1969

Teachers' judgments of the academic achievement of children with and without characteristics of inattention, impulsivity, and hyperactivity.

Lyndsay Jenkins and Michelle Demaray

Teachers' Judgments of Academic Achievement of Children with and Without Characteristics of Inattention, Impulsivity, and Hyperactivity

Lyndsay N. Jenkins

Eastern Illinois University

Michelle K. Demaray

Northern Illinois University

Abstract

Some evidence suggests that teachers may have a negative bias against and be less accurate in academic judgments of students with behavioral characteristics of Attention-Deficit/Hyperactivity Disorder (ADHD). Thus, examining the accuracy of their judgments is an important area of investigation. The current study examined the accuracy of teachers' judgments on Curriculum-Based Measures (CBM) of reading, math, and writing for students with and without characteristics of inattention, impulsivity, and hyperactivity. Participants included 72 elementary students, 30 with characteristics of inattention, impulsivity, and hyperactivity (IIH Group) and 42 without these characteristics (Comparison Group). Overall, there were no major differences between groups in teachers' accuracy. Regarding indirect judgments (i.e., ratings of academic skills on a rating scale), teachers' ratings were positively related to reading and math CBM for both groups. There were few significant differences in teacher judgments between the groups of students with and without symptoms of IIH on indirect ratings of achievement, but when examining overestimation and accuracy of direct ratings, a different picture emerged. Teachers overestimated performance on reading for both groups (75 and 43 words, Comparison and IIH, respectively); however, this difference was not statistically significant. For math, teachers overestimated for both groups (7.5 and 11.5 points, Comparison and IIH, respective) and they overestimated significantly more for the Comparison group.

Keywords: Teacher judgments; academic achievement; curriculum-based measurement; ADHD

Teachers' Judgments of the Academic Achievement of Children with and Without Characteristics of Inattention, Impulsivity, and Hyperactivity

Teachers' judgments of students' academic skills inform decisions about curriculum and differentiation of instruction and can lead to referrals for additional instruction or special education evaluations. In schools using a Response to Intervention service delivery model, teachers' judgments are used as supplementary information in addition to school-wide academic screening processes (Fuchs & Fuchs, 2006). Even in schools that use universal screening, teachers often serve as a "gate keeper" for initiating supplemental academic instruction and/or special education referrals. Students with behaviors such as inattention, impulsivity, and hyperactivity (i.e., characteristics of Attention-Deficit/Hyperactivity Disorder; ADHD), including individuals with subclinical and clinical levels of ADHD, often have academic difficulties, which may necessitate supplementary academic support (i.e., Tier 2 or Tier 3 levels of academic instruction). Though not all children with inattention, impulsivity, and hyperactivity warrant an ADHD diagnosis, these behaviors can have a negative effect on educational performance. The current study included students with high levels of symptoms of inattention, impulsivity, and hyperactivity rather than only individuals with diagnoses of ADHD because educators may be challenged by working with students with these symptoms. Even if a formal diagnosis of ADHD has not been given, inattention, impulsivity, and hyperactivity can be challenging behaviors to manage.

Though it is generally accepted that teachers are fairly accurate in global academic judgments (Demaray & Elliott, 1998; Hoge & Coladarci, 1989; Südkamp, Kaiser, & Möller, 2011), some evidence suggests that teachers may be less accurate in their academic judgments of students with behavioral difficulties (Bennett, Gottesman, Rock & Cerullo, 1993). In order to provide the best services to students with characteristics of ADHD, school psychologists often rely on teachers'

judgments of academic performance, but researchers have not examined the accuracy of teacher judgments of students with characteristics of ADHD. Thus, it is important to explore whether educators are making accurate academic judgments for all students.

A meta-analysis by Südkamp and colleagues (2011) found that teachers generally have moderate to high levels of accuracy when judging students' academic performance; however, most studies focused primarily on global ratings of academic ability using rating scales or norm-referenced tests. Several conclusions can be drawn from studies regarding teachers' accuracy in predicting academic performance. First, teachers are fairly accurate judges when predicting global academic functioning via indirect measures (Demaray & Elliott, 1998; Hoge & Coladarci, 1989; Südkamp et al., 2011), with many studies reporting correlations between .60 and .70 among teacher judgments and actual achievement (Demaray & Elliott, 1998; Feinberg & Shapiro, 2003; Feinberg & Shapiro, 2009; Hoge & Coladarci, 1989). Second, accuracy of teacher judgments of reading and math is similar. (Südkamp et al., 2011). Third, teachers are less accurate in predictions on Curriculum-Based Measurement (CBM) assessments compared to global indirect assessments (e.g., teacher-rated academic skills rating scales) and tend to overestimate student performance (Eckert, Dunn, Codding, Begeny, & Kleinmann, 2006; Feinberg & Shapiro, 2003; Hamilton & Shinn, 2003). Fourth, no known studies have examined teachers' judgments of writing using CBM assessment procedures. Overall, teachers' accuracy in judging global academic performance is moderate, but an area that needs to be explored further is whether students' characteristics impact the accuracy of global academic judgments, as well as accuracy on CBM assessments.

