novice professionals may be more or less likely to make a diagnosis than veteran school psychologists. This study had the school psychologists use a Likert scale to indicate their diagnostic impressions. Perhaps requiring a dichotomous diagnostic decision might yield different results. It would also be interesting to see how behavior rating scale scores influence the diagnostic decisions of other mental health professionals (e.g., clinical psychologists, school social workers).

The use of another behavior rating scale besides the *BASC-2* might also provide different results on its influence on practitioners' diagnostic decisions. For example, the *Conners 3* (Conners, 2008) provides a percentage of likelihood of a particular diagnosis. Having an instrument that specifically states there is, for example, an 85% likelihood of ADHD might have more of a direct influence on a diagnosis than a clinically significant standard score. If other behavior rating scales are revised or created to provide the same type of information, such results might have a greater influence on school psychologists' and other practitioners' diagnostic impressions. In that respect, the results of this study could be considered baseline data. If such reporting of results on behavior rating scales become commonplace, future research could replicate this study to assess how results presented in such a form impact diagnostic impressions. A greater influence on diagnostic impressions would be unfortunate, as all the limitations of behavior rating scales described by McConaughy and Ritter (2008) and others would still exist. School psychologists and other practitioners would still need to obtain a variety of data from multiple sources for diagnostic purposes.

Accurate diagnoses are important in the fields of education and psychology. Both under- and over-diagnosing can be problematic for numerous reasons. This study and a few others (e.g., Barnard-Brak et al., 2013) are starting to examine what influences school psychologists' diagnostic decisions. Not only are additional studies needed with school psychologists, but with other diagnosticians as well. For example, it is unknown how these results might generalize to other developmental disorders or to clinical psychologists or psychiatrists. Professionals in different settings (e.g., private practice, hospital) may rely on different sources or types of information to make diagnostic decisions. Clearly, additional research is needed to fully understand how mental health professionals make diagnostic decisions.

## References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Angello, L. M., Volpe, R. J., DiPerna, J. C., Gureasko-Moore, S. P., Gureasko-Moore, D. P., Nebrig, M. R., & Ota, K. (2003). Assessment of attention-deficit/hyperactivity disorder: An evaluation of six published rating scales. *School Psychology Review, 32*, 241-262.
- Barnard-Brak, L., Stevens, T., Robinson, E., & Holt, A. (2013). School psychologist diagnostic decision-making: A pilot study. *Psychology in the Schools, 50*, 567-576. <http://dx.doi.org/doi.org/10.1002/pits.21694>

- Carter, A. S., Briggs-Gowan, M. J., & Davis, N. O. (2004). Assessment of young children's social-emotional development and psychopathology: Recent advances and recommendations for practice. *Journal of Child Psychology and Psychiatry*, *45*, 109-134. <http://dx.doi.org/doi.org/10.1046/j.0021-9630.2003.00316.x>
- Chafouleas, S., Riley-Tillman, T. C., & Sugai, G. (2007). *School-based behavioral assessment: Informing intervention and instruction*. New York, NY: Guilford.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Conners, K. (2008). *Conners third edition*. Los Angeles, CA: Western Psychological Services.
- Curtis, M. J., Grier, J. E., & Hunley, S. A. (2004). The changing face of school psychology: Trends in data and projections for the future. *School Psychology Review*, *33*, 49-66.
- Curtis, M. J., Lopez, A. D., Castillo, J. M., Batsche, G. M., Minch, D., & Smith, J. C. (2008). The status of school psychology: Demographic characteristics, employment conditions, professional practices, and continuing professional development. *Communique*, *36*, 27-29.
- Elliott, S. N., Busse, R. T., & Gresham, F. M. (1993). Behavior rating scales: Issues of use and development. *School Psychology Review*, *22*, 313-321.
- Gnys, J., Willis, W. G., & Faust, D. (1995). School psychologists' diagnoses of learning disabilities: A study of illusory correlation. *Journal of School Psychology*, *33*, 59-73. [http://dx.doi.org/doi.org/10.1016/0022-4405\(94\)00023-2](http://dx.doi.org/doi.org/10.1016/0022-4405(94)00023-2)
- Groenier, M., Beerthuis, V., Pieters, J., Witteman, C., & Swinkels, J. (2011). Psychologists' diagnostic processes during a diagnostic interview. *Psychology*, *2*, 917-924. <http://dx.doi.org/doi.org/10.4236/psych.2011.29138>
- Kavale, K., & Andreassen, E. (1984). Factors in diagnosing the learning disabled: Analysis of judgmental policies. *Journal of Learning Disabilities*, *17*, 273-278. <http://dx.doi.org/doi.org/10.1177/002221948401700504>
- Kretchman, D. (2007). *Incremental validity: Impact on decision making in schools*. (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses. (3276991)
- Lowe, J., Pomerantz, A. M., & Pettibone, J. C. (2007). The influence of payment method on psychologists' diagnostic decisions: Expanding the range of presenting problems. *Ethics and Behavior*, *17*, 83-93. <http://dx.doi.org/doi.org/10.1080/10508420701310141>
- López, S. R. (1989). Patient variable biases in clinical judgment: Conceptual overview and methodological considerations. *Psychological Bulletin*, *106*, 184-203. <http://dx.doi.org/doi.org/10.1037/0033-2909.106.2.184>
- McConaughy, S. H., & Ritter, D. R. (2008). Best practices in multimethod assessment of emotional and behavioral disorders. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 697-715). Bethesda, MD: National Association of School

Psychologists.

Merrell, K. W. (2008). *Behavioral, social, and emotional assessment of children and adolescents* (3rd ed.). New York, NY: Lawrence Erlbaum Associates.

Muroff, J. R., Jackson, J. S., Mowbray, C. T., & Himle, J. A. (2007). The influence of gender, patient volume and time on clinical diagnostic decision making in psychiatric emergency services. *General Hospital Psychiatry*, 29, 481-488.  
<http://dx.doi.org/doi.org/10.1016/j.genhosppsy.2007.08.008>

Reid, R., & Maag, J. W. (1994). How many fidgets in a pretty much: A critique of behavior rating scales for identifying students with ADHD. *Journal of School Psychology*, 32, 339-354.  
[http://dx.doi.org/doi.org/10.1016/0022-4405\(94\)90032-9](http://dx.doi.org/doi.org/10.1016/0022-4405(94)90032-9)

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd ed.). Circle Pines, MN: AGS Publishing.

Tobin, R. M., Schneider, W. J., Reck, S. G., & Landau, S. (2008). Best practices in the assessment of children with Attention Deficit Hyperactivity Disorder: Linking assessment to response to intervention. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 617-631). Bethesda, MD: National Association of School Psychologists.

Sattler, J. M. (2002). *Assessment of children: Behavioral and clinical applications* (4<sup>th</sup> ed.). San Diego, CA: Jerome M. Sattler, Publisher.

Shapiro, E. S., & Heick, P. F. (2004). School psychologist assessment practices in the evaluation of students referred for social/behavioral/emotional problems. *Psychology in the Schools*, 41, 551-561. <http://dx.doi.org/doi.org/10.1002/pits.10176>

Sox, H. C., Higgins, M. C., & Owens, D. K. (2013). *Medical decision making* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Spengler, P. M., White, M. J., Egisdóttir, S., Maugherman, A. S., Anderson, L. A., Cook, R. S., ... Rush, J. D. (2009). The meta-analysis of clinical judgment project: Effects of experience on judgment accuracy. *The Counseling Psychologist*, 37, 350-399.  
<http://dx.doi.org/doi.org/10.1177/0011000006295149>

---

**Appendix: Scenarios used for data collection.**

**Scenarios 1 & 2** (The only difference between these two scenarios is the BASC-2 score noted in parentheses in the last bullet point.)

An 8-year-old male student is referred by his teacher for a psychoeducational evaluation due to academic concerns. Assessment information includes the following:

- Teacher interview indicates the student:
  - has difficulty keeping track of his things
  - often forgets to turn in homework assignments
  - he has trouble staying in his seat
  - often asks to sharpen his pencil or get something from his backpack
- Parent interview indicates:
  - he is rather active
  - he needs a lot of re-direction in order to comply with directions or complete household chores
- A direct classroom observation indicates the target student was on-task 40% of the intervals observed, while a peer was on-task 85% of the same time period.
- The *Behavior Assessment System for Children, Second Edition* (BASC-2) completed by the teacher resulted in T scores of 73 (**Scenario 2 = 57**) on both the Hyperactivity and Attention Problems scales.

**Scenarios 3 & 4** (The only difference between these two scenarios is the BASC-2 score noted in parentheses in the last bullet point.)

An 8-year-old male student is referred by his teacher for a psychoeducational evaluation due to academic concerns. Assessment information includes the following:

- Teacher interview indicates the student:
  - is usually a good student
  - he sometimes has difficulty keeping track of his things
- Parent interview indicates:
  - he follows directions
  - sometimes he is rather active
- A direct classroom observation indicates the target student was on-task 85% of the intervals observed, while a peer was on-task 85% of the same time period.
- The *Behavior Assessment System for Children, Second Edition* (BASC-2) completed by the teacher resulted in T scores of 73 (**Scenario 4 = 57**) on both the Hyperactivity and Attention Problems scales.

**Copyright Disclaimer**

Copyright reserved by the author(s).

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).