

Syracuse University

SURFACE at Syracuse University

Theses - ALL

12-7-2022

Teachers' Perceptions of Students' Writing Skills: An Examination of the Agreement Between Teachers' Judgments and Curriculum-Based Measurement in Written Expression

Emily Ann Watts
Syracuse University

Follow this and additional works at: <https://surface.syr.edu/thesis>



Part of the [Education Commons](#), and the [Psychology Commons](#)

Recommended Citation

Watts, Emily Ann, "Teachers' Perceptions of Students' Writing Skills: An Examination of the Agreement Between Teachers' Judgments and Curriculum-Based Measurement in Written Expression" (2022).
Theses - ALL. 657.
<https://surface.syr.edu/thesis/657>

This Thesis is brought to you for free and open access by SURFACE at Syracuse University. It has been accepted for inclusion in Theses - ALL by an authorized administrator of SURFACE at Syracuse University. For more information, please contact surface@syr.edu.

Abstract

Nearly 75% of students in the United States of America are not meeting grade-level standards in the area of writing (NCES, 2012; Persky et al., 2003), despite this skill impacting students' performances in other academic areas (Ray et al., 2016), and limiting students' access to higher education (Addison & McGee, 2010), and opportunities for jobs in the adult workforce (National Commission on Writing, 2005). Because difficulties with early writing skills are associated with later writing skills deficits (Juel, 1988), it is crucial that educators accurately identify students in need of additional support in order to provide them with appropriate instruction. Two common methods for identifying students are through teacher referral and standardized assessments such as Curriculum-Based Measurement-Written Expression (CBM-WE). Although levels of agreement between teacher referral and CBM-WE were examined in the past, this study extended the literature by conducting kappa analyses to investigate levels of agreement in order to take chance into account. In addition, due to differences in students' performances on national assessments based on gender (Reilly et al., 2019), as well as differences between national and state normative data, levels of agreement were investigated as a function of gender and normative type. Results of this study suggest that levels of agreement may vary based on the CBM-WE scoring metric used and the student's gender, but no evidence was found to suggest using national or local norms impacted agreement. Furthermore, this study revealed poor levels of agreement for female students with writing skills below the 10th percentile across, suggesting that schools may need to use converging data sources to accurately identify female students in need of support. Limitations, directions for future research, and further implications are also discussed.

Teachers' Perceptions of Students' Writing Skills: An Examination of the Agreement Between
Teachers' Judgments and Curriculum-Based Measurement in Written Expression

By

Emily Watts

B.A., University at Buffalo, 2005
M.S., Rochester Institute of Technology, 2010

Thesis

Submitted in partial fulfillment of the requirements for the degree of
Master of Science in Psychology

Syracuse University
December 2022

Copyright © Emily Ann Watts 2022

All Rights Reserved

Table of Contents

Introduction.....	1
Prevalence of Writing Difficulties Displayed by Students.....	2
Theoretical Framework of Writing Development.....	3
Problem Identification Methods.....	7
Gender Disparities in CBM-WE Performances.....	12
Agreement Between Teacher Judgments and Academics.....	19
Purpose of Study.....	25
Method.....	27
Participants.....	27
Measures.....	28
Outcome Variables.....	30
Procedures.....	31
Results.....	33
Descriptive Results.....	34
Major Analyses.....	36
Discussion.....	44
Agreements of Teachers' Rating and Students' Writing Skills.....	45
Relationship Between Teachers' Ratings and Students' Writing Skills.....	46
Gender-based Analyses.....	46
Limitations.....	49
Implications for Future Research and Practice.....	50
Conclusions.....	52

Teachers' Perceptions of Students' Writing Skills: An Examination of the Agreement Between Teachers' Judgments and Curriculum-Based Measurement in Written Expression

Writing is an important skill that is required to communicate with others and record one's experiences, discoveries, and thoughts (Persky et al., 2003). To be successful with written communication, it is imperative that school-aged students develop proficient writing skills (Addison & McGee, 2010; National Commission on Writing, 2004). Written language is used to communicate through notes, essays, emails, PowerPoint presentations, and other mediums and is used in one's life across time and settings. Therefore, difficulties in writing can impact a student's life in many ways.

It is crucial for students to develop writing skills early on, as difficulties with early writing skills are associated with later writing skills deficits (Juel, 1988). Not only might difficulties in writing impact students' skills to meet standards in English Language Arts (ELA) and put them at risk for developing future writing skills (Juel, 1988), but these difficulties may also impact performance across content areas. In a survey of 159 middle school teachers in the areas of language arts, science, and social studies, 99.4% of teachers reported requiring students to write short response answers to assess their knowledge in the subject (Ray et al., 2016). These survey results also indicated that 96.3% of the participating teachers required students to take notes during class, and 95.7% of the teachers required students to complete worksheets.

The need for proficient writing skills may continue into adulthood through either higher education or the adult workforce. In a survey of 554 high school teachers and professors in higher education, over 96% of respondents indicated that writing will be somewhat important to very important for students' future successes (Addison & McGee, 2010). Writing skills are also needed for gainful employment. In a survey of state human resource directors, over 90% of

respondents reported that an applicant's writing skills are almost always considered when hiring employees (National Commission on Writing, 2005). Similarly, in a survey of 120 major corporations in the United States of America, results indicated that over two-thirds of salaried workers in those corporations have writing responsibilities and that inadequate writing skills are often a barrier to promotions (National Commission on Writing, 2004). Clearly, students need to develop proficient writing skills to meet the demands of primary and secondary education, higher education, and the adult workforce.

Given the importance of writing and the need to develop these skills early, Common Core State Standards now have an emphasis on developing students writing skills as early as kindergarten (Shanahan, 2015). These standards for student performance are currently being implemented in 41 states, the District of Columbia, and four territories nationwide (Common Core State Standards Initiative, 2018). Clearly, writing skills are an important component of academic success.

Prevalence of Writing Difficulties Displayed by Students

Despite the need to develop proficient writing skills, a majority of students nationwide do not meet grade-level competency standards in the area of writing. In 2011, the National Center for Education Statistics (NCES) examined the writing performance of a large, nationally representative sample of eighth- and twelfth-grade students through the National Assessment of Educational Progress (NAEP). The students were instructed to write with the use of a computer for one of three essay purposes: to persuade, to explain, or to convey an experience. Students were provided with standard technological tools, such as spellcheck, copy and paste, and a thesaurus. The results of this assessment revealed that nearly three-fourths of eighth-grade students performed below grade-level proficiency standards. This trend was also observed in

high school, as 73% of twelfth-grade students also performed below grade-level proficiency standards (NCES, 2012). These results indicate that most students graduating high school in the United States of America are demonstrating only partial mastery of the prerequisite knowledge and skills required to communicate their thoughts and knowledge through written language (NCES), thus impacting their abilities to be successful in higher education or the adult workforce (Addison & McGee, 2010; National Commission on Writing, 2004).

Students in elementary school also demonstrate difficulties in the area of writing. In an earlier NAEP assessment conducted by the NCES in 2002 that included elementary-aged students, only 28% of fourth-grade students met grade-level proficiency standards in the area of writing (Persky et al., 2003). This indicates that only about a quarter of fourth-grade students supported and developed their main ideas in a manner that showed a clear understanding of the assigned writing task and the audience they were expected to address (Persky et al., 2003). Given the large number of students not meeting grade-level proficiency standards, it is crucial that educators base their practices on empirically-supported theoretical models of writing development.

Theoretical Framework of Writing Development

Writing is a complex behavior that is developed by using separate and interacting systems of language (Berninger, 2000). Writing requires many skills, such as spelling, handwriting, generating ideas for writing, organizing thoughts, and creating text based on the thoughts while maintaining focus and attention to the task (Berninger & Winn, 2006). Given this complexity, dozens of empirical studies over the past four decades have studied how writing develops. From this body of empirical work, prominent theories of writing development have emerged.

Model of Adult Writing

Initial theoretical work on writing development was conducted with adults. Hayes and Flower (1980) proposed that writing is goal-directed and highlighted the cognitive processes used in the writing process by adults: planning, translation, and revision. Planning is a pre-writing process in which the writer plans what to write by brainstorming, retrieving information from their long-term memory, and generating ideas. Individuals then translate their thoughts and ideas into written language. Finally, writers engage in a revision process in which they evaluate and review what they have written in order to make appropriate changes. These processes are complex and interact with one another (Hayes & Flower, 1980). Writers may not follow these in successive order, but rather use them more fluidly while going back and forth between each process throughout their writing (Hayes & Flower, 1986). For example, a writer might be prompted while revising to plan for further writing to more accurately accomplish their writing goals. Although this framework aimed to explain the process of writing development, this theory focused on the cognitive processes in adults rather than addressing how children develop writing behaviors, such as transcription and text generation, which prompted further investigation into the process of writing in children.

Model of Children's Writing Development

Based on Hayes and Flower's work, Berninger and colleagues conducted a series of rigorous empirical studies to systematically investigate the development of children's writing. Results of these studies led Berninger (2000) to theorize that, first, children develop an aural language system to understand the language that they hear, thereby developing receptive language skills. Then, children develop an oral system for expressive skills to communicate orally. Next, children begin to develop a language system for written language. In order to do

this, they develop a system to process written language for reading. Then, children develop a language system for producing written language. These four language systems are both distinct and interacting systems, leading to the complexity of writing development (Berninger, 2000).