Since CBM is a direct assessment of academic skills, it is an appropriate tool to utilize in investigating teachers' accuracy of their students' reading, math, and writing skills. Compared to the extensive literature regarding the technical adequacy of CBM (Shinn, 1998), there is little information

about teachers' accuracy in predicting performance on CBM. Moreover, even less information exists regarding the accuracy of teachers' judgments of CBM performance for students with varying ability levels and behavior challenges. Südkamp and colleagues (2011) stated the majority of studies examined teacher judgments on global measures of academic achievement, such as standardized achievement tests. Much less research has examined teachers' abilities to judge performance on direct measures of *specific* skills, such as reading fluency or math computation, which are measured by CBM assessments.

Teachers can accurately differentiate between students with and without learning disabilities (Gresham, Reschly, & Carey, 1987; Shinn, Tindal, & Spira, 1987), but when predicting lower achieving students' performance on a standardized test (Demaray & Elliott, 1998) or on oral reading fluency tasks (Begeny, Eckert, Montarello, & Storie, 2008), teachers are less accurate. Feinberg and Shapiro (2009) compared teachers' accuracy in predicting oral reading fluency CBM scores for students with differing achievement levels. They reported that teachers were less accurate at predicting performance on reading CBM measures for students with lower levels of achievement. Thus, another logical question is, can teachers accurately judge the academic performance for students with symptoms of inattention, impulsivity, and hyperactivity?

Teacher Judgment Accuracy for Students with characteristics of ADHD

To date, no study has examined teachers' accuracy in predicting performance on CBM measures in reading, math, and writing for children with characteristics of ADHD. This is a concern because there is evidence to suggest that some teachers have a negative bias against students with characteristics of ADHD. For example, Green, Beszterczey, Katzenstein, Park, and Goring (2002) found that teachers reported greater levels of stress and frustration when teaching and interacting with students with ADHD compared to their classmates without ADHD. Additionally, Eisenberg and

Schneider (2007) found that teachers have more negative perceptions of student academic skills when the student displayed characteristics of ADHD. Therefore, if teachers have negative feelings towards students with high levels of inattention, impulsivity, and hyperactivity, they may not be as accurate in their judgments of students with these behaviors.

The Current Study

The current study compared the accuracy of teacher judgments of academic skills for students with and without high levels of inattention, impulsivity, and hyperactivity using both direct (i.e., Curriculum-Based Measurement) and indirect (i.e., teacher-rated academic skills in reading, math, and critical thinking) academic assessments. Several research questions were addressed. First, are teachers less accurate in predicting performance for children without high levels of inattention, impulsivity, and hyperactivity (IIH) via an *indirect* measure of academic skills? Literature suggests that teachers are moderately accurate (Hoge & Coladarci, 1989; Südkamp et al., 2011), thus a similar level of accuracy was predicted in the current study when associating reading, math, and writing CBM scores with teacher-rated academic skills. Additionally, it was predicted that teachers would be less accurate overall for the IIH group (Bennett et al., 1993). Second, are teachers less accurate in predicting performance for children with high levels of IIH on *direct* academic assessments, including Reading-CBM, Math-CBM, and Writing-CBM? The limited amount of research in this area suggests that accuracy is lower for the group of students with high levels of inattention, impulsivity, and hyperactivity on global academic assessments (Eckert et al., 2006; Hamilton & Shinn; 2003; Südkamp et al., 2011), and for students with lower academic abilities (Feinberg & Shapiro, 2009). Third, do teachers overestimate students' performance on curriculum-based measurement and is the overestimation greater for students with or without high levels of IIH? Two previous studies suggest that teachers overestimate students' performance on reading (Hamilton & Shinn, 2003) and math

(Eckert et al., 2006) CBM; however, overestimation of students with and without behavioral difficulties has not been explored.

Method

Participants

Participants included 72 students from 21 different classrooms in three rural elementary schools (see Table 1 for demographic information). Children were categorized into two groups: students with high levels of parent-rated symptoms of inattention, impulsivity, and hyperactivity (IIH Group; $n = 30$) and students with low levels of these symptoms (Comparison Group; $n = 42$) (see Procedures for group inclusion criteria). There were 18 general education teachers (17 females, 1 male) who taught third ($n = 7$), fourth ($n = 3$), or fifth ($n = 8$) grade; half of the teachers had a Bachelor's degree and half had a Master's degree. The teachers reported a range of 1 to 32 years of teaching experience ($M = 10.25$ years).

Measures

Inattentive, impulsive, and hyperactive behaviors were measured via the ADHD-IV (DuPaul, Power, Anastopoulos, & Reid, 1998). Raters indicated frequency of 18 ADHD-characteristic behaviors on a 4-point scale ranging from 0 *(never or rarely*) to 3 (*very often*). The current study utilized the Home Version of the scale. Normative data for the ADHD-IV were gathered from 4000 teachers and 4500 parents of children in grades K - 12. The Home version has strong internal consistency (.92) and strong test-retest reliability (.85). Inter-rater agreement between Home and School versions was moderate, .41. There were significant correlations with the Conner's Parent Rating Scale (Conners, 1989) ranging from .10 to .81 (DuPaul et al., 1998). Factor analysis of the ADHD-IV Rating Scale shows a clear two-factor structure, which corresponds to the two subtypes of ADHD in the DSM-IV (DuPaul et al., 1998).

Teacher's indirect judgments of academic skills were measured via Academic Competence Evaluation Scales (DiPerna & Elliott, 2000), which is a norm-referenced rating scale for evaluating academic functioning of students. The ACES measures two domains of academic functioning: academic enablers and academic skills (DiPerna & Elliott, 2000). The Academic Skills scale of the teacher-rated version of the ACES was used, which consists of three subscales: Reading/Language Arts, Math, and Critical Thinking. These items are rated on a 5-point scale ranging from 1 (*far below*) to 5 (*far above*).

The ACES was standardized on a large, national sample of teachers and students. Reliability for the Academic Skills scale is demonstrated through strong internal consistency (alphas of .99 for all grade clusters) and a strong test-retest correlation of .96. Inter-rater correlations were adequate, ranging from .61 to .99. Validity for the ACES was demonstrated through factor analysis and correlations with similar measures. For example, correlations ranged from .38 to .87 and .56 to .90 between the ACES and Iowa Test of Basic Skills (ITBS; Hoover, Hieronymus, Frisbie, & Dunbar, 1993) and grade point averages, respectively (DiPerna & Elliott, 2000).

CBM probes were used to directly assess academic skills in reading, math, and writing. CBM is a set of standardized, short, timed tests, or probes, used to monitor a student's academic progress in basic skills (Shinn, 2002). Three types were used in the current study: Reading-CBM, Math-CBM, and Writing-CBM. A brief description and psychometric information for each is described below. Information regarding score calculation and use is in the Procedures section.

To assess reading skills, a total of three AIMSweb Reading-CBM (R-CBM) measures were administered to participants at their respective grade level. The mean scores of these three probes were used in analyses. An extensive literature has established strong technical adequacy for R-CBM. For example, Tindal, Germann, and Deno (1983) and Tindal, Marston, and Deno (1983) found coefficients

for test-retest reliability ranging from .92 to .97, and alternate form reliability ranging from .89 to .94. Marston (1989) reported in his review that criterion-related validity coefficients ranged between .73 and .81.

To assess math skills, participants completed three mixed-skill (addition, subtraction, multiplication, and division) math probes at their respective grade level and were given 2 minutes to complete each probe. The mean scores of these three math probes were used in analyses. The math probes were taken from Monitoring Basic Skills Progress (Fuchs, Hamlett, & Fuchs, 1998). The technical adequacy of Math-CBM (M-CBM) has also been explored. Tindal, Germann, and Deno (1983) reported test-retest coefficients ranging from .78 to .93 and alternate form reliability coefficients ranging from .48 to .72. Coefficients between .26 and .67 are reported when comparing mathematics fluency tasks with standardized mathematics achievement tests (Skiba, Magnusson, Marston, & Erickson, 1986).

Finally, students were presented with a story starter, which was "One day your family let you go to the pet store and pick out any pet you wanted. Write a story about that pet." The participants had 1 minute to think and 3 minutes to write the story. Percentage of Correct Word Sequences was calculated and used in statistical analyses. Percentages of Correct Word Sequences reflect the structural correctness of a written passage, such as grammar, punctuation, and capitalization. Strong evidence of reliability has been found for Correct Word Sequences, including test-retest coefficients, parallel-test correlations, split-half reliability coefficients, and interrater correlations (Amato & Watkins, 2011; McMaster & Espin, 2007; Videen, Deno, & Marston, 1982). For the current study, undergraduate research assistants were trained to score the writing probes and reliability checks were collected. In order to determine interscorer reliability, approximately 10 percent of the writing probes

were randomly selected to be scored by a different rater. The reliability for the percentage of Correct Word Sequences was .988 ($p < .01$).

Procedures

Data were collected as part of a larger study that investigated the relations among academic enablers and academic achievement in children with inattention, impulsivity, and hyperactivity. Teacher participants sent consent forms and ADHD-IV rating scales home with all their students. If the parents agreed to participate, they returned the consent form and the completed rating scale. Students also provided assent for their participation in the study.

Participants in the IIH group included all students with a score at or above the 90th percentile for the Total Score on the parent-rated form. For research purposes, DuPaul, Anastopoulos, Power, Reid, Ikeda, and McGoey (1998) recommend using an ADHD-IV score above the 90th or 93rd percentile combined with other data. The authors chose to include students with scores above the 90th percentile in the High IIH group. Participants in the Comparison Group consisted of students below the 90th percentile on the parent-rated ADHD-IV. The IIH Group ($M = 32.5$, $SD = 8.29$) and Comparison Group ($M = 5.1$, $SD = 4.15$) had significantly different Parent ADHD-IV Total scores, $t\ (70) = -18.47$, $p < .001$, with the IIH Group obtaining higher scores than the Comparison Group. The comparison participants were chosen at random for a maximum of four students per classroom. Teachers rated between one and four students ($M = 3.29$).

Students were given CBM probes in small groups (M-CBM and W-CBM) or individually (R-CBM) by a trained school psychology graduate student and research assistants. Teachers completed the ACES Academic Skills rating scale and predicted performance on the CBM measures. For R-CBM, teachers were given exact copies of the passages. They indicated how far the student would read in one minute and identified words the student would skip, struggle with for more than 3 seconds,

mispronounce, or replace with an incorrect word. For each of the three M-CBM probes, teachers predicted if the student would get each of the 25 items correct or incorrect. An item was considered incorrect if the teacher predicted the student would incorrectly solve the problem, skip the problem, or fail to complete that portion of the math sheet. For W-CBM, teachers indicated how much of a passage would be structurally correct including grammar, punctuation, and capitalization, also known as percentage of correct writing sequences.

In order to answer the study's research questions, several different analyses were used. Correlations were used to analyze the relation between student CBM scores and teacher-rated academic skills. When examining accuracy for the CBM measures, correlational analyses indicate the degree of the association between the actual total score and the predicted total score. Percent agreement calculations were used to determine if teachers' predictions and students' performances matched on the individual items, which is a more precise measurement of accuracy. For example, a teacher may indicate that a student will answer eight math problems correctly and the student may answer eight math problems correctly, however, the student may correctly answer eight different problems than what the teacher predicted. Percent agreement is based on the number of matches for actual and predicted performance at the item level.

Actual and Predicted Scores were used in correlational analyses and "matches" were defined for percent agreement analyses. For R-CBM, Actual Scores were the total number of words each participant read aloud correctly in one minute (i.e., the total number of words attempted minus any errors). Predicted Scores were the total number of words a teacher predicted a student would read correctly in one minute. For the purpose of calculating percent agreement, a "match" between students and teachers was defined. For R-CBM, a match occurred when 1) a student read a word correctly and the teacher predicted that it would be correct before the point at which the teacher predicted the student

stopped reading, or 2) when a student read a word incorrectly and their teacher predicted an error. A match did not occur if 1) the student read a word correctly and the teacher predicted an error, or 2) if the student had an error and the teacher predicted that the word would be correct.

Correlations and matches were also used for M-CBM analyses. Actual Scores included the total number of math problems correct in two minutes on each probe for each participant. Predicted Scores consisted of the total number of items the teacher thought a student would get correct. Matches occurred when 1) a student was correct and the teacher predicted that the student would be correct, or 2) when the student was incorrect and the teacher predicted that problem would be incorrect (or not attempted). A match did not occur when 1) the student was correct and the teacher predicted that they would solve the problem incorrectly (or not complete it), or 2) when the student was incorrect and the teacher predicted that the problem would be correct.

Finally, for W-CBM, teachers were asked to indicate the percentage of the passage that would be structurally correct (percentage of correct writing sequences). Matches could not be defined for writing measures because teachers had no way of knowing what the student wrote.

Results

Teacher Judgment of Reading-CBM

To examine the association between R-CBM scores and a teacher-rated indirect measure of reading skills, mean R-CBM scores and ACES Reading Skills were correlated. The correlation was significant for both groups, Comparison ($r = .64$, $p < .001$) and IIH ($r = .64$, $p < .001$), and there was no significant difference between the two correlations ($z = .03$, $p = .97$). A similar pattern emerged in the association between R-CBM and ACES Total Academic Skills, Comparison Group, $r = .52$, $p < .001$, and IIH Group, $r = .44$, $p = .015$, with no significant difference between the correlations ($z = .41$, $p = .68$), when using an r to z transformation.

To compare scores for the direct assessment of reading, the correlation between the mean Actual R-CBM score and the mean teacher-predicted R-CBM score was significant for the Comparison Group ($r = .48$, $p < .01$), but not for the IIH Group ($r = .33$, $p = .07$). However, these two correlations were not significantly different from each other ($z = .70$, $p = .48$). When comparing Actual and Predicted mean scores, teachers overestimated R-CBM performance for both groups. Though teachers overestimated more for the Comparison group (75 words) than the IIH Group (43 words), there was not a statistically significant difference in the overestimation, $F(1, 70) = 1.22$, $p = .273$. When examining how accurate the teachers were in their judgment of students' ability to read the individual words on all reading probes combined, teachers were significantly more accurate for students in the Comparison Group (94% agreement) compared to the IIH Group (89% agreement), $F(1, 70) = 5.41$, $p = .02$, which was a small effect size (Cohen's $d = .26$).

Teacher Judgment of Math-CBM

For math, the correlation between the students' actual Math CBM score and teacher-rated ACES Math Skills was significant for the Comparison Group ($r = .33$, $p = .04$), but not for the IIH Group ($r = .34$, $p = .072$), however, there was not a significant difference between the correlations ($z = .05$, $p = .96$). The correlation between M-CBM and teacher-rated ACES Total Academic Skills was not significant for either group, Comparison ($r = .28$, $p = .08$) and IIH ($r = .15$, $p = .43$).

For math direct assessment, the correlations between the mean Actual M-CBM score and the mean teacher-predicted M-CBM score was significant for the Comparison Group ($r = .54$, $p < .001$) and for the IIH Group ($r = .35$, $p = .05$). These two correlations were not significantly different from each other ($z = 1.58$, $p = .11$). Teachers overestimated M-CBM scores for both groups, but they overestimated significantly more for the Comparison group (11.5 points) than the IIH Group (7.5 points), $F(1, 70) = 9.19$, $p = .003$. Teacher's accuracy in predicting performance on the individual

problems was tested by comparing percent agreement. Teachers were significantly more accurate for the IIH group (66%) than the Comparison Group (58%), $F(1, 70) = 5.76$, $p = .02$, which was a large effect size (Cohen's $d = .73$).

Teacher Judgment of Writing-CBM

Since the ACES does not assess writing skills, but Reading, Math, and Critical Thinking, a correlation between Correct Writing Sequences and a teacher-rated ACES Score was not performed. The correlations between Actual percentage of Correct Word Sequences (Comparison Group 46%, IIH Group 32%) and teacher-predicted percentage of Correct Word Sequences (Comparison Group 84%, IIH Group 65%) was significant for the Comparison Group ($r = .62$, $p < .001$), but not for the IIH Group ($r = .24$, $p = .22$). When testing to see if these correlations were significantly different from each other, results indicated that the correlations were significantly different from each other ($z = 1.89$, $p = .05$). Percent agreement and overestimation could not be analyzed for W-CBM since the teachers did not know what the student would write.

Discussion

Professionals and parents regularly ask teachers how students are doing in the classroom. In order to provide the best services to students, educators and school psychologists often rely on teacher judgments. Much of the literature regarding teachers' judgments of academic performance has focused on global, indirect assessments (Hoge & Coladarci, 1989; Südkamp et al., 2011). The few studies looking at how well teachers predict performance on direct measures, like CBM, have found that teachers are less accurate at predicting performance on specific items and that they tend to overestimate performance on both reading and math (Eckert et al., 2006; Feinberg & Shapiro, 2009; Hamilton & Shinn; 2003; Südkamp et al., 2011). It is important to understand if teachers accurately predict student performance on direct measures in addition to indirect and standardized measures of

achievement. Additionally, previous research found that teachers are less accurate at predicting academic performance for children with behavioral difficulties (Bennett et al., 1993). Children with characteristics of ADHD have behavioral difficulties and unique learning needs which may suggest that teachers are less accurate at predicting academic performance for these children. To explore this further, the current study asked three primary research questions: First, are teachers more accurate in predicting performance for children without high levels of IIH via an indirect measure of academic skills? Second, are teachers less accurate in predicting performance for children with high levels of IIH on direct academic assessments, including Reading-CBM, Math-CBM, and Writing-CBM? Third, do teachers overestimate their predictions of performance on CBM and is the overestimation greater for students with or without high levels of IIH? Generally, there were few differences in accuracy of judgments between the Comparison and IIH groups on the indirect assessments, but differences in accuracy on direct assessments were found. There were also differences in how much teachers overestimated on the CBM direct measures.

On indirect assessments, correlations for reading were moderate to strong for students with and without high levels of IIH, but were lower for math, mostly in the weak range. When examining the pattern of results for students with and without symptoms of IIH, there were few significant differences between the groups on correlational analyses. Writing correlations were different between groups, with the comparison group having a moderate correlation, but a weak correlation for the students with symptoms of IIH. Previous work has also reported moderate to strong correlations for R-CBM (Feinberg & Shapiro, 2003; 2009) and lower correlations for math (Eckert et al., 2006). Writing has not been explored previously.

On direct assessments (i.e., Reading, Math, and Writing CBM), teachers had higher accuracy for reading (94% and 89%, Comparison and IIH, respectively) than for math (58% and 66%,

Comparison and IIH, respectively). This would suggest that predicting performance for Reading-CBM may be easier to judge than Math-CBM. Second, accuracy in teacher judgments of reading performance was significantly higher for the comparison group (94%) than the students with symptoms of IIH (89%). The opposite was true for math in that teacher accuracy was significantly higher for the students with symptoms of IIH (66%) than the students without these symptoms (58%). Teachers overestimated significantly more on Math-CBM students in the comparison group. Teachers predicted that these students would answer many problems correct (average of 15 problems correct in 2 minutes); however, the students actually had many incorrect responses (average score of 4 problems correct in 2 minutes). In essence, they overestimated more for the comparison group, which made them less accurate. It is not clear why teachers would be more accurate in predicting math performance for this group of students, which was opposite of the hypothesized results. For W-CBM, there was a significant difference, in the expected direction, for the association between the Comparison and IIH groups.

Previous studies have also reported overestimation of student performance on CBM measures (Eckert et al., 2006; Feinberg & Shapiro, 2003; Hamilton & Shinn, 2003). In the current study, teachers overestimated performance on reading for the comparison group by 75 words and by 43 words for students with high levels of IIH symptoms; however, this difference was not statistically significant. For math, teachers overestimated for both groups (7.5 and 11.5 points for students with and without symptoms of IIH, respectively) and they significantly overestimated more for the students in the comparison group. Other studies have reported overestimation of 22-40 words on reading CBM probes (Feinberg & Shapiro, 2003; Hamilton & Shinn, 2003). Overestimation on math CBM has not been explored at the item level previously; however, when Eckert and colleagues asked teachers to indicate if students would fall into the Mastery, Instructional, or Frustration level, they tended to

indicate that the majority of students would be at the mastery level when in fact the majority was at Instructional (Eckert et al., 2006). This was true for both reading and math assessments.

Implications for Practice

In practice, this and other studies suggest that teacher judgments of students' global skills are generally accurate, even for children with symptoms of inattention, impulsivity, and hyperactivity. However, when asking teachers to predict performance on specific CBM items, they are more likely to overestimate performance for students with and without these symptoms. Given the popularity of CBM and the importance placed on teachers' judgments of students' skills, it is essential that teachers be well-versed in students' skills as measured by CBM assessments. While multi-tiered systems of intervention rely heavily on student-derived screening data, professional judgment from teachers is still a critical piece of information. Best practice requires that schools collect data from multiple sources (i.e., standardized assessments, screening measures, classroom performance, teacher interviews) and that school psychologists want to be confident that the information they are gathering is accurate and valid. This is true for teacher reports as well.

Overall, there were few significant differences in teacher judgments between the groups of students with and without symptoms of IIH in the correlational analyses, but when examining overestimation and accuracy, a different picture emerges. Using different types of analyses allows for a more comprehensive picture about the accuracy of teachers' academic judgments. Correlational analyses alone suggest that there are few differences between teachers' indirect judgments of reading and math skills for students with and without symptoms of IIH; however, closer examination of overestimation and accuracy suggests that teachers are not equally accurate in their predictions. Additionally, though many studies examining teacher accuracy in predicting academic performance

suggest that they are accurate judges, it seems that the more specific the skills, teachers' judgments are less accurate.

Future Research

Future research should investigate the tendency of teachers to overestimate performance and examine accuracy of judgments of other educators. Perhaps surveying special education teachers or interventionists would reveal that they are better judges since they likely spend more time working individually with their students. Future studies could also assess teachers' accuracy on predicting CBM performance at baseline, provide in depth instruction about CBM performance, and then assess teachers' accuracy after the instruction to determine the role of familiarity with CBM on judgment accuracy. It is also important to investigate teacher accuracy with other areas of special need or perhaps children with formal diagnoses of ADHD.

Limitations

The study had several limitations. First, the study's student sample was relatively small and homogenous. The results demonstrated a trend that teachers were more accurate judges of students without symptoms of IIH; the small sample size may have limited the ability in some of the analyses to demonstrate that these differences were statistically significant.

Moreover, the study was focused on teacher judgments and only included 18 elementary teachers. Future research should include a larger number of teachers and extend the age range to higher grades, including middle and high school. In addition, the small teacher sample did not allow for any investigation of differences among teachers in their accuracy (e.g., gender, training). Additionally, variables related to school, teacher, or classroom characteristics could have influenced the results but was unavailable (e.g., school type, school size) or the sample was too small to investigate the potential impact on the results (e.g., teacher gender differences, teacher degree).

No direct information was collected on teachers' use or familiarity with CBM. Teacher familiarity with CBM assessment may influence their ability to judge their students' performance on these measures. For example, teachers who regularly use CBM may be better able to judge how far a child in their classroom would read in 1 minute. There were also some difficulties trying to capture teachers' judgments. For example, on reading CBM, teachers indicated how far they thought they would read in one minute. If they underestimated how far the child would read in 1 minute, no data were collected on any words after that point. Lastly, no global ratings were collected on writing performance since the ACES does not have a writing subscale.

It is important to note that although at times we found differences in teacher judgments that were significantly different, these differences in accuracy may not have as much practical significance. For example, when examining how accurate the teachers were in their judgment of students' scores on all reading probes combined, teachers were significantly more accurate for students in the Comparison Group (94% agreement) compared to the IIH Group (89% agreement). Although this finding was statistically significant, the difference may be minor in a practical setting. In an applied setting, a teacher would be able to correctly predict 47 out of 50 words for Comparison students versus 45 out of 50 for students with characteristics of ADHD. Although significantly different, this difference may have no real effect on classroom instruction.

Compliance with Ethical Standards

Conflict of Interest: Both authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

Amato, J. M., & Watkins, M. W. (2011). The predictive validity of CBM writing indices for eighth-grade students. *The Journal of Special Education*, *44*(4), 195-204. doi: 10.1177/0022466909333516

Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination o the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly, 23*(1), 43-55. doi: 10.1037/1045-3830.23.1.43

Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology*, *85*(2), 347. doi: 10.1037/0022-0663.85.2.347

Conners, C. K. (1989). *Manual for Conner's Rating Scales.* N. Tonawanda, NY: Multi-Health Systems.

Demaray, M. K., & Elliott, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly, 13,* 8–24. doi: 10.1037/h0088969

DiPerna, J. C., & Elliott, S. N. (2000). *Academic competence evaluation scales.* San Antonio, TX: The Psychological Corporation.

DuPaul, G. J., Anastopoulos, A. D., Power, T. J., Reid, R., Ikeda, M. J., & McGoey, K. E. (1998). Parent ratings of Attention-Deficit/Hyperactivity Disorder symptoms: Factor structure and normative data. *Journal of Psychopathology and Behavioral Assessment, 20,* 83-102.

DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (1998). *ADHD rating scale - IV.* New York: Guildford Press.

Eckert, T. L., Dunn, E. K., Codding, R. S., Begeny, J. C., & Kleinmann, A. E. (2006). Assessment of mathematics and reading performance: An examination of student performance and teacher report. *Psychology in the Schools, 43,* 247-265. doi: 10.1002/pits.20147

Eisenberg, D., & Schneider, H. (2007). Perceptions of academic skills of children diagnosed with ADHD. *Journal of Attention Disorders*, *10*(4), 390-397. doi: 10.1177/1087054706292105

Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly, 18*, 52– 65. doi: 10.1521/scpq.18.1.52.20876

Feinberg, A. B., & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research, 102*(6), 453-462. doi: 10.3200/JOER.102.6.453-462

Fuchs, L. S., & Fuchs, D. (2006). Implementing responsiveness-to-intervention to identify learning disabilities. *Perspectives on Dyslexia, 32,* 39–43.

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1998). *Monitoring Basic Skills Progress* (2nd Edition). Pro-Ed, Inc.

Green, R. W., Beszterczey, S. K., Katzenstein, T., Park, K., & Goring, J. (2002). Are students with ADHD more stressful to teach? Patterns of stress in an elementary school sample. *Journal of Emotional and Behavioural Disorders, 10,* 79–89. doi: 10.1177/10634266020100020201

Gresham, F. M., Reschly, D. J., & Carey, M. P. (1987). Teachers as "tests": Classification accuracy and concurrent validation in the identification of learning disabled children. *School Psychology Review, 16,* 543-553.

Hamilton, C., & Shinn, M. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgments of reading comprehension and oral reading skills. *School Psychology Review, 32,* 228–240.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*, 297–313. doi: 10.3102/00346543059003297

Hoover, H., Hieronymus, A., Frisbie, D., & Dunbar, S. (1993). *Iowa Test of Basic Skills.* Chicago: Riverside.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What is it and why do it? In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18–78). New York: Guilford Press.

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing: A literature review. *The Journal of Special Education*, *41*(2), 68-84.

Shinn, M. R. (Ed.) (1998). *Advanced applications of curriculum-based measurement.* New York: Guilford Press.

Shinn, M. R. (2002). Best practices in using curriculum-based measurement in a problem solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 671-697). Bethesda, MD: NASP.

Shinn, M. R., Tindal, G., & Spira, D. (1987). Special education referrals as an index of teacher tolerance: Are teachers imperfect tests? *Exceptional Children, 54*, 34-40.

Skiba, R., Magnusson, D., Marston, D., & Erickson, K. (1986). *The assessment of mathematics performance in special education: Achievement tests, proficiency tests, or formative evaluation?* Minneapolis, MN: Special Services, Minneapolis Public Schools.

Südkamp, A., Kaiser, J., & Möller, J. (2011). Accuracy of teachers' judgment of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743-762. doi: 10.1037/a0027627

Tindal, G., Germann, G., & Deno, S. L. (1983). *Descriptive research on the Pine County norms: A compilation of findings* (Research Report No. 132). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement* (Research Report No. 109). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Videen, J., Deno, S., & Marston, D. (1982). *Correct word sequences: A valid indicator of proficiency in written expression* (Rep. No. 84). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

Table 1

Demographic Characteristics of Participants by Group

	Total $N = 72$		Symptoms of IIH $n = 30$		Comparison $n = 42$	
	n	%	*n*	%	*n*	%
Gender						
Boy	35	48.6	14	46.7	21	50
Girl	37	51.4	16	53.3	21	50
Race/Ethnicity						
African American	2	2.8	1	3.3	1	2.4
Hispanic American	1	1.4	1	3.3	0	0
White	69	95.8	28	93.3	41	97.6
Grade						
Third	32	44.4	15	50	17	40.5
Fourth	9	12.5	5	16.7	4	9.5
Fifth	31	43.1	10	33.3	21	50.0

Note. IIH = inattention, impulsivity, and hyperactivity.

Table 2

Means and Standard Deviations of Main Study Variables by Group

Group	Variable	*N*	Minimum	Maximum	*M*	*SD*
Comparison	ADHD-IV Total	42	.00	16.20	5.10	4.15
	Student R-CBM	42	25.00	184.67	127.52	36.19
	Teacher R-CBM	42	30.00	331.00	198.96	72.69
	Student M-CBM	42	.33	13.00	3.73	2.53
	Teacher M-CBM	42	.00	24.67	15.29	6.66
	Student W-CBM	42	33.00	100.00	85.93	14.18
	Teacher W-CBM	42	10.00	100.00	80.07	19.32
	ACES Total	39	61.00	159.00	116.07	21.30
	ACES Reading	40	17.00	55.00	39.32	7.94
	ACES Math	40	14.00	40.00	28.10	5.88
Symptoms of IIH	ADHD-IV Total	30	16.00	49.00	32.50	8.28
	Student R-CBM	30	33.67	165.00	93.31	31.05
	Teacher R-CBM	30	39.00	298.00	148.24	63.49
	Student M-CBM	30	.33	6.33	3.12	1.77
	Teacher M-CBM	30	1.00	21.00	10.64	5.79
	Student W-CBM	29	31.00	100.00	68.33	16.64
	Teacher W-CBM	29	10.00	100.00	62.76	22.48
	ACES Total	29	38.00	131.00	85.86	23.01
	ACES Reading	29	15.00	44.00	28.55	8.01
	ACES Math	29	8.00	35.00	21.10	6.79

Note. IIH = inattention, impulsivity, and hyperactivity.

Table 3

Correlations among Actual CBM, Predicted CBM, and ACES subscales.

	Predicted R-CBM		ACES Reading		ACES Total	
	IIH	Comparison	IIH	Comparison	IIH	Comparison
Actual R-CBM	.33	.48***	.64***	.64***	.44**	.52***
	Predicted M-CBM		ACES Math		ACES Total	
	IIH	Comparison	IIH	Comparison	IIH	Comparison
Actual M-CBM	.35	.54***	.34	.33*	.15	.28
	Predicted W-CBM					
	IIH	Comparison				
Actual W-CBM	.24	.62***				

Note: * $p < .05$, * $p < .01$, *** $p < .001$.