

5-2010

Diagnostic Decision-Making: How Much Do Behavior Rating Scales Influence School Psychologists?

Lesley Ann Higgins

Western Kentucky University, lesley.higgins06@gmail.com

Follow this and additional works at: <http://digitalcommons.wku.edu/theses>



Part of the [Child Psychology Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), and the [Educational Psychology Commons](#)

Recommended Citation

Higgins, Lesley Ann, "Diagnostic Decision-Making: How Much Do Behavior Rating Scales Influence School Psychologists?" (2010). *Masters Theses & Specialist Projects*. Paper 157.
<http://digitalcommons.wku.edu/theses/157>

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.

DIAGNOSTIC DECISION-MAKING: HOW MUCH DO BEHAVIOR RATING
SCALES INFLUENCE SCHOOL PSYCHOLOGISTS?

A Thesis
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, KY

In Partial Fulfillment
Of the Requirements of the Degree
Specialist in Education

By
Lesley Ann Higgins
May 2010

DIAGNOSTIC DECISION-MAKING: HOW MUCH DO BEHAVIOR RATING
SCALES INFLUENCE SCHOOL PSYCHOLOGISTS?

Date Recommended __May 11, 2010__

Carl Myers
Director of Thesis

Reagan Brown

Elizabeth Jones

Dean, Graduate Studies and Research

Date

Acknowledgements

There are so many individuals I would like to thank for contributing in the success of this project and for my success in graduate school. First, I would like to thank God for always watching over me and guiding my life. There were times that I felt like everything was out of control, but my faith always kept me strong. Second, I would like to thank Dr. Carl Myers, my thesis director, for all his hard work and dedication throughout this process. Dr. Myers, you pushed me and gave me the constructive criticism and support that I needed. I could not have done this without you! I also want to thank my other committee members, Dr. Elizabeth Jones and Dr. Reagan Brown for their guidance in improving this project. Most importantly, I would like to thank my husband Matt for always believing in me and motivating me to reach my goals. I am fortunate to have a husband who was able to support me financially and emotionally throughout these three years of graduate school, so that I could have a career that I love! Lastly, I would like to thank my parents and sister for always being there for me during my times of doubt and believing that I could reach my dreams. Their support and guidance helped me get to where I am today and for that I will be eternally grateful! Thank you all!

Table of Contents

	Page
List of Tables.....	iii
Abstract.....	iv
Introduction.....	3
Literature Review.....	5
Method.....	17
Results.....	21
Discussion.....	25
References.....	28
Appendix A: Human Subjects Review Board Approval.....	32
Appendix B: Four Scenarios.....	34
Appendix C: Postcard for Responses.....	39

List of Tables

	Page
Table 1: Participant Demographics.....	18
Table 2: Mean Ratings on Participants' Use of the BASC-2 and Frequency of Providing ADHD Assessment Information.....	22
Table 3: Mean Ratings on Scenarios.....	22
Table 4: Results of <i>t</i> -test Pairings.....	23

DIAGNOSTIC DECISION-MAKING: HOW MUCH DO BEHAVIOR RATING SCALES INFLUENCE SCHOOL PSYCHOLOGISTS?

Lesley Ann Higgins

May 2010

40 pages

Directed by: Dr. Carl Myers, Dr. Elizabeth Jones, and Dr. Reagan Brown

Department of Psychology

Western Kentucky University

Behavior rating scales are commonly used as part of the evaluation process throughout the field of psychology. Behavior rating scales help assess social, emotional, and/or behavioral problems in children, adolescents, and teens. Behavior rating scales indicate the severity of problem behaviors compared to a normative sample. Four scenarios were developed that varied scores on a behavior rating scale and the amount of other information that supported a specific diagnosis. A rating of the likelihood of a diagnosis was requested to see how much influence behavior rating scale scores have on diagnostic decision-making. Each of the four scenarios was sent to 200 school psychologists across the country for a total of 800 potential participants. An overall response rate of 37.5% was achieved. The findings revealed that behavior rating scales do have some influence on school psychologists' diagnostic decision-making. However, school psychologists put more weight on other supporting information, such as classroom observations and teacher and parent reports, than on behavior rating scale scores when making a diagnostic decision.

Introduction

Behavior rating scales are standardized, norm-referenced instruments that are commonly used across the field of psychology to evaluate perceptions of an individual's behavior or social-emotional functioning. Because behavior rating scales are norm-referenced, psychologists are able to use obtained standard scores to compare problem behaviors of their client to others of the same age and even same gender. Behavior rating scales measure behaviors such as depression, withdrawal, anxiety, hyperactivity, attention problems, and somatization. Behavior rating scale scores are often used as part of an assessment when diagnosing disorders such as Attention-Deficit/Hyperactivity Disorder (ADHD).

When initially developed in the 1970s and 1980s, behavior rating scales rarely had sound statistical properties due to limited norm samples, poor item development, and limited statistical analyses (Merrell, 2001). The reliability and validity of behavior rating scales have improved over the past couple of decades (Sattler, 2002). With these improvements, psychologists have increased their usage of behavior rating scales (Shapiro & Heick, 2004). Behavior rating scales are also used more frequently because of a number of advantages they present, such as not being as time consuming as direct observations. However, best practices dictate that behavior rating scales should still be used in conjunction with direct observations, parent and teacher reports, and other assessment results. Behavior rating scales should never be used in isolation to make diagnostic decisions (Reid & Maag, 1994).

While it is recommended that psychologists do not use results from behavior rating scales in isolation for diagnostic purposes, it is unknown how much influence a

behavior rating scale score has on psychologists' diagnostic decision-making. A review of the literature did not find any studies that directly addressed this issue. The present study examined the influence of behavior rating scale scores on school psychologists' diagnostic decision-making related to the possibility of ADHD. A national sample of 800 school psychology practitioners were sent scenarios that contained varied assessment information and behavior rating scale scores and were asked to rate the likelihood of ADHD.

Literature Review

Description of Behavior Rating Scales

Behavior rating scales use a standardized format to develop summary judgments about an individual's behavioral characteristics (Merrell, 2008). Behavior rating scales measure perceptions of specified behaviors rather than provide any direct measurement of the behaviors as might be done through systematic behavioral observations. Behavior rating scales have been developed for children as young as 18 months of age through adulthood. An example of the type of item on a behavior rating scale may be, "Is OK when things do not go his or her way" (Reynolds & Kamphaus, 2004). Typically, the rater is to indicate the frequency of occurrence for each item on a continuum (e.g., Never, Sometimes, Often, Almost Always). Behavior rating scales should be completed by individuals who know the child the best, such as parents and teachers (Merrell, 2008). Although measuring perceptions of behavior, behavior rating scales are considered a relatively objective assessment method, yielding more reliable data than either unstructured clinical interviews or projective-expressive techniques (Merrell, 2008).

Behavior rating scales are considered relatively objective because norms, typically based on large representative samples of children, have been developed to provide standard scores on the ratings. When behavior rating scales are developed, the items on the scales are picked through content analysis based on theoretical constructs, and groups of items that comprise specific behavioral constructs are determined through factor analytic procedures (Sattler, 2002). The user of a behavior rating scale can then determine a standard score on each of the various behavioral constructs for an individual child, allowing an interpretation of the severity of the child's behavior in relation to other

children of the same age, and possibly even the same gender, from the general population. The information obtained from behavior rating scales can be used to assist practitioners with diagnostic decision-making, such as treatment integrity and progress monitoring.

There are two types of behavior rating scales, multidimensional scales and unidimensional scales. Multidimensional rating scales evaluate a child's behavior across several constructs or areas of adjustment (Wingenfeld, 2002). These scales assess a variety of areas such as adaptive skills, attention problems, aggression, or any number of social or school problems. These multidimensional scales provide the psychologist with a way to screen for a number of problems and disorders rather quickly (McConaughy & Ritter, 2008). Unidimensional rating scales are disorder or problem specific scales; such scales focus on one specific behavior construct area such as Attention Deficit Hyperactivity Disorder (ADHD) or social skills (Wingenfeld, 2002). These scales are an important component in the assessment of a diagnostic category because they allow psychologists standardized ratings on the degree to which a child exhibits certain key behaviors (McConaughy & Ritter, 2008).

It is also important to clarify the difference between a rating scale and a checklist (Merrell, 2008). These two types of scales are related; however, they are not the same. A behavior checklist is useful for identifying behavioral problems or competencies. It lists a number of behavioral descriptors and if the rater thinks the symptom is present, he or she simply "checks" the item. The results of checklists are used to provide qualitative information about a person. As previously described, behavior rating scales provide a standardized way of estimating the degree to which a behavior is present (Merrell, 2008).

Over the past 15 years, the use of behavior rating scales by psychologists has increased and the use of projective tests has declined when assessing social and emotional behaviors in children (Shapiro & Heick, 2004). This shift in assessment practices has occurred due to several factors, with a primary factor being that projective techniques are thought to result in unreliable data (Merrell, 2008). Indeed, by their very nature, projective tests require the psychologist to draw an inference about a person's emotional status or personality based on ambiguous stimuli. As a result, relying on subjective inferences tends to lead to unreliable data more so than the use of behavior rating scales. Historically, behavior rating scales were seen as a last resort by clinicians because of their poor psychometric properties, but their current widespread usage is a result of the research base and increased technical adequacy put behind behavior rating scales (Shapiro & Heick, 2004). Merrell (2008) also noted that acceptance among clinicians has led to more widespread usage of behavior rating scales.

Advantages of Behavior Rating Scales

According to Merrell (2001), there are six primary advantages for using behavior rating scales. The first one being that behavior rating scales are less expensive because they require less professional time and training to learn to administer. Second, behavior rating scales are able to provide data on low-frequency but severe behaviors, which can often go unobserved through direct observations. Third, behavior rating scales give more reliable data than unstructured interviews because of their objective format. Fourth, behavior rating scales can be used to assess children with low verbal skills or uncooperative attitudes. Fifth, behavior rating scales can tap into others' observations that are based in the child's environment (e.g., home, school) over long periods of time. The

sixth advantage is that behavior rating scales capture the judgments and observations of people who are invested in the child's well-being, such as parents and teachers.

There are numerous other advantages to using behavior rating scales as described by McConaughy and Ritter (2008). Behavior rating scales are standardized and give quantifiable information that can yield indexes of reliability and validity. Behavior rating scales can be used across a wide age range; therefore, making them an economical resource for practitioners. Rating scales exist for children as young as 18 months and can be used for students up through early adulthood. The majority of the behavior rating scales can be completed in 10 to 15 minutes and can be scored quickly by hand or computer. Behavior rating scales can provide data on a broad range of potential problems. Normative data provides a standard for judging the severity of problems by comparing an individual to a large sample of other children representative of the general population. Standardized behavior rating scales can also be used to compare similar data from multiple informants, such as parents, teachers, and even the students themselves.

Behavior rating scales are noted to be advantageous because they are easy to administer, relatively objective, time efficient, and provide useful data for screenings and evaluations (Angello et al., 2003; Chafouleas, Riley-Tillman, & Sugai, 2007; Elliot, Busse, & Gresham, 1993). It takes much less time for a parent and/or teacher to complete a behavior rating scale than it does for a school psychologist to obtain a representative picture of a student's behaviors through observations or interviews. Even when a school psychologist directly observes a behavior of concern, it becomes a subjective opinion as to whether the observed level or severity of the target behavior (e.g., activity level) is at a significantly high level.

Disadvantages of Behavior Rating Scales

Although behavior rating scales can provide good 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 about behaviors of concern (Angello et al., 2003; McConaughy & Ritter, 2008). Such information is critical to determine an appropriate behavioral intervention plan.

Behavior rating scales, like any measurement device, are not truly objective measures of an individual's competencies and problems because they involve people's perceptions of problems. "Such scales, despite their apparent objectivity, are simply quantifications of adult opinions. School psychologists must be wary of the seductive quality of this pseudo objectivity" (Reid & Maag, 1994, p. 348). Reid and Maag were warning psychologists not to rely on scores from behavior rating scales for diagnostic purposes. Thus, another critical limitation of behavior rating scales is that they cannot be used solely to make a formal diagnosis (Chafouleas et al., 2007; Sattler, 2002). As noted by Reid and Maag (1994), "there is no magic number on any scale that invariably means a student should be diagnosed" (p. 348).

The reliance on people's perceptions is also problematic for a number of other reasons as well. Behavior ratings can be impaired or influenced by the rater's memory, values, attitudes, and motivations, as well as situational factors (McConaughy & Ritter,

2008). Some parents or teachers may intentionally make their ratings of the child better or worse for varying reasons. Teachers see children in different contexts while at school. For example, a reading teacher may see more problem behaviors from a child than an art teacher. Therefore, their ratings may differ greatly because of the context in which they see the child. Differences between raters can be unintentional as well because people's expectations for certain behaviors and tolerance for misbehavior can differ greatly (Sattler, 2002). A number of studies have documented differences between raters (e.g., two teachers, parents and teachers, mothers and fathers) on behavior rating scales completed on the same children (Bingham, Loukas, Fitzgerald, & Zucker, 2003; Cai, Kaiser, & Hancock, 2004; Mandal, Olmi, & Wilczynski, 1999). Psychologists are told to expect rating scales to reveal different levels of problem behaviors reported by parents versus teachers or one teacher versus another teacher and mother versus father (McConaughy & Ritter, 2008).

The use of behavior rating scales with ethnically diverse populations has also been questioned and may be a limitation. While some authors reported behavior rating scales show no significant difference associated with ethnicity (Goh, 1997; Hosterman, DuPaul, & Jitendra, 2008), others noted there was inconclusive evidence regarding the appropriateness of behavior rating scale use with culturally diverse populations (Angello et al., 2003; Epstein, March, Conners, & Jackson, 1998; Reid et al., 2000). The primary criticism is that some behavior rating scales are not constructed to adequately represent diverse populations (Manz, Fantuzzo, & McDermott, 1999).

Use of Behavior Rating Scales

Behavior rating scales have traditionally been used as part of the screening and identification process for children referred for special education services (Chafouleas et al., 2007; Elliot et al., 1993). In fact, rating scales along with interviews and observations are found to be the one of the primary assessment methods used in 60% to 90% of cases (Shapiro & Heick, 2004). According to Merrell (2008), there are three best practices related to the use of behavior rating scales. First, behavior rating scales can be used for screening and identification purposes. Second, behavior rating scale data should be gathered from multiple raters across different settings to provide a broader picture of a child's behavior. Third, behavior rating scales can also be used to monitor progress of interventions; they provide evidence of the effectiveness of planned interventions. No one seems to dispute the second "best practice" described by Merrell (2008). However, concerns have been raised in the literature related to the first and third "best practice."

Although Merrell (2008) indicated that a "best practice" use of behavior rating scales is for progress monitoring of interventions, a number of researchers have raised concerns about such a use. During the early years of behavior rating scales, Wilson and Prentice-Dunn (1981) stated that overall standard scores on behavior rating scales could not be used to relate change directly to specific treatments. Other authors have continued to express the same cautions. Hosp, Howell, and Hosp (2003) assessed the usefulness of behavior rating scales for monitoring student progress and found that educators are not able to determine if the decreasing problem behavior is being replaced with a positive one. Chafouleas et al. (2007) stated, "Most behavior rating scales are not designed to be sensitive to incremental change in behavior. Thus, their use is limited to long-term

monitoring, such as in an evaluative capacity” (p. 113). More direct means of measuring specific behaviors of concern are recommended for progress monitoring purposes.

The use of behavior rating scales for screening and identification purposes is the most widespread use of the instruments (Merrell, 2008). 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). To address this concern “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). When behavior rating scales are used in conjunction with other assessments the concerns over their validity are lessened (Merrell, Streeter, & Boelter, 2001).

Influences on Diagnostic Decision-Making

As previously mentioned, behavior rating scales are primarily and frequently used as part of the screening and identification process for children referred for special education services (Elliot et al., 1993; Merrell, 2008). Furthermore, many authors caution against using behavior rating scales alone for making diagnostic or placement decisions (Carter et al., 2004; Chafouleas et al., 2007; Reid & Maag, 1994; Sattler, 2002). Although a multi-method assessment approach is recommended for diagnostic assessments (Angello et al., 2003), it is unknown how results from behavior rating scales inform and influence psychologists’ diagnostic decision-making processes. An electronic review of

the literature could not find any studies that examined how much influence a standard score on a behavior rating scale might have on a psychologist's diagnosis.

A review of the literature found some researchers have examined the broader topic of what influences a psychologist's diagnosis. For example, psychologists in the clinical setting may be influenced by the method of payment (insurance versus out-of-pocket) when diagnosing a client with a disorder (Lowe, Pomerantz, & Pettibone, 2007). Lowe et al. gave psychologists a survey in which the two hypothetical clients were described as paying either through managed care or out-of-pocket. The first vignette described a client with symptoms of Social Phobia that fell short of the diagnostic criteria and the second vignette described a client with symptoms of Attention-Deficit Hyperactivity Disorder (ADHD) that also fell short of the diagnostic criteria. Each of the vignettes was designed so that the symptoms could not be the result of a recent stressor to ensure that Adjustment Disorder was not a possibility. The results indicated that 51% of participants assigned an ADHD diagnosis to the client paying with managed care, whereas only 27% assigned the diagnosis to an identical client paying out-of-pocket. As regards to the social phobia vignette, the results indicated that 92% of clients paying with managed care were diagnosed, whereas only 69% of clients paying out-of-pocket were diagnosed (Lowe et al., 2007).

In a study by Gnys, Willis, and Faust (1995), researchers surveyed nationally certified school psychologists to see if diagnostic decisions related to learning disabilities were based on false beliefs. Each participant was given an information packet and a brief decision-making questionnaire. There were six different versions of the packets and each participant received only one packet with a cover page that requested participation and

ensured anonymity. Each packet included an examinee's hypothetical scores on the Wechsler Intelligence Scale for Children-Revised (WISC-R), Woodcock-Johnson Psycho-Educational Battery (WJPB), and the Child Behavior Checklist (CBCL), as well as a brief description of each of these tests and some identifying information about the examinee. The packets included summary pages that varied in reporting high, medium, or low intersubtest scatter on the WISC-R. The participants were then asked to rate the probability that the child represented by the materials was learning disabled on a scale from 0% (not probable) to 100% (certain).

Gnys et al. (1995) hypothesized that school psychologists' diagnoses would be influenced by an illusory belief in the association between level of WISC-R intersubtest scatter and a learning disability. They also predicted that the effects of this illusory belief would not be canceled by the availability of valid diagnostic information like the academic achievement data from the WJPB. The results of their survey supported both hypotheses. The participants were more likely to diagnose a learning disability when there was a high degree of intersubtest scatter on the WISC-R. Additionally, the school psychologists were much more likely to diagnose a learning disability when academic achievement was low than when it was at an average level. Their results demonstrated that certain factors (i.e., belief in an illusory correlation between a learning disability and the WISC-R intersubtest scatter, achievement level) influenced their diagnostic decision-making (Gnys et al., 1995).

Purpose of the Present Study

As previously noted, the use of behavior rating scales by psychologists has increased greatly over the past 15 years (Shapiro & Heick, 2004). The increased use in

behavior rating scales is due to their many advantages they present related to assessment. The primary use of behavior rating scales is for screening and diagnostic purposes (Merrell, 2008). Despite cautions that a psychologist should not rely solely on a score from a behavior rating scale for diagnostic purposes, it appears the developers of recent behavior rating scales are encouraging the use of such scales for diagnostic purposes. For example, the Clinical Assessment of Behavior (CAB, Bracken & Keith, 2004) is a multidimensional behavior rating scale for children that goes beyond providing scores on typical behavioral clusters such as aggression or activity level. The CAB provides standard scores on specific disabilities (i.e., learning disability, mental retardation, attention-deficit/hyperactivity and autism spectrum behaviors). Another example is the most recent edition of the Conners 3 behavior rating scale (Conners, 2008). The computerized scoring provides a comparison of the child's behaviors to DSM-IV diagnostic criteria of ADHD and even provides a percentage of likelihood that the child has ADHD.

It is unknown if any psychologists use the results of a behavior rating scale to directly diagnose specific disorders (e.g., a psychologist concludes a child has ADHD simply because of a high score on the ADHD scale). It would be difficult to determine accurately, based on psychologists' self-ratings, if such inappropriate practices exist. A broader question is how much influence does a behavior rating scale have on a psychologist's decision-making processes? Two studies were identified that evaluated influences on psychologists' diagnostic decisions. Both studies found that irrelevant information influenced diagnoses. No studies were identified that looked at how much influence standard scores on behavior rating scales have on diagnostic decision-making.

The purpose of the current research project is to see how much influence a behavior rating scale has on school psychologists' diagnostic decision-making.

The overall research question for this study asks how much influence does a behavior rating scale have on school psychologists' diagnostic decision-making. School psychologists will be asked to rate the likelihood of ADHD after reading one of four scenarios that contain a variety of pieces of information consistent or inconsistent with an ADHD diagnosis. Behavior rating scale results will vary in each of those scenarios.

Hypothesis 1. When all information consistently supports a diagnosis of ADHD, school psychologists are more likely to rate the student as having ADHD. This condition is essentially a control condition to determine school psychologists' likelihood of diagnosing ADHD based on a limited amount of information.

Hypothesis 2. When all information does not support a diagnosis of ADHD, school psychologists are unlikely to rate the student as having ADHD. This condition is essentially a control condition to determine school psychologists' likelihood of not diagnosing ADHD given little information to support such a diagnosis.

Hypothesis 3. School psychologists' diagnostic impressions are heavily influenced by scores on a behavior rating scale. When behavior rating scale results are clinically significant, such high scores will outweigh other pieces of assessment information.

Method

Participants

Approval from Western Kentucky University's Human Subjects Review Board was obtained to conduct this study (see Appendix A). The potential sample of participants included 800 randomly selected school psychology practitioners that were members of the National Association of School Psychologists. Each of the four scenarios was sent to 200 school psychologists across the United States. Usable responses were received from 300 participants for a return rate of 37.5%. Four letters (0.5%) were returned because of an incorrect address and 12 postcards (1.5%) were returned that did not include a rating of the scenario. Demographic information regarding participants' years of experience, gender, and highest degree are presented in Table 1.

All four groups are fairly similar across all demographic variables, such as gender, type of degree, and years of experience. A one-way ANOVA including each of those variables found no significant differences between the four groups. The respondents were predominately female (79.9%), although that percentage is similar to statistics from the National Association of School Psychologists (NASP, 2008), which indicated that 74% of school psychologists are female. Similarly, 70.7% of this sample had non-doctoral degrees and NASP (2008) reported that 7 out of 10 school psychologists hold non-doctoral degrees. Therefore, because this sample's demographics regarding gender and highest degree are comparable to the national statistics provided by NASP, it is assumed a representative sample was obtained and the results can be generalized to the field as a whole.

Table 1

Participant Demographics

	Group 1 (n = 73)	Group 2 (n = 61)	Group 3 (n = 83)	Group 4 (n = 83)	Total (n = 300)
Gender					
Males	23.6%	16.4%	22.9%	16.9%	20.1%
Females	76.4%	83.6%	77.1%	83.1%	79.9%
Degree					
Masters	21.9%	18.0%	15.7%	26.5%	20.7%
Specialist	49.3%	54.1%	48.2%	49.4%	50.0%
Doctorate	28.8%	27.9%	36.1%	24.1%	29.3%
Mean Years of Experience					
	14.6	15.7	15.1	15.4	15.2

Instrument

Four scenarios were created to assess the influence of behavior rating scales on school psychologists' diagnostic decision making. The final scenarios are presented in Appendix B. All scenarios include information from teacher and parent interviews, as well as systematic classroom observation data and standard scores on the "hyperactivity" and "attention problems" scales from a commonly used behavior rating scale (i.e., BASC-2). Scenario 1 includes parent and teacher interviews and a systematic observation that support an ADHD diagnosis; however, the scores from a behavior rating scale do not support an ADHD diagnosis. Scenario 2 includes the same parent and teacher interview

information and systematic observation information as Scenario 1; however, the behavior rating scale scores do support an ADHD diagnosis. In Scenario 3, none of the information provided supports an ADHD diagnosis. Scenario 4 contains the same information as Scenario 3 except that the behavior rating scale scores are the only bits of information that do support an ADHD diagnosis. T scores of 57 and 73 were chosen for the BASC-2 information. 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. After reading the scenario information, participants were asked to rate the likelihood the student has Attention-Deficit/Hyperactivity Disorder 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 local school psychologists. In addition to asking the pilot study group to complete the ratings, they were also asked to provide feedback on the scenarios and survey questions. Mean scores varied among each of the scenarios, suggesting the different information contained in each of the scenarios influenced ratings. In particular, Scenario 2 (where all information suggested ADHD) resulted in the highest mean score while Scenario 3 (where none of the information suggested ADHD) had the lowest mean score. Some respondents expressed concerns regarding the wording of one of the questions, “Have you previously diagnosed children as having Attention Deficit/Hyperactivity Disorder (ADHD)?” The school psychologists indicated that they technically do not “diagnose” ADHD in the schools and thought the question was irrelevant or they were uncertain how to respond to this question. Therefore, the question was altered to read, “Typically, how often do you provide assessment information for an ADHD evaluation?” Several

respondents also made comments about the student's grades in the scenarios and discussed how "adverse affect" influenced their ratings. (In special education, an "adverse affect" must be present before services can be provided.) It was not the intent of this study to have an adverse affect variable included in the information; therefore, the sentence about grades was excluded from each scenario.

Procedure

Address labels from 800 randomly selected members of the National Association of School Psychologists (NASP) were obtained from the NASP headquarters. It was requested that only school psychology practitioners be included and that others, such as trainers or students, be excluded. After developing and field testing the scenarios, each participant was mailed a cover letter explaining the purpose of the study and informed consent procedures, one scenario, and a pre-stamped postcard to indicate responses (see Appendix C). Each of the scenarios was mailed to 200 potential participants. Address labels were randomly assigned to the envelopes mailed to the NASP members. The returned postcards contained no identifying information, other than the scenario number. Therefore, no follow-up mailings could be conducted for non-respondents.

Results

A national sample of school psychology practitioners received one of four scenarios. Each of the scenarios varied behavior rating scale scores (i.e., average range or clinically significant) and other information suggesting the presence or absence of ADHD behaviors. The main survey question asked the school psychologists to indicate the likelihood of the student in the scenario having ADHD on a six point Likert scale. In addition, the school psychologists were also asked to rate how often they used the behavior rating scale mentioned in the scenarios (i.e., BASC-2) and they were asked how frequently they provided assessment information for an ADHD evaluation. As can be seen in Table 2, all groups had equivalent levels of experience with using the BASC-2, with mean scores at a “fairly frequent” level. In addition, all groups had equivalent levels of experience providing assessment information for ADHD evaluations. A one-way ANOVA using each of those variables found no significant differences among the four groups.

To address the research hypotheses, mean ratings for each of the scenarios are presented in Table 3 and results from a series of *t*-tests comparing all combinations of pairs of scenarios are presented in Table 4. Hypothesis 1 predicted that when all information consistently supports the diagnosis of ADHD, the ratings of the likelihood of ADHD would be the highest. This hypothesis was supported as that scenario (Scenario 2) had the highest mean rating of any of the scenarios. Although the mean score of 3.69 was significantly higher than ratings from the other three scenarios ($p < .001$), it was somewhat surprising that the mean rating was not even higher, given that all information

Table 2

*Mean Ratings on Participants' Use of the BASC-2 and Frequency of Providing ADHD**Assessment Information*

	Group 1 (n = 73)	Group 2 (n = 61)	Group 3 (n = 83)	Group 4 (n = 83)	Total (n = 300)
Use of BASC-2 ^a	2.93	3.13	2.79	2.76	2.89
ADHD assessments ^b	3.84	3.78	3.74	3.73	3.77

^aThese ratings were on a five point scale where 1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Fairly Frequent, and 5 = Often. ^bThese ratings were on a five point scale where 1 = very rarely/never, 2 = 1-3 times a year, 3 = 4-6 times a year, 4 = 7-9 times a year, and 5 = 10 or more times a year.

Table 3

Mean Ratings on Scenarios

Scenarios	Mean Rating	SD
1: Everything but the rating scales suggest ADHD	3.13	.83
2: Everything suggests ADHD	3.69	.81
3: Little information suggests ADHD	1.71	.61
4: Little information but the rating scales suggest ADHD	2.33	.68

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. All pairs of ratings were statistically significantly different at $p < .001$.

Table 4

Results of t-test Pairings

Group Pairing	<i>t</i> -value	Significance Level
1 vs. 2	3.93	.000
1 vs. 3	12.27	.000
1 vs. 4	6.60	.000
2 vs. 3	16.75	.000
2 vs. 4	10.91	.000
3 vs. 4	6.18	.000

supported a diagnosis of ADHD. A rating of 3.69 would indicate ratings between a “moderate possibility” and “likely.” Such a rating would suggest that school psychologists are being cautious about applying a diagnostic label even when the available information suggests the diagnosis.

Hypothesis 2 predicted that when all information consistently does not support the diagnosis of ADHD, the ratings of the likelihood of ADHD would be the lowest. This hypothesis was supported as that scenario (Scenario 3) had the lowest mean rating of any of the scenarios. The ratings for Scenario 3 were significantly lower than all other scenario ratings ($p < .001$).

Hypothesis 3 predicted that scores on a behavior rating scale would heavily influence school psychologists’ diagnostic impressions. This hypothesis was not supported. When the scenario included various pieces of information that supported

ADHD, but the behavior rating scale scores were in the average range, the mean rating of 3.13 for that scenario (Scenario 1) was significantly higher than the mean rating of 2.33 for Scenario 4, where little information but the rating scale scores suggested ADHD ($p < .001$). Such results suggest that school psychologists are putting more weight on various pieces of assessment information than on behavior rating scale scores.

Discussion

Behavior rating scales are becoming increasingly popular as an assessment tool with school psychologists. Recent behavior rating scales (e.g., Conners 3, Conners, 2008) seem to be moving toward providing specific diagnostic results. It is unknown, however, how the results from a behavior rating scale might influence a psychologist's decision about the presence or absence of a diagnosis. The purpose of the current research project was to see how much influence behavior rating scale scores have on the decision-making process of school psychologists.

Four scenarios were developed that varied information supporting or not supporting the diagnosis of ADHD. One scenario was designed where all information consistently supported a diagnosis of ADHD and another scenario had none of the information supporting a diagnosis of ADHD. For those scenarios, results were as predicted. School psychologists are most likely to indicate the presence of ADHD when all assessment information supports the diagnosis. School psychologists are very unlikely to rate the student as having ADHD when none of the information supports the diagnosis. It was noted that even when all information supports a diagnosis of ADHD, school psychologists' ratings of the likelihood of ADHD were only moderately strong with a mean of 3.69 on a five point Likert scale. Obtaining only moderate ratings for this scenario were initially surprising, but are probably very appropriate. Even though all the information in the scenarios supported the diagnosis of ADHD, it still contained a very limited amount of information. School psychologists in this sample appear to have been appropriately cautious in their diagnostic decision-making.

It was predicted that scores on a behavior rating scale would heavily influence school psychologists' diagnostic impressions. Such a hypothesis was not supported. The results from this study indicated that school psychologists put more weight in other supporting pieces of information than in behavior rating scale scores when indicating the likelihood of a student having ADHD. The results of this study did indicate that behavior rating scale scores do influence school psychologists' diagnostic decisions; however, the behavior rating scale scores do not overshadow other information that may support or not support a particular diagnosis. As previously noted, when behavior rating scales are used for diagnostic purposes they should 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). 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).

Limitations. A limitation of this study is that information was collected 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 37.5% response rate. However, despite the relatively low response rate, the participants' demographics seemed to reflect the population of school psychologists as a whole.

Strengths. A strength of this study was that it included a randomly selected national sample of school psychologists and the sample was comparative to

characteristics (e.g., gender, highest degree) of the field of school psychology. Another strength of this study is that it appears that no one else has previously conducted research examining the influence of behavior rating scale results on psychologists' diagnostic decision-making. Therefore, this research makes a unique contribution to the school psychology literature.

Summary. Behavior rating scales are a prominent part of a school psychologist's assessment procedures. An implication of this study is the knowledge and assurance that school psychologists are using all available pieces of assessment information to make diagnostic decisions about students. It is unclear, however, how the newer generation of behavior rating scales, that include scores on the likelihood of DSM-IV disorders, might influence school psychologists' diagnostic decisions. For future research, it would be interesting to replicate this study using a behavior rating scale such as the Conners 3 (Conners, 2008) in the scenarios.

References

- Angello, L. M., Volpe, R. J., Diperna, J. C., Gureasko-Moore, S. P., Gureasko-Moore, D. P., Nebrig, M. R., et al. (2003). Assessment of attention-deficit/hyperactivity disorder: An evaluation of six published rating scales. *School Psychology Review*, 32, 241-262.
- Bingham, C. R., Loukas, A., Fitzgerald, H. E., & Zucker, R. A. (2003). Parental ratings of son's behavior problems in high-risk families: Convergent validity, internal structure, and interparent agreement. *Journal of Personality Assessment*, 80, 237-251.
- Bracken, B. A., & Keith, L. K. (2004). *Clinical assessment of behavior professional manual*. Lutz, FL: Psychological Assessment Resources.
- Cai, X., Kaiser, A. P., & Hancock, T. B. (2004). Parent and teacher agreement on Child Behavior Checklist items in a sample of preschoolers from low-income and predominately African American families. *Journal of Clinical Child and Adolescent Psychology*, 33, 303-312.
- 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.
- Chafouleas, S., Riley-Tillman, T. C., & Sugai, G. (2007). *School-based behavioral assessment: Informing intervention and instruction*. New York: Guilford.
- Conners, K. (2008). *Conners third edition*. Los Angeles: Western Psychological Services.

- Elliot, S. N., Busse, R. T., & Gresham, F. M. (1993). Behavior rating scales: Issues of use and development. *School Psychology Review, 22*, 313-321.
- Epstein, J. N., March, J. S., Conners, C. K., & Jackson, D. L. (1998). Racial differences on the Conners Teacher Rating Scale. *Journal of Abnormal Psychology, 26*, 109-118.
- 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.
- Goh, D. S. (1997). Clinical utility of the Devereux behavior rating scale school form among culturally diverse children. *Psychology in the Schools, 34*, 301-308.
- Hosp, J. L., Howell, K. W., & Hosp, M. K. (2003). Characteristics of behavior rating scales: Implications for practice in assessment and behavioral support. *Journal of Positive Behavior Interventions, 5*, 201-208.
- Hosterman, S. J., DuPaul, G. J., & Jitendra, A. K. (2008). Teacher ratings of ADHD symptoms in ethnic minority students: Bias or behavioral difference? *School Psychology Quarterly, 23*, 418-435.
- 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.
- Mandal, R. L., Olmi, D. J., & Wilczynski, S. M. (1999). Behavior rating scales: Concordance between multiple informants in the diagnosis of attention-deficit/hyperactivity disorder. *Journal of Attention Disorders, 3*, 97-103.

- Manz, P. H., Fantuzzo, J. W., & McDermott, P. A. (1999). The parent version of the preschool social skills rating scale: An analysis of its use with low-income, ethnic minority children. *School Psychology Review*, 28, 493-504.
- 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. (2001). Assessment of children's social skills: Recent developments, best practices, and new directions. *Exceptionality*, 9, 3-18.
- Merrell, K. W. (2008). *Behavioral, social, and emotional assessment of children and adolescents* (3rd ed.). New York: Lawrence Erlbaum Associates.
- Merrell, K. W., Streeter, A. L., & Boelter, E. W. (2001). Validity of the home and community social behavior scales: Comparisons with five behavior-rating scales. *Psychology in the Schools*, 38, 313-325.
- National Association of School Psychologists. (2008). *Degree type of practicing school psychologists by gender and ethnicity*. Unpublished data from the 2004-05 NASP Membership Survey.
- 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.
- Reid, R., Riccio, C. A., Kessler, R. H., DuPaul, G. J., Power, T. J., Anastopoulos, A. D., et al. (2000). Gender and ethnic differences in ADHD as assessed by behavior ratings. *Journal of Emotional and Behavioral Disorders*, 8, 38-48.

- Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd edition). Circle Pines, MN: AGS Publishing.
- Sattler, J. M. (2002). *Assessment of children: Behavioral and clinical applications* (4th 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*(5), 551-561.
- Wilson, D. R., & Prentice-Dunn, S. (1981). Rating scales in the assessment of child behavior. *Journal of Clinical Child Psychology, 10*, 121-125
- Wingenfeld, S. A. (2002). Assessment of behavioral and emotional difficulties in children and adolescents. *Peabody Journal of Education, 77*, 85-105.

Appendix A

Human Subjects Review Board Approval



A LEADING AMERICAN UNIVERSITY WITH INTERNATIONAL REACH
HUMAN SUBJECTS REVIEW BOARD

In future correspondence, please refer to HS10-156, January 25, 2010

Lesley Higgins
c/o Dr. Carl Myers
Psychology
WKU

Lesley Higgins:

Your research project, *The Influence of Behavior Rating Scales on School Psychologists' diagnostics decision-making*, was reviewed by the HSRB and it has been determined that risks to subjects are: (1) minimized and reasonable; and that (2) research procedures are consistent with a sound research design and do not expose the subjects to unnecessary risk. Reviewers determined that: (1) benefits to subjects are considered along with the importance of the topic and that outcomes are reasonable; (2) selection of subjects is equitable; and (3) the purposes of the research and the research setting is amenable to subjects' welfare and producing desired outcomes; that indications of coercion or prejudice are absent, and that participation is clearly voluntary.

1. In addition, the IRB found that you need to orient participants as follows: (1) signed informed consent is not required; (2) Provision is made for collecting, using and storing data in a manner that protects the safety and privacy of the subjects and the confidentiality of the data. (3) Appropriate safeguards are included to protect the rights and welfare of the subjects.

This project is therefore approved at the Exempt from Full Board Review Level.

2. Please note that the institution is not responsible for any actions regarding this protocol before approval. If you expand the project at a later date to use other instruments please re-apply. Copies of your request for human subjects review, your application, and this approval, are maintained in the Office of Sponsored Programs at the above address. Please report any changes to this approved protocol to this office. A Continuing Review protocol will be sent to you in the future to determine the status of the project. Also, please use the stamped approval forms to assure participants of compliance with The Office of Human Research Protections regulations.

Sincerely,

Paul J. Mooney, M.S.T.M.
Compliance Coordinator
Office of Sponsored Programs
Western Kentucky University



cc: HS file number Higgins HS10-156

Appendix B

Four Scenarios

Scenario 1

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 57 on both the Hyperactivity and Attention Problems scales.

While the presence or absence of a disorder cannot be made solely on the information provided, given the information in the scenario, Please use the following Likert scale and circle your response on the enclosed postcard and drop it in the mail.

1=Not at all

4=Likely

2=Slight Possibility

5=Very Likely

3=Moderate Possibility

6=Definitely

Scenario 2

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 on both the Hyperactivity and Attention Problems scales.

While the presence or absence of a disorder cannot be made solely on the information provided, given the information in the scenario, what do you think is the likelihood that this student has Attention-Deficit/Hyperactivity Disorder? Please use the following Likert scale and circle your response on the enclosed postcard and drop it in the mail.

1=Not at all

4=Likely

2=Slight Possibility

5=Very Likely

3=Moderate Possibility

6=Definitely

Scenario 3

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 57 on both the Hyperactivity and Attention Problems scales.

While the presence or absence of a disorder cannot be made solely on the information provided, given the information in the scenario, what do you think is the likelihood that this student has Attention-Deficit/Hyperactivity Disorder? Please use the following Likert scale and circle your response on the enclosed postcard and drop it in the mail.

1=Not at all

4=Likely

2=Slight Possibility

5=Very Likely

3=Moderate Possibility

6=Definitely

Scenario 4

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 on both the Hyperactivity and Attention Problems scales.

While the presence or absence of a disorder cannot be made solely on the information provided, given the information in the scenario, what do you think is the likelihood that this student has Attention-Deficit/Hyperactivity Disorder? Please use the following Likert scale and circle your response on the enclosed postcard and drop it in the mail.

1=Not at all

4=Likely

2=Slight Possibility

5=Very Likely

3=Moderate Possibility

6=Definitely

Appendix C

Postcard for Responses

Circle your response to the scenario here:

(Not at all) 1-----2-----3-----4-----5-----6 (Definitely)

Years of experience as a school psychologist: _____

Gender: Male Female

Type of Degree: Masters Specialist Doctorate

State where employed: _____

How often do you use the *Behavior Assessment System for Children, 2nd Edition* (BASC-2)? *(circle a response below)*

Often Fairly Frequently Sometimes Rarely Never

Typically, how often do you provide assessment information for an ADHD evaluation? *(circle a response below)*

Very rarely/never 1-3 times/year 4-6 times/year

7-9 times/year 10 or more times/year