Berninger (2000) initially proposed that there are two components of the writing process that are particularly relevant to children's writing development—transcription and text generation. Later, Berninger and Winn (2006) expanded this model to also include executive function and memory. This model proposes that these systems are interrelated and function together to support writing (Abbott & Berninger, 1993). In order for children to move from an oral language system to a written language system, all four components are needed to work in concert (Berninger, 2000).

Transcription is the first component of the writing system to develop (Berninger & Winn, 2006). First, a child learns how to write specific letters (i.e., handwriting), and then learns how to form words with letters (i.e., spelling) correctly. These are separate skills but are correlated factors (Abbott & Berninger, 1993). In addition, even though some children may demonstrate difficulties with handwriting and spelling together, other children demonstrate difficulties with either spelling only or handwriting only (Berninger et al., 1998). As demonstrated by Kim et al. (2011), oral language, spelling, and letter writing fluency were found to be uniquely and positively related to writing skills in kindergarten students. Both handwriting and spelling are skills required to translate oral language into written language.

However, simply having the skills to write and spell words correctly is not sufficient to produce written language. Once words can be formed, children have to construct text through text generation (Berninger, 2000). Text generation has multiple dimensions—fluency, discourse structures, and quality (Berninger, 2000). Fluency is how quickly and efficiently a student can

write and is often measured by how many letters or words are produced in a brief writing period (Berninger, 2000). Discourse structure is the way in which a text is organized, both locally, at the sentence level, and globally, at the overall topic level (Berninger, 2000). Quality of writing may be difficult to define, is often assessed using holistic ratings, and can vary depending on the age of the writer, topic, and reader audience (Berninger, 2000).

Cognitive processes involved in the writing process include executive functioning and memory (Berninger & Winn, 2006). Executive functioning is needed for planning, reviewing of text, and text revision (Hayes & Flower, 1980), and it provides the self-regulation needed for task initiation, task maintenance, and task completion (Harris & Graham, 1992). Short-term memory is needed for word recognition and transcription (Swanson & Berninger, 1996), while working memory is used when generating text (Swanson & Berninger, 1996). Long-term memory is also needed in the writing process, as this provides the writer with access to previously learned knowledge of writing topics and letter/word transcription (McCutchen, 1986). Although these processes—transcription, text generation, executive functioning, and memory—work separately and in unison to support the complex writing process (Berninger, 2000), if transcription skills (i.e., spelling and handwriting) are underdeveloped, executive functioning and memory will be devoted to transcribing text instead of generating text (Kim et al., 2015). When children lack accuracy and fluency in their transcription skills, their memory becomes overtaxed (McCutchen, 2000). This limits cognitive resources that can be used for generating text. Therefore, writing measures that assess accuracy and fluency with transcription skills may help to identify those students in need of additional support in the area of writing.

Problem Identification Methods

Given the complexity of how written language develops, it may not be surprising that a majority of students' writing skills in the United States of America do not meet grade-level proficiency standards (Persky et al., 2003). Thus, it is crucial that educators aim to correctly identify students at risk for not developing proficiency with their writing skills in order to provide them with appropriate instruction and intervention. One model that may provide a framework for this is Response to Intervention (RTI).

RTI is a problem-solving process used to evaluate instructional approaches and ensure that all students are provided with and responding to high-quality instruction (Shinn, 2005). RTI uses a three-tiered approach to provide evidence-based general education to all students (Tier I) and to provide high-quality intervention and remediation services to support students identified as needing additional academic support (Tiers II and III). A core feature of RTI is the use of data to guide instructional decision-making. Within an RTI model, universal screening of students' basic academic skills is used to identify those students in need of additional academic support (Deno, 2015).

One tool used to screen students' academic skills is curriculum-based measurement (CBM). CBMs are an indicator of basic skill development in specific skill areas. CBMs are defined as measures involving "direct observation and recording of student performance in response to selected curriculum materials [that] are emphasized as a basis for collecting information" (Deno, 2003, p. 4). Used as a universal screening tool, CBMs allow educators to assess the academic skills of the entire student population in a setting to identify students at risk for academic difficulties.

There are several advantages to using CBMs in schools. CBMs can be used for problem identification in multiple academic areas, such as reading (CBM-R), mathematics (CBM-M), spelling (CBM-S), and written expression (CBM-WE). CBMs are brief production measures (typically fluency-based), making them sensitive to small changes over time (Espin et al., 2000). This sensitivity allows for the use of CBM to monitor students' progress in response to interventions.

Perhaps due to these advantages, the popularity of CBMs and their use in schools is increasing. In a recent survey of 3,218 general and special education teachers in four Midwest states, 76.1% of respondents reported using CBM for screening purposes (Swain & Hagaman, 2020). This has significantly increased since the same survey was conducted in 1997, when only 45% of respondents indicated that their school utilized CBMs (Swain & Hagaman, 2020). Several software programs have been developed to assist schools in providing academic assessments with CBMs. For example, AIMSWeb® (2018) is one web-based program that supports schools in providing universal screening, progress monitoring, and data management tools for students in grades K-12, including CBM-WE to screen and monitor students' writing performances.

CBM-WE

CBM-WE can be used to screen students' written expression skills. When using CBM-WE, students are provided with a writing prompt, are given time to think, and then are asked to produce a writing sample. Several studies have examined the various aspects of CBM-WE procedures, such as writing prompt selection, writing duration, number of prompts, and scoring procedures (Jewell & Malecki, 2005; McMaster et al., 2009; Romig et al., 2017; Romig et al., 2020).

CBM-WE Administration. Prior to administering CBM-WE to students to assess their writing performance, educators must first decide on the type of prompt provided to students. AIMSWeb® (2018) CBM-WE probes use story stems to prompt students' writing. A recent meta-analysis (Romig et al., 2020) examined the criterion validity of various types of writing prompts (i.e., story stem, picture, expository, text copying, picture-word, and picture-story). Across all writing prompts, results indicated that CBM-WE outcomes were moderately correlated ($r = .48$ to $.67$) with state- or commercially-developed writing assessments. Although the authors were not able to examine whether there were statistically significant differences in criterion validity between prompt types due to a low number of studies, the confidence intervals of all the prompts had significant overlap, indicating minimal differences in criterion validity between writing prompts. This suggests that unless future work suggests otherwise, researchers and educators can select prompts based on assessment goals rather than based on a concern of differences in the prompts' criterion validity.

Once educators have decided on a prompt type, they then have to decide the duration in which students will write. AIMSWeb® (2018) uses 3-minute writing durations to assess students' writing skills. In their meta-analysis, Romig et al. (2020) also examined the criterion validity of various writing durations (i.e., 1.5, 3, 5, 6, 7, and 10 minutes). Across all writing durations, results indicated that CBM-WE outcomes were moderately correlated ($r = .48$ to $.64$) with state- or commercially-developed writing assessments. Similar to prompt type, the authors were unable to examine whether there were statistically significant differences in correlation magnitudes between writing durations; however, again, the confidence intervals of all of the durations had significant overlap, indicating a minimal difference in criterion validity between writing durations. Similar findings were evidenced when the authors examined writing duration

by grade level (i.e., K-5 and 6-12) and obtained similar correlation coefficients between the results of CBM-WE and state- or commercially-developed writing assessments ($r = .53$ to $.67$). These results suggest that regardless of grade level, writing durations ranging from 1.5 to 10 minutes will likely yield similar levels of criterion validity.

Within RTI models their writing performances can be compared to normative data to identify their level of writing risk: (a) below the 10th percentile, (b) within the 10th to 24th percentile, (c) within the 25th to 75th percentile, (d) within the 76th to 90th percentile, and (e) above the 90th percentile (Shinn, 2005). Typically, students whose performance falls below the 25th percentile are considered to be at risk for academic difficulty and in need of intervention support (Shinn, 2005). A common procedure for categorizing students' academic performances is through the use of national norms, and AIMSWeb® (2018) provides national norms for TWW, WSC, and CWS. Alternatively, schools can use local norms (i.e., comparing students to other students in their classroom or school instead of the national population) as a basis for comparison (Shinn, 2005).

Previous research examining the diagnostic accuracy of national and local norms in the area of reading suggests that the use of local norms has higher predictive validity for second-through fifth-grade students' performances on state-developed assessments (Sandberg Patton et al., 2014). Not only might the use of local norms have higher predictive validity for state-developed assessments in the area of reading, but local norms may also better align with teacher and parent perceptions of students' academic skills than national norms. Teachers are more likely to agree with students' scores on academic outcome measures when they are asked to compare their students against a sample of students in the classroom or in the district than if they are asked to judge their students' performances on academic measures against a national sample of

students. Mowry and Farran (2016) examined the extent to which fifth- and sixth-grade students' scores on a math achievement test were related to teachers' and parents' perceptions. Results indicated that when students' scores (i.e., within the sample and therefore using local norming) were grouped into five categories (i.e., far below average, below average, average, above average, and far above average) based on teachers' judgments, there were significant differences between the groups' scores on the math assessment, indicating an alignment between teachers' judgments and the sample performance. However, the teachers' judgments did not match the national norms provided by the publishers of the math assessment. In addition, it has been suggested that the use of local norms may be a more accurate method to assess individual students, whereas the use of national norms for comparison may be more reflective of systems-level problems (Habedank-Stewart & Kaminski, 2002). For example, if when using national norms as a comparison group, educators realize that a majority of their students' scores fall below average, this may indicate a need to change the instructional plan for all students as opposed to identifying some students in need of additional academic support. Not only does the source of comparison affect what information students' scores on CBM-WE imply, but the type of scoring does as well.

Scoring Indices. Once CBM-WE is administered and a writing sample is obtained, it can be scored using a variety of metrics. Scoring within CBM-WE typically includes the following metrics: total words written (TWW), words spelled correctly (WSC), correct writing sequences (CWS; i.e., the number of word-word and word-punctuation sequences that have correct spelling, capitalization, punctuation, grammar, and syntax; Jewell & Malecki, 2005), and correct minus incorrect writing sequences (CIWS). Notably, these metrics all measure the writing fluency component of text generation as theorized by Berninger and Winn (2006); however, there are

subtle differences between each metric. For example, TWW is assessed by counting the total number of words a student writes regardless of accuracy, and thereby focuses exclusively on writing productivity. WSC is measured by the number of words a student correctly spells, thereby focusing on transcription accuracy within the context of writing productivity. CWS, calculated by adding the number of word-word and word-punctuation sequences a student writes, focuses on accurate writing productivity regardless of the number of errors in a writing sample., whereas CIWS takes the number of writing errors into account.

A recent meta-analysis (Romig et al., 2017) examined the criterion validity of TWW, WSC, CWS, and CIWS. Results indicated that all of the scoring indices yielded moderate correlations with commercially-developed tests or state- and district-developed achievement tools ($r = .37$ to $.60$), with overall correlations of $.37$ for TWW, $.44$ for WSC, and $.51$ for CWS. CIWS had an overall correlation of $.60$. Although a major limitation of this study was that the authors only used the highest correlation reported by each study, CWS and CIWS had the highest criterion validity estimates.

Gender Disparities in CBM-WE Performances

Previous research has identified discrepancies in students' CBM-WE performances as a function of gender (Farrington et al., 2014; Jewell & Malecki, 2005; Malecki & Jewell, 2003), which aligns with national writing assessment results. NAEP results indicated that 82% of eighth-grade males did not meet grade-level proficiency standards, whereas 64% of eighth-grade females did not meet grade-level proficiency standards. These results suggest that females tend to outperform their male counterparts in the area of writing (NCES, 2012).

Farrington et al. (2014) administered CBM-WE probes to students in third- through twelfth-grade ($n = 1,240$) at three time-points throughout a single school year. Across all grades,

and at all time points, females wrote more words and produced more CWS when compared to their male counterparts. These results were similar to findings reported by Weiss et al. (2019), in which the authors found that females wrote more words, spelled more words correctly, and had a higher rate of CWS. Furthermore, the authors found that females tended to write longer words, have more variety of words, and have more correct punctuation marks within their writing samples. These gender disparities were observed even after controlling for other factors. In a study with 492 second-grade students, being male was found to be a significant, negative predictor of writing performance, even after controlling for reading and oral language skills, spelling skills, handwriting fluency, and letter-naming fluency (Kim et al., 2014). There is also evidence that males and females generalize writing gains developed through intervention differently. In a study examining the generalization effects of performance feedback and goal setting interventions, Hier et al. (2019) found that females demonstrated higher CWS on measures of generalization after receiving the intervention, even when controlling for baseline writing performance.

Although females may outperform males in the area of writing, as several studies have indicated (Farrington et al., 2014; Jewell & Malecki, 2005; Malecki & Jewell, 2003), there is conflicting evidence regarding writing growth rates between genders. In one study with 89 second- through fifth-grade students, females developed their writing skills at a different pace than males, as females' linear growth for CWS and CIWS on writing tasks exceeded males' growth (McMaster et al., 2017). However, in a previous study with 672 second- through fifth-grade students conducted by Keller-Margulis et al. (2015), the results indicated that although males' writing performances on CBM-WE were initially lower than females' writing performances, there were few differences between genders in the rate of growth across the

school year. Although males across all grades demonstrated greater growth rates on %CWS, the only significant difference in CWS growth rates between genders occurred in third grade, where females demonstrated greater CWS growth than males. For other scoring metrics (TWW and WSC), growth rates did not differ across grades or between genders. Therefore, it is possible that gender differences in writing performances might be dependent upon the type of writing produced or how writing is assessed.

Gender Disparities on National Assessments. Given that gender discrepancies have been inconsistently found in studies examining student writing performance on CBM-WE (Farrington et al., 2014; Jewell & Malecki, 2005; Keller-Margulis et al., 2015; Malecki & Jewell, 2003; McMaster et al., 2017), it is important to consider how these gender differences may or may not be occurring on other assessment methods. Using the data collected by the National Assessment of Educational Progress (NAEP), a large, nationally representative longitudinal assessment of fourth-, eighth-, and twelfth-grade students, Reilly et al. (2019) examined students' writing performances from 1988- 2011 ($N = 3.9$ million students) to evaluate the magnitude of gender discrepancies. The authors found there to be an overall medium effect size ($d = .54$) when examining gender differences in students' performances on the NAEP. There were no significant differences between grade levels. Furthermore, girls were more likely to obtain proficient scores by twelfth grade (63.3%) than boys (36.7%). This significant trend over the past three decades of females outperforming their male counterparts is remarkable given the inconsistent findings in studies examining gender discrepancies on CBM-WE.

The ambiguities in research studies and national assessments examining possible gender disparities in students' performances on writing assessments justify further examination. One possible explanation for these discrepancies may be related to how the data are collected for

CBM-WE research compared to national assessments. Research studies examining students' writing skills on CBM-WE measures are sample-based (instead of population-based) and therefore represent results from samples in a particular region and at a particular time. Therefore, there may be more variability from study to study and may not be reflective of the student population in this nation. Overall, although previous literature suggests there is a need to consider students' gender in writing assessment research, it should be noted that gender discrepancies in CBM-WE student performances have been inconsistent and need to be further explored.

It is also possible that the type of scoring metric used may contribute to gender discrepancies in students' writing performances on CBM-WE, as well as why the gender differences are consistently found on national assessments (Reilly et al., 2019). In an investigation of 946 first- through eighth-grade students' writing samples, females wrote significantly more TWW, WSC, CWS, and CIWS (Malecki & Jewell, 2003). The authors concluded that females tend to demonstrate stronger writing skills regardless of what scoring indices are used. However, all of the scoring indices in that study involved production. Upon further investigation of these gender differences, Jewell and Malecki (2005) found that although females performed higher on all production-dependent metrics, there were no gender differences when examining production-independent or accurate-production metrics. Thus, even though females may write more TWW and WSC, males and females may be equally accurate in their writing (i.e., CIWS, %CWS). The type of scoring metric used to assess students' CBM-WE samples may influence whether gender differences are identified, and similarly, why these differences are consistently found on national assessments as they do not use these different scoring metrics but rather teachers use a holistic scoring rubric. Therefore, it is important to

consider to the agreement between educators' perceptions of students' writing performance in relation to results purported by CBM-WE measures when evaluating students' writing samples.

Teacher Judgments

Although the use of CBMs for instructional decision-making is on the rise in several states (Swain & Hagaman, 2020), it is unclear to what extent those data are supplemented with teacher judgments about students' writing skills and how much of a role that plays in instructional decision-making. Historically, teachers' judgments of students' performances have been one of the primary methods used to identify students in need of special education services (Ysseldyke et al., 1982). Teacher judgments and ratings continue to be used in the evaluation of students' academic skills (Ritchey & Coker, 2014). Consequently, teachers are expected to be proficient in identifying problems within students' academic performances (Wagner et al., 2017). Teacher characteristics, such as self-reported levels of efficacy in student engagement and instructional strategies, influence the likelihood they will refer students for additional intervention support (Randall, 2013). Therefore, teachers' judgments of student skills can play a role in how and whether a student receives additional academic support.

Teacher Judgment in Writing. Despite the importance of teacher judgments, there is no standard agreement among teachers on what constitutes good writing. Elementary teachers' judgments of students' writing skills are strongly influenced by students' use of writing conventions (i.e., spelling, capitalization, punctuation) and writing organization (i.e., whether it is coherent and logical; McFarland & Wallace, 2008), but these judgments often lack reliability (Swartz et al., 1999), suggesting that teachers may have diverse, subjective, and complex opinions about students' writing skills.

How a teacher assesses a student's writing skills may be different from how CBM-WE is used by educators to assess a student's writing skills. CBM-WE is a general outcome measure used to evaluate writing fluency, but research indicates that teachers may be considering other factors than just writing fluency when judging students' writing skills. Olinghouse and Leaird (2008) examined how written vocabulary was considered in teachers' judgments of fourth-grade students' writing skills on narrative writing samples. The authors found that vocabulary diversity within the students' writing samples explained 20.6% of the variance in teacher judgments, but TWW (i.e., measure of writing fluency) only explained 8.5% of the variance in teacher judgments. In another study examining teachers' judgments of students' analytic writing skills, Wang et al. (2018) found that 98% of the sample's fifth-grade teachers considered whether students accurately addressed the writing prompt, 95% of teachers considered whether students used evidence to support their claims, and 93% considered whether the student writing sample provided an adequate explanation. Although CBM-WEs administered in winter and spring of academic schoolyears have been demonstrated to be significantly correlated with other assessments of writing, such as state English Language Arts exams ($r = .29, .34$; Coddington et al., 2015), teachers may be considering additional factors of writing, other than fluency, when assessing students' writing skills, such as vocabulary, voice, and organization.

Gender Discrepancies in Teacher Judgments. Despite evidence that students' CBM-WE performances may differ as a function of gender (Fearington et al., 2014; Jewell & Malecki, 2005; Malecki & Jewell, 2003), less research has examined differences in teachers' judgments of male and female writing. This is particularly relevant given the decades-long trend of gender discrepancies on NAEP assessments (Reilly et al., 2019), which are scored in a holistic manner by educators. In addition, gender stereotypes might play a role in how a teacher judges a

student's writing skills. In a review of gender stereotypes, Ellemers (2018) concluded that gender stereotypes impact teachers' judgments in a variety of subject areas. Thus, it is possible, given the lack of direct evidence for writing specifically, that teachers may have unconscious beliefs about how a male or female writes.

One study (Beard & Burrell, 2010) that did examine whether teachers' judgments of students' writing skills varied as a function of gender found conflicting evidence. The authors examined teachers' holistic judgments of students' writing skills among 112 students aged 9-11 years old in London. Teachers were asked to rate students' narrative and persuasive writing samples on the following factors: overall writing skills, writing fluency, effective opening, content, and language use. Results suggested that the teachers judged females' overall writing skills to be higher than males. Furthermore, teachers rated females' performances on narrative essay types and overall writing fluency as higher than males. However, on persuasive essays, teachers' judgments of males' writing skills in the areas of effective opening, content, and language use were higher compared to females' writing performances.

Stagg Peterson and Kennedy (2006) examined whether students' gender played a role in teachers' responses to student essays. The authors took writing samples from 22 sixth-grade students and asked 11 classroom teachers to guess the gender of the student authors. Based on these results, the researchers picked the two student authors with the greatest level of uncertainty from teachers regarding the ascribed gender of the student writer. With writing samples from these two students (i.e., one male and one female), Stagg Peterson and Kennedy manipulated one essay written by the male student and presented it to half of the teachers with a female student name and to half of the teachers with a male student name. The same procedure was done for an essay written by a female student. Then, the teachers' comments were coded into seven

categories (correction and criticism, commands, closed questions, praise, open-ended questions, reader response, and lessons or suggestion). Results of a two-way ANOVA indicated that when the teachers were given an essay with a name typically assigned to males, teachers tended to provide more corrections and criticism than when there was a typically female-assigned name, $F(1,96) = 4.25, p < .05$. Teachers also tended to give more explanations, suggestions, and lessons when the student was perceived to be a male as compared to when the student was perceived as a female, $F(1,102) = 7.92, p < .05$. Based on these results, it appears that not only might a student's gender impact how a teacher judges their writing, but also the type of feedback that a student receives. More research needs to be conducted to further examine possible differences in how teachers judge students' writing skills across genders.

Agreement Between Teacher Judgments and Academics

Because correct identification of the problem impacts every stage of the problem-solving process and given teachers' crucial roles in problem identification (Eckert & Arbolino, 2005), an examination of whether teachers and CBMs identify the same students in need of additional academic support is necessary. Previous research suggests results of academic assessments and teachers' judgments of students' academic skills have a high overall agreement. A meta-analysis of 75 studies (Südkamp et al., 2012) was conducted to assess the agreement between teachers' judgments of student academic achievement and the results of students' performances on measures of academic achievement. Results indicated that the relationship between teachers' judgments and students' academic performances was high ($r = .63$). Upon further investigation, the analyses indicated that significantly higher effect sizes were obtained in studies in which the teachers were informed about the specific academic test being used and in which way it assessed the students' skills when compared to studies in which the teachers were not informed about the

specific academic test utilized ($Zr = .76$ and $.61$, respectively). Based on those results, it is likely that when teachers are informed about what specific skills measures assess, they are more likely to agree with the results of academic assessments. Given that no teachers reported using CBM to assess writing-related skills in a recent survey (Swain & Hagaman, 2020), it may be possible that teachers are less familiar with CBMs related to writing, leading to higher disagreements with measures of written expression. One major limitation of the meta-analysis conducted by Südkamp et al. was that the authors broadly examined the relationship between teachers' judgments and students' academic performances using correlational analyses. Thus, this study does not provide a precise indication of agreement, but rather examined to what extent the two methods were related.

Agreement Between Teacher Judgments and CBM

In one study examining the level of agreement, the authors assessed the agreement between teacher judgments and second-grade students' performances on CBM-M and CBM-R (Eckert et al., 2006). Based on established guidelines (Fuchs & Deno, 1986), each student's instructional levels in math and reading were determined. Teachers were asked to estimate each student's general academic skills and instructional level, and to compare each student's skills to classroom norms. These authors examined both the correlational relationship and percent agreement between teacher judgment and CBM results. In the area of mathematics, the correlation between teachers' judgments and CBM-R results were negligible to low (median $r = .16$). When examining the total percent agreement between teachers' judgments and CBM-M, although agreement for single-digit addition was high across all instructional levels, the total percent agreement was low when CBM-M assessed other math skills such as subtraction or advanced addition for students in the frustrational or mastery levels for these skills. Furthermore,

in the area of reading, teachers' judgments and CBM-R performances were moderately to highly correlated (median $r = .72$). However, upon further investigation using total percent agreement, teachers' judgments and CBM-R results were higher when assessing students' skills on grade-level or above grade-level materials, but lower when assessing below grade-level materials (i.e., first-grade material). The authors suggested that these differing levels of agreement between grade-level materials may have occurred because the teachers were predicting students to be at the instructional level for second-grade material because they were second-grade students, rather than accurately assessing the students' reading levels.

Similar findings were obtained when Begeny et al. (2008) examined how teachers' judgments agreed with first- through third-grade students' CBM-R performance. They identified similar discrepancies between the agreement of teacher judgments and CBM-R results for students with high oral reading skills and students with low oral reading skills, but the authors did not find any significant differences in agreement level based on the grade of the students. Specifically, first- through third-grade teachers had higher rates of agreement with CBM results when rating students with strong reading skills and lower rates of agreement when assessing oral reading skills of students with weaker reading skills (Begeny et al., 2008). Although the authors asked teachers to judge students' writing skills relative to other students in the class, they utilized national norms established in previous normative studies. Teachers may have rated their students differently if they had been informed that students' performances would be compared to how students across the nation perform in reading. In addition, it is possible that teachers' lack of familiarity with national norms may have impacted the results.

Although both studies (Begeny et al., 2008; Eckert et al., 2006) contributed to the literature regarding how teacher judgments and CBM are related, they did not examine how

agreement is related to the area of writing. In addition, both studies took the position that CBM measures are the accurate estimates of students' skills in the areas of reading and mathematics. This may be appropriate given the binary nature of reading and mathematics; either a student reads a word and answers a math problem correctly or they do not. However, in the area of writing, students' skills may be more difficult to judge. CBM-WE is a measure of writing fluency and does not take factors such as writing organization, persuasiveness, or overall writing quality into account, whereas teachers' judgments of students' writing samples may consider these as well as additional factors when assessing their students' writing samples. For example, teachers may consider vocabulary diversity (Olinghouse & Leaird, 2008), providing development of ideas (Wang et al., 2018), and using evidence to support claims (Wang et al., 2018). Therefore, when considering to what extent CBM-WE and teacher judgments identify the same students in need of additional academic support, the level of agreement should be emphasized.

Agreement Between Teacher Judgments and CBM-WE

Initial work examining the relationship between teachers' judgments of students' writing skills, and CBM-WE results was conducted by Videen et al. (1982). The authors examined the correlation between CWS on students' writing samples with teachers' holistic ratings of students' writing skills and found these two factors to be significantly related ($r = .85$). Espin et al. (2000) extended this work by examining the predictive validity of CBM-WE with teachers' ratings of students' writing skills. Results indicated significant, moderate relationships between teachers' judgments with TWW, CWS, and CIWS ($r = .46$ to $.66$). However, in both of these studies, the authors examined to what extent teachers' judgments and students' writing performances were related rather than to what extent they both identified the same students in need of support.

In a study conducted in 2002, Gansle et al. examined the relationship between teachers' rankings and students' CBM-WE results. Teachers involved in this study were asked to rank their students' writing skills from first to last within each classroom. The results of this study did not find a significant relationship between teachers' rankings and TWW, but did find a significant, yet small, relationship between teachers' rankings and WSC. The results of Gansle et al. (2002) suggest that level of agreement between CBM-WE results and teachers' judgments may vary depending on the scoring metric used. McFarland and Wallace (2008) examined which CBM-WE scoring metric would best predict teachers' ratings and obtained similar results in which WSC significantly predicted teachers' judgments of students' writing skills.

In a study with 170 second- and third-grade students, Ritchey and Coker (2013) examined the relationship between teachers' judgments of students' overall writing skills with CBM-WE. All of the scoring metrics (TWW, WSC, and CWS) on the story-prompt type CBM-WE were significantly, moderately correlated ($r = .30$ to $.48$) with teachers' judgments of students' writing skills. Furthermore, the authors examined to what extent the two methods (i.e., teachers' judgments and CBM-WE results) identified students at risk for writing difficulties and found that false negatives increased when teachers' ratings were used to identify students with writing difficulties. However, this study used national norms to categorize students' writing performances, which may have impacted teachers' abilities to accurately judge students' skills.

Taken together, prior research suggests that although there are psychometric benefits of CBMs (Romig et al., 2017; Romig et al., 2020), there may not be a strong agreement between those direct measures of students' performances and teachers' judgments of students' academic skills. This is problematic because a lack of agreement between a teacher's judgments and the assessment measures used for universal screening can make correctly identifying students in

need of additional support challenging and potentially negatively impact student outcomes (Eckert & Arbolino, 2005). Although previous research has examined the agreement between CBMs and teacher judgments in the area of writing, studies often examined only the relationship between the two methods rather than the agreement level, and they did not consider how the agreement level may have varied due to the type of norm used. Furthermore, given the gender discrepancies commonly identified in assessments of students' writing performance, there is a need to investigate to what extent teacher judgments and CBM-WE level of agreement vary as a function of gender. Finally, it is possible that teachers' judgments of students' academic skills are formed based on the typical writing performances of students in that school, thus indicating a need to examine to what extent levels of agreements between teacher judgments and CBM-WE vary when local or national norms are used.

Use of Kappa Statistic to Examine Agreement

Previous literature examining teacher judgments and CBMs have often used two methods—correlation and percent agreement—to assess the level of agreement (Begeny et al., Eckert et al., 2006; Espin et al., 2000; Gansle et al., 2002; Ritchey & Coker, 2013; Videen et al., 1982). Although correlation examines how two variables are related, it does not provide information regarding the level of agreement between two variables. For example, when Espin et al. (2000) investigated whether teacher ratings were related to the students' CWS on CBM-WE narrative writing samples, the authors reported a significant correlation between the two factors ($r = .76$). This statistic demonstrates that overall, as teacher judgments increase, so do students' CWS scores on CBM-WE assessments, but it does not provide any information about whether the teacher rating and a specific CBM-WE score agrees. Furthermore, although percent agreement allows for an examination of how two variables agree, it does not consider chance. It

is possible that the two methods agree by chance and the kappa statistic takes that into account. Utilizing the kappa statistic, which considers both observed agreements and expected agreements (i.e., chance), allows for a more precise examination of agreement (Viera & Garrett, 2005). In their seminal article, Landis and Koch (1977) recommended the following guidelines for interpreting kappa estimates: (a) poor agreement (less than 0.00), (b) slight (0.00 to 0.20), (c) fair agreement (0.21 to 0.40), (d) moderate agreement (0.41 to 0.60) (e) substantial agreement (0.61 to 0.80), and (f) almost perfect agreement (0.81 to 1.00). Given the advantages of the kappa statistic, it may be beneficial to incorporate in studies attempting to examine to what extent teacher judgments and CBM-WE methods agree.

Purpose of the Study

The purpose of this study was to extend the previous research investigating level of agreement between teachers' judgments of students' writing skills and CBM-WE assessment scores by examining whether teachers' judgments and CBM-WE results identify the same level of risk (i.e., below the 10th percentile, within the 10th to 24th percentile, within the 25th to 75th percentile, within the 76th to 90th percentile, and above the 90th percentile) for female and male students. This was examined using both national and local norms. The following primary research questions guided this study:

1. What is the level of agreement between teachers' ratings of students' writing skills and TWW, WSC, and CWS on CBM-WE writing assessments? To date, no research has used kappa coefficients to examine agreement, and therefore there is not enough prior empirical evidence to substantiate a hypothesis regarding the specific levels of agreement. However, based on previous work, it was hypothesized that Kappa coefficients would have more instances of fair, moderate, or substantial levels of agreement when using the CWS scoring

metric compared to the TWW and WSC scoring metrics (Espin et al., 2000; Ritchey & Coker, 2013; Romig et al., 2017).

2. What is the association between teachers' ratings of student writing skills and students' CBM-WE performances using national and local norms? Based on prior empirical work (Eckert et al., 2006; Espin et al., 2000; Gansle et al., 2002; McFarland & Wallace, 2008; & Ritchey & Coker, 2013), it was hypothesized that there would be moderate correlation coefficients between teachers' ratings of student writing skills and students' CBM-WE performances. In addition, based on prior studies (Espin et al., 2000; Ritchey & Coker, 2013; Romig et al., 2017), it was hypothesized that this association would be higher when using the CWS scoring metric compared to the TWW and WSC scoring metrics. Further, based on work by Mowry and Farran (2016), it was hypothesized that the association would be higher when using local norms compared to when using national norms.
3. Given the observed differences in writing performances between males and females in previous literature, the third research question examined the levels of agreement as well as the associations between teachers' ratings of female and male students' TWW, WSC, and CWS using national and local norms. Because there is no known prior work examining teachers' levels of agreements or association with CBM-WE metrics no a priori hypotheses regarding the levels of agreement or associations for female and male students using local and national norms were developed.
4. According to CBM-WE performance, are males more likely to be identified as in need of additional writing support than females? This question will be explored using both national and local norms. Based on previous research (Farrington et al., 2014; Kim et al., 2014; Weiss et al., 2019), it was hypothesized that males would be more likely to be identified as in

need of additional writing support than females according to CBM-WE performance using both national and local norms.

5. According to teacher judgments, are males more likely to be identified in need of additional writing support than females? This question will be explored using both national and local norms. Based on previous literature (Beard & Burrell, 2010; Stagg Peterson & Kennedy, 2006), it was hypothesized that teachers would be more likely to identify males in need of additional writing support than females using both national and local norms.

Method

This study used data that were previously collected as part of a larger study that examined the effects of generalization training on writing generalization outcomes for students (Hier et al., 2020). The data used for this project were collected during the pre-intervention assessment and therefore were collected prior to any implementation of intervention procedures.

Student Participants

All students ($n = 276$) across first-, second-, and third-grade classrooms were recruited to participate in the study. After receiving approval from the university's Institutional Review Board and the participating school, a total of 276 student participants met the following criteria: (a) parental consent was obtained, (b) student assent was obtained, and (c) the student did not experience severe motor deficits that interfered with writing ability. However, as described in the results, 13 student participants were excluded from the study due to missing data, resulting in a final student participant sample of 263.

Student participants came from a rural elementary school in the northeast United States of America. A total of 81 first-grade students from five classrooms, 97 second-grade students from five classrooms, and 85 third-grade students from four classrooms participated, resulting in

a total student participant sample of 263. Overall, 53.6% of the students who participated were female, and the majority of students were White (82.5%), with fewer students identifying as Native American or Alaska Native (15.2%), multiracial (1.5%), and Black (0.8%). Additional participant demographic information is presented in Table 1.

Teacher Participants

All of the first- ($n = 5$), second- ($n = 5$), and third-grade ($n = 4$) teachers in the school participated in the study. The participants' teachers were all female and had similar levels of university training (i.e., a master's degree). Years of teaching experience ranged from 11 to 45 years ($M = 23.3$, $SD = 9.6$), while years of teaching at their current schools ranged from 2 to 45 years ($M = 20.7$, $SD = 10.79$). When all teachers were asked about additional trainings or certificates, three teachers reported having teaching certifications in literacy, one reported being certified to teach English for seventh to twelfth grade, and one reported being certified to teach pre-kindergarten to sixth grade.

Measures

CBM-WE

Students' writing performances on CBM-WE probes were examined using AIMSWeb® procedures (Powell-Smith & Shinn, 2004). Specifically, story stem prompts were used, and students were given 1 minute to plan and 3 minutes to write a story. Each CBM-WE probe (see Appendix A) was included in a student writing packet that contained other writing measures for the larger study (Hier et al., 2020). The CBM-WE probe consisted of two pages. The first page contained a story stem and lines for the students to write on, and the second page contained blank lines in case the student needed additional space. Additional blank pages were available if students needed more pages for writing.

Story stems with strong alternate-form reliability ($r = .73$ to $.90$) were used (McMaster et al., 2010). Due to the purpose of the broader study from which these data were collected (Hier et al., 2020), story stems were modified in two ways: (a) to prompt a simple sentence and (b) to change the subject form from first- to third-person. Original and modified versions of the story stems are presented in Table 2. Meta-analytic findings indicate that for kindergarten through fifth-grade, story stem prompts, on average, are significantly and moderately correlated with state-developed and commercially-developed writing assessments ($r = .40$; Romig et al., 2020). Furthermore, the use of a three-minute writing duration for students in grades kindergarten through fifth grade is, on average, significantly correlated with state-developed and commercially-developed tests ($r = .39$; Romig et al., 2020).

Teacher Judgments of Students' Writing Skills

Based on the teaching rating scale used by Begeny et al. (2008), a measure was created by the author to obtain the teachers' judgments of their students' writing performances (see Appendix B). For each teacher, this questionnaire listed all their current students and asked the teacher to rate each student's writing performance relative to their peers in the class using a 5-point Likert-type scale. I aligned the Likert-type scale to the writing risk categories used in RTI models (Shinn, 2005) to indicate students' need for supplemental interventions. Specifically, scale responses of '*much worse*' were categorized as students falling below the 10th percentile, scale responses of '*somewhat worse*' were categorized as students falling within the 10th to 24th percentile, scale responses of the '*about the same*' were categorized as students falling within the 25th to 75th percentile, scale responses of '*somewhat better*' were categorized as students falling within the 76th to 90th percentile, and scale responses of '*much better*' were categorized

as students falling above the 90th percentile. Because this measure was developed for this study, the psychometric properties have not been previously examined.

Outcome Variables

CBM-WE Student Writing Performances

The median scores of the three CBM-WE probes were used to estimate students' writing performances. Because CBM-WE are brief assessments that are sensitive to small changes over time (Espin et al., 2000), such as test error, the median score was used and is consistent with AIMSweb administration procedures (Powell-Smith & Shinn, 2004). To examine how levels of agreement might differ based on scoring metric used, TWW, WSC, and CWS were computed. A recent meta-analysis (Romig et al., 2017) examined the criterion validity of TWW, WSC, and CWS. Results indicated that all the scoring indices yielded moderate correlations with commercially developed tests or state- and district-developed achievement tools ($r = .37$ to $.51$).

TWW. Using AIMSWeb® standardized scoring procedures (Powell-Smith & Shinn, 2004), a TWW is defined as “any letter or group of letters separated by a space is defined as a word, even if the word is misspelled or is a nonsense word” (p. 9). The amount of TWW by each student on each of their writing samples were summed. The median score for each student was used for the analyses.

WSC. Using AIMSWeb® standardized scoring procedures (Powell-Smith & Shinn, 2004), a WSC is defined as any word that is spelled correct using “low-inference judgment” (p. 13). Similar with TWW, WSC were summed for each writing sample, and the median score for each student was used for data analyses.

CWS. As with the other scoring metrics, AIMSWeb® standardized scoring procedures (Powell-Smith & Shinn, 2004) were utilized to assess the amount of CWS in students' writing

samples. A CWS is defined as any “two adjacent writing units (words and punctuation) that are correct within the context of what is written” (Powell-Smith & Shinn, 2004, p. 11). CWS was totaled for each writing sample, and the students’ median scores were used for analyses.

Norms

Based on common instructional decision-making practices within RTI models (Shinn, 2005), students’ writing performances of TWW, WSC, and CWS were compared to AIMSWeb®, national norms from the 2015-2016 academic year. In addition, local norms were derived by computing the mean and standard deviation of the school’s population for TWW, WSC, and CWS to calculate percentiles (Hartman & Fuller, 1998).

Procedures

This study was carried out by one school psychology assistant professor, five school psychology graduate research assistants, and five undergraduate psychology research assistants. All of the research assistants were trained to 100% proficiency in administration of the CBM-WE measure. The data were collected in each of the students’ general education classrooms across two sessions in one week, each lasting approximately 30 minutes. These sessions occurred in the eighth month of the ten-month 2017 – 2018 school year (i.e., April). During the sessions, brief breaks were provided to limit handwriting fatigue.

At the beginning of each session, using procedural scripts, the lead research assistant gained the attention of the students by introducing themselves and stating the behavioral expectations of students during the session (i.e., “Listen to directions and instruction. Keep your packets closed until you are told to open them. Stay on the page we’re working on until you’re told to flip the page.”). Then, research assistants followed standardized AIMSWeb® procedures (Powell-Smith & Shinn, 2004) to administer the CBM-WE assessments (see Appendices C and

D). Specifically, students were provided with a story starter (see Appendix A) and one minute to plan what they wanted to write. Then, the students were given three minutes to write. During this time, research assistants walked around the room and provided verbal prompts to any students who stopped writing (i.e., told “keep writing the best story you can”). A total of three CBM-WE probes were administered so a median score could be obtained.

Simultaneously, as research assistants conducted the sessions with the students, teachers were asked to report their professional and educational experiences, such as how long they have been teaching and what additional training they have received.

Procedural Integrity

Procedural scripts were utilized to promote and assess procedural integrity (see Appendices C and D) and research assistants evaluated the implementation of the procedures using checklists for 68.85% ($n = 19$) of the total student data collection sessions ($n = 28$). Procedural integrity was calculated by dividing the number of steps completed by the total number of steps and multiplied by 100. The mean procedural integrity was 97.89% (range, 88% to 100%) across all sessions.

Interscorer Agreement

The CBM-WE probes were scored by two graduate school psychology students and one psychology undergraduate student. A two-hour training was provided to the scorers in which modeling and performance feedback were provided. Follow-up trainings with the scorers were completed until they demonstrated 100% proficiency scoring mock student writing samples according to AIMSWeb® scoring procedures (Powell-Smith & Shinn, 2004). No scoring for this study was completed until proficiency was met.

Interscorer agreement was assessed for 33% ($n = 91$) of the CBM-WE probes to ensure the data were reliably scored. Percent agreement between scorers was determined on a word-by-word basis and calculated as the number of agreements in scoring and dividing it by the total number of agreements and disagreements multiplied by 100 (Watkins & Pacheco, 2000). The overall average interscorer agreement was 96.51% (range, 72.73% to 100%) and the overall average Kappa was 0.91 (range, 0.38 to 1.0).

Results

Data Preparation

The primary researcher and undergraduate research assistants were responsible for entering the data into Microsoft Excel spreadsheets. The data were entered twice and checked for accuracy. Then, the data were transferred into IBM SPSS Statistics 26 (IBM Corp., 2017). SPSS was used to conduct all descriptive, correlational, and chi-squares analyses. In addition, local norms were created for each grade by computing percentiles. SPSS was also used to calculate the total cell frequencies for the kappa coefficients, and then Microsoft Excel was used to compute the kappa coefficients.

Data Inspection

Due to student absences (i.e., unable to collect any CBM data), there were 3 (1.09%) students whose outcome variable data were incomplete. Additionally, gender data were not available for 10 (3.66%) students. To examine whether these data were missing completely at random, a Little's test was conducted using SPSS. Results of a Little's test was not significant, $\chi^2(1, n = 276) = 3.647, p = 0.056$, thus indicating that the data were missing completely at random. Listwise deletions were done, and subsequent data analyses were conducted with the data from the remaining 263 participants.

Descriptive Statistics

Table 3 presents the mean, standard deviation, skewness, and kurtosis for the CBM-WE assessment data for each grade for TWW, WSC, and CWS. Overall, first grade students, on average, wrote 15.39 words, spelled 12.62 words correctly, and had 8.88 CWS within their writing samples. Second-grade students, on average, wrote 30.68 words, spelled 28.15 words correctly, and wrote 23.42 CWS. Third-grade students, on average, wrote 39.08 words, spelled 36.32 words correctly, and included 31.71 CWS on their CBM-WE sample.

The results of the local norm data, as well as the national norm data (AIMSweb®, 2018) appear in Table 4. Regardless of the norm type used (i.e., national or local), the majority of students across all grades had CBM-WE scores that fell within the average range across all scoring metrics (range, 50.6% to 61.6%), whereas teachers identified 31.6% of the students as having average writing skills and 35.4% of the students as having less than average (i.e., below average and well-below average) writing skills. When using national norms, less students were identified as having less than average writing skills based on their CBM-WE performances. Specifically, when using TWW scoring metric, 31.5% of students had scores less than average, whereas 18.9% and 19.7% were identified when using WSC and CWS scoring metrics, respectively. When using local norms, 23.5% of students had scores less than average ranges when using TWW, 22.8% when using WSC, and 24.7% when using CWS scoring metric. Overall, these descriptive results of the CBM-WE data suggest that students' scores were normally distributed.

Table 5 presents the frequencies of students in each risk category as identified by either CBM-WE assessment or teacher judgments. The first-grade teachers overall rated 32.1% ($n = 26$) of their students as having average writing skills, with fewer students being identified as

having well-below average (11.1%, $n = 9$), below average (17.3%, $n = 14$), above average (28.4%, $n = 23$), or well-above average (11.1%, $n = 9$) writing skills as rated by their teachers. Within the group of second-grade students, 30.9% ($n = 30$) of them were rated by their teachers as having average writing skills, with fewer students being identified as having well-below average (11.3%, $n = 11$), below average (25.8%, $n = 25$), above average (17.5%, $n = 17$), or well-above average (14.4%, $n = 14$) writing skills as rated by their teachers. For third-grade students, 31.8% ($n = 27$) were rated by their teachers as having average writing skills, whereas fewer students were rated by their teachers as having well-below average (11.8%, $n = 10$), below average (28.2%, $n = 24$), above average (20%, $n = 17$), or well-above average (8.2%, $n = 7$) writing skills. Similar to CBM-WE data, these results suggest that teacher judgments of student's writing skills were also normally distributed.

One-sample t-tests were conducted to examine whether this sample obtained scores significantly different from the national population (AIMSWeb®, 2018). First-grade students wrote significantly less words ($M = 15.39$, $SD = 7.31$), had less correctly spelled words ($M = 12.62$, $SD = 6.67$), and had less CWS ($M = 8.88$, $SD = 5.63$) in their writing samples than the population as whole, $t(80) = -5.68$, $p < 0.001$; $t(80) = -4.56$, $p < 0.001$; $t(80) = -11.38$, $p < 0.001$, respectively. Second-grade students did not have significantly more words ($M = 30.68$, $SD = 12.18$) in their writing samples when compared to the national population, $t(96) = -1.07$, $p = 0.289$). However, second-grade students did write more WSC ($M = 28.15$, $SD = 11.78$) and CWS ($M = 23.42$, $SD = 12.1$) than the national population, $t(96) = 3.48$, $p = 0.001$; $t(96) = -7.8$, $p < 0.001$. Third-grade students in this sample did not have significantly different TWW ($M = 39.08$, $SD = 11.68$), WSC ($M = 36.32$, $SD = 11.52$), or CWS ($M = 31.71$, $SD = 12.54$), $t(84) = 0.06$, $p = 0.95$; $t(84) = 2.66$, $p = 0.009$; $t(84) = 1.26$, $p = 0.213$, respectively. Overall, these one-sample t-

tests suggest that first grade students' writing samples were significantly different from the national norm, and third-grade students' writing samples were not significantly different from the national population.

Major Analyses

Research Question 1

To determine the levels of agreement between teachers' rating of students' writing skills and students' CBM-WE performances using national and local norms, kappa coefficients were calculated for the three scoring metrics, and for the five risk categories (see Table 6).

TWW. For the overall sample, across both sets of norms, kappa coefficients ranged from poor to fair agreement when using the TWW metric (range, -0.03 to 0.22). None of the kappa coefficients obtained using TWW indicated moderate levels of agreement.

WSC. For the overall sample using national and local norms, kappa coefficients ranged from slight to fair agreement when using the WSC metric (range, 0.04 to 0.35). For both national and local norms, levels of agreement were fair for students with percentile scores below the 10th percentile or above the 90th percentile. Slight agreement levels were observed for students whose scores fell between the 10th and 90th percentiles.

CWS. When using the CWS scoring metric, levels of agreement ranged from slight to moderate (range, 0.11 to 0.46) across both sets of norms. Across both sets of norms, moderate agreement was indicated for students with scores above the 90th percentile. Levels of agreement were fair when identifying students whose scores fell below the 25th percentile for both types of norms, as well as for students whose scores fell between 25th and 75th percentiles when using local norms. None of the kappa coefficients obtained using CWS indicated poor agreement.

TWW, WSC, and CWS. Across both types of norms, the TWW scoring metric was associated most frequently with poor level of agreement when compared to kappa levels obtained using WSC or CWS. Conversely, CWS was the only scoring metric to obtain moderate levels of agreement. Partly congruent with my hypothesis, CWS had more instances of fair or moderate agreement levels. However, none of the scoring metrics were associated with substantial or almost perfect levels of agreement with teachers' ratings.

Research Question 2

To determine the association between teachers' ratings of student writing skills and students' CBM-WE performances using national and local norms, Spearman rank order correlations were computed for the following: (a) teacher ratings and CBM-WE TWW scores, (b) teacher ratings and CBM-WE WSC scores, and (c) teacher ratings and CBM-WE CWS scores. For each of these, the overall correlations using national and local norms were computed.

For the overall sample, there were moderate, positive correlations between teacher ratings and CBM-WE performances using TWW and WSC metrics based on national norms, $r = .36, p < .001$ and $r = .43, p < .001$, respectively. There was also a moderate, positive correlation between teacher ratings and the interval ranking of TWW based on local norms, $r = .42, p < .001$. Between teacher ratings and interval ranking of CWS based on national norms, there was a large, positive correlation, $r = .52, p < .001$. Similarly, there were large, positive correlations between teacher ratings and interval rankings of WSC and CWS based on local norms, $r = .54, p < .001$ and $r = .62, p < .001$, respectively.

To test my hypothesis that associations would be higher when using the CWS scoring metric compared to the TWW and WSC scoring metrics, Fisher's r to z transformations were conducted and an online calculator (Lowry, 2001) was used. Specifically, correlation coefficients

were converted to z scores (i.e., normally distributed) and confidence intervals are compared for overlap. Partly congruent with my hypothesis, the relationship between teachers' rating and CWS was larger than TWW across both national and local norms ($z = -2.47, p = 0.006$; $z = -3.07, p = 0.001$). However, the difference between the correlations using WSC and CWS was not significant for either set of norms.

National and Local Norms. In order to determine whether there were differences in the correlations derived from national and local norms, Fisher's r to z transformation were conducted and an online calculator (Lowry, 2001) was used to test the significance of the difference between the transformed correlations. Contrary to my hypotheses that correlations would be larger when using local norms than when using national norms, the results indicated that there were no statistically significant differences on any of the writing metrics. For the TWW metric, there was no statistically significant difference in the relationship between teacher ratings when using national or local norms ($z = -0.81, p = 0.418$), nor were there statistically significant differences using the WSC metric ($z = -1.55, p = 0.1211$) or the CWS metric ($z = 0.057, p = 0.1141$).

Research Question 3

To examine whether students' gender impacted the levels of agreement or the associations between teachers' ratings of students' TWW, WSC, and CWS using national and local norms, descriptive comparisons based on gender were conducted for the levels of agreement and statistical testing was conducted for the associations. Prior to examining the levels of agreement and associations, a series of ANOVAs were conducted to determine whether there were gender differences in students writing performance within each grade. All the statistical assumptions underlying the ANOVA were met except for the assumption of homogeneity across

some of the CBM-WE metrics. As a result, Welch corrections were applied to address this violation (Delacre et al., 2019).

For first-grade students, the results of ANOVAs using a Welch correction indicated that there were statistically significant differences between the performances of males and females for TWW, $F(1, 75.62) = 35.14, p < .001, \eta_p^2 = 0.08$, WSC, $F(1, 76.24) = 30.41, p < .001, \eta_p^2 = 0.08$, and CWS, $F(1, 77.99) = 19.84, p < .001, \eta_p^2 = 0.07$, with first-grade female students writing more words (TWW), having more correctly spelled words (WSC), and more CWSs when compared to their male counterparts. For second- and third-grade students, the assumption of homogeneity was met for CBM-WE performances when using TWW and WSC but was violated when using the CWS CBM-WE metric. Results of those ANOVAs (with a Welch correction for CWS) indicated that there were statistically significant differences between the performances of second-grade males and females for TWW, $F(1, 94.41) = 18.28, p < .001$, WSC, $F(1, 92.19) = 21.46, p < .001$, and CWS, $F(1, 88.00) = 18.29, p < .001$, with second-grade female students having more TWW, WSC, and CWS in their writing samples than second-grade male students. Similar results were obtained when investigating the performances of third-grade students. Results of ANOVAs (with a Welch correction for CWS) indicated that there were statistically significant differences between the performances of third-grade males and females for TWW, $F(1, 83) = 7.93, p = .006$, WSC, $F(1, 83) = 8.98, p = .004$, and CWS, $F(1, 81.97) = 9.91, p = .002$, with third-grade female students having more TWW, WSC, and CWS on the CBM-WE than third-grade male students. In summary, the results of these analyses indicated that female students outperformed male students on the CBM-WE assessment across all grades and all three CBM-WE metrics.

Levels of Agreement for Females Only. *TWW*. Levels of agreement ranged from poor to slight agreement when using the *TWW* scoring metric (range, -0.05 to 0.18). Across both sets of norms, poor agreement was indicated when identifying which female students' scores fell below the 10th percentile. Similarly, across both types of norms, when asked to identify which female students' scores below above 90th percentile, levels of agreement were in the slight agreement range.

***WSC*.** Levels of agreement varied from poor to fair agreement (range, -0.04 to 0.35) when using the *WSC* scoring metric. For both types of norms, poor agreement was observed when detecting female students who demonstrated performance below the 10th percentile. Fair agreement was indicated when using local norms to compare ratings of female students whose scores fell above the 90th percentile.

***CWS*.** Levels of agreement ranged from poor to moderate t when using the *CWS* metric across norms (range, -0.05 to 0.44). Congruent with the *TWW* and *WSC* results, levels of agreement were poor with female students' scores below the 10th percentile. Moderate agreement was observed for female students whose scores were above the 90th percentile.

***TWW, WSC, and CWS*.** Across all scoring metrics and both sets of norms, there were poor levels of agreement when identifying female students' whose writing skills fell below the 10th percentile. The only instances of moderate levels of agreement that were observed were between teachers' ratings of female with writing performance above the 90th percentile.

Levels of Agreement for Males Only. *TWW*. Levels of agreement varied from poor to fair when assessing male students' writing skills using the *TWW* scoring metric (range, -0.06 to 0.28). Across local and national norms, fair agreement was indicated for male students with scores

below the 10th percentile, whereas poor agreement was observed for male students with scores at the 25th percentile or greater.

WSC. When using the WSC scoring metric, poor to fair levels of agreement were observed (range, -0.09 to 0.35). Fair agreement was observed for male students with scores below the 10th percentile, as well as for male students with scores above the 90th percentile when using national norms only.

CWS. When using the CWS scoring metric, slight to moderate level of agreement were observed (range, 0.09 to 0.48). Moderate agreement was observed for male students with scores below the 10th percentile, as well as for male students with scores above the 90th percentile when using local norms only.

TWW, WSC, and CWS. For male students across both set of norms and all scoring metrics, there were fair to moderate levels of agreement between teachers' ratings and CBM-WE score metrics for students with scores below the 10th percentile. Across both sets of norms, the only instances of poor agreement between teachers' ratings and CBM-WE scoring metrics occurred for male students with scores above the 25th percentile for TWW and above the 75th percentile for WSC. Congruent with my hypothesis, levels of agreements were higher when comparing agreement between teacher ratings and CWS scores in comparison to TWW and WSC scores.

Correlations Between Teachers' Rating of Female Students and CBM-WE Scores.

When examining female students only, there were small, positive correlations between teacher ratings of TWW and WSC based on national norms, $r = .26, p < .001$ and $r = .28, p < .001$, respectively (see Table 9). There was a medium, positive correlation between teacher ratings of CWS based on national norms, $r = .44, p < .001$. There were also medium, positive correlations between teacher ratings of TWW and WSC based on local norms, $r = .31, p < .001$ and $r = .44, p$

< .001, respectively. For females only, the only large, positive correlation was between teacher ratings of CWS based on local norms, $r = .53, p < .001$.

National and Local Norms. Results of statistical testing indicated that the relationships between teacher ratings and CBM-WE scores for female students using national and local norms were not significantly different for TWW, $z = -0.53, p = 0.596$, WSC, $z = -1.45, p = 0.147$, or CWS, $z = -0.93, p = 0.352$.

Correlations Between Teachers' Rating of Male Students and CBM-WE Scores.

When examining male students only, there were small, positive correlations between teacher ratings of TWW based on both national and local norms, $r = .24, p = .009$ and $r = .29, p = .001$, respectively (see Table 9). There were medium, positive correlations between teacher ratings and WSC based on national and local norms, $r = .38, p < .001$ and $r = .46, p < .001$. Similarly, there was a medium, positive correlation between teacher ratings of CWS based on national norms, $r = .49, p < .001$. For male students only, the only large, positive correlation found was between teacher ratings of CWS based on local norms, $r = .59, p < .001$.

National and Local Norms. Results of statistical testing indicated that the relationships between teacher ratings and CBM-WE scores for male students using national and local norms were not significantly different for TWW, $z = -0.41, p = 0.682$, WSC, $z = -0.75, p = 0.453$, or CWS, $z = -1.11, p = 0.267$.

Correlational Analyses Between Female and Male Students. To determine whether the relationships between teacher ratings and CBM-WE performances were significantly different based on the students' gender, a series of statistical tests were conducted. Results, as shown in table 9, indicated that regardless of the norm type used (i.e., national or local) or scoring metric

used (i.e., TWW, WSC, or CWS), there were no differences in the correlational coefficients as a function of gender.

Research Question 4

To determine whether males were more likely to be identified as in need of additional writing support than females according to CBM-WE performance, chi-square tests of independence were conducted between gender and CBM-WE performance using TWW, WSC, and CWS. For the purposes of this research question, students' CBM-WE performance was dichotomized into two categories: (a) below the 25th percentile and (b) at or above 25th percentile. There were no violations of the statistical assumptions underlying the chi-square analyses (i.e., no cells will have an expected count of less than five).

Results of chi-square tests revealed a significant association between gender and CBM-WE TWW results for both national, $\chi^2(1, n = 263) = 24.22, p < .001, phi = .30$, and local norms $\chi^2(1, n = 263) = 28.23, p < .001, phi = .33$. Significant associations between gender and writing risk were also observed for CBM-WE WSC results across national, $\chi^2(1, n = 263) = 22.17, p < .001, phi = .29$, and local norms, $\chi^2(1, n = 263) = 19.97, p < .001, phi = .28$, as well as for CBM-WE CWS results for national, $\chi^2(1, n = 263) = 15.99, p < .001, phi = .25$, and local norms, $\chi^2(1, n = 263) = 18.12, p < .001, phi = .26$. Taken together these results indicate that regardless of the CBM-WE metric used and regardless of the type of norm used, males were more likely to be identified in need of additional writing supports (i.e., falling below the 25th percentile) than females, and are commensurate with my hypothesis.

Research Question 5

To determine whether teachers were more likely to identify males in need of additional writing support than females, chi-square tests of independence were conducted between gender

and teacher rating. The teachers' rating was dichotomized into two categories: (a) much worse or somewhat worse and (b) about the same, somewhat better, or much better. There were no violations of the statistical assumptions underlying the chi-square analyses.

Results of this analysis indicated that there was a significant association between gender and teacher rating, $\chi^2(1, n = 263) = 44.9, p < .001, phi = .26$. Congruent with my hypothesis, results revealed that male students were more likely to be identified by their teachers as in need of additional writing supports (i.e., falling below 25th percentile) compared to their female counterparts.

Discussion

Given that writing is an important skill needed for other academic subject areas (Ray et al., 2016), success in higher education (Addison & McGee, 2010), and being hired in the adult work force (National Commission on Writing, 2005), it is imperative that students develop adequate writing skills. As a part of the RTI process, educators must first identify students in need of additional support to provide them with high-quality interventions (Deno, 2015). The purpose of this study was to examine two primary methods that may be used as part of the RTI process—teacher judgments and universal screening using CBM-WE—used in schools to identify students in need of additional academic support in the area of writing. This study aimed to do this in four ways: (a) examine levels of agreement, (b) examine the relationships between the two variables, (c) examine whether gender affects levels of agreement and relationships between variables, (d) investigate whether teacher judgments identify more male students than female students as having less than average writing skills, and (e) investigate whether performances on CBM-WE identify more male students than female students as having less than average writing skills.

Agreement of Teachers' Rating and Students' Writing Skills

To examine levels of agreement between teachers' ratings and students' CBM-WE performances, kappa coefficients were calculated for the three scoring metrics and for each risk category. Results indicated that agreement levels ranged between poor agreement and moderate agreement. For the overall sample, CWS was the only scoring metric to indicate levels of moderate agreement. Conversely, TWW was the only scoring metric to have a poor level of agreement, thus indicating that CWS may be more aligned with teachers' ratings of students' writing skills. To date, this is the only study known to this author that uses Kappa analyses to explore the agreement between teachers' rating and CBM-WE scores. Given that the equation for Kappa takes chance into account, it was likely that the findings of this study would be lower than levels of agreement reported in prior studies using correlational analyses (Espin et al., 2000; Gansle et al., 2002; McFarland & Wallace, 2008; Ritchey & Coker, 2013). However, these findings are important because they suggest that when chance is considered, poor agreement is found when using TWW for the overall sample, thereby revealing instances of disagreement that may have been hidden using correlational analyses only. Furthermore, these results suggest that CWS has moderate levels of agreement suggesting that this scoring metric may better align with teachers' perceptions of students' writing skills.

When comparing levels of agreement for the overall sample, kappa coefficients were similar between national and local norms when using CWS and WSC. However, when using TWW, there were poor to slight levels of agreement using national norms, whereas there were slight to fair levels of agreement when using local norms. These findings suggest that when using TWW, teachers' ratings may better align with local norms of the school when compared to using national norms.

Relationship Between Teachers' Ratings and Students' Writing Skills

Results of correlational analyses indicated that there were overall significant, positive, moderate to large associations between teachers' ratings and students' writing performances across all scoring metrics. These findings are consistent with several other studies that found significant, positive correlations between teachers' ratings of students' skills and students' performances on CBM-WE (Espin et al., 2000; McFarland & Wallace, 2008; Videen et al., 1982). TWW was the only scoring metric that did not have large correlations with teachers' ratings and the associations between TWW and teachers' ratings were significantly lower than the relationships between CWS and teachers' ratings across both types of norms. Taken together, these findings suggest that TWW may not be the best indicator of teachers' perceptions of students' writing skills, which is somewhat consistent with Gansle et al.'s (2002) findings that TWW was not significantly related to teachers' ratings.

The results of subsequent analyses indicated that the relationship between teachers' ratings and students' writing skills did not differ based on whether national or local norms were used, although the magnitude of the correlations were lower for some of the writing metrics. These results were surprising given previous work suggesting that the use of local norms may have higher predictive validity with state-developed assessments (Sandberg Patton et al., 2014). However, one factor that may account for this is the similarity between national and local norms for this sample of students. For example, third-grade students' CBM-WE scores in this study did not differ from the national population on any scoring metric. It may be possible that in settings where the local norms significantly differ from the national norms, similar analyses may produce different findings.

Gender-based Analyses

Given the inconsistent findings in previous research examining writing performances between male and female students (Farrington et al., 2014; Hier et al., 2019; Jewell & Malecki, 2005; Keller-Margulis et al., 2015; Kim et al., 2015; Malecki & Jewell, 2003; McMaster et al., 2017; Weiss et al., 2019), this study extended the literature by examining the agreement levels and relationship between CBM-WE performances and teachers' ratings based on gender, as well as determine whether male students were more likely to be identified as in need of additional writing support (i.e., below 25th percentile) than female students, using either students' performances on CBM-WE or teacher judgment. In this study, male students were more likely to be identified as needing additional writing support than female students using both CBM-WE data and teacher judgment data. This is not surprising given that female students are more likely than males to obtain proficient scores on national assessments over the past three decades (Reilly, 2019), in addition to female students typically outperforming their male counterparts on CBM-WE assessments (Farrington et al., 2014; Kim et al., 2015; Weiss et al., 2019).

Levels of Agreement for Female Students

For female students, kappa analyses indicate that agreement levels ranged between poor agreement to moderate agreement. For TWW, levels of agreement across all risk categories were poor when using national norms. When using local norms and TWW, there was poor agreement level for students below the 10th percentile, but slight agreement for students at and above the 10th percentile. Similar findings were found for WSC across both types of norms. Moderate agreement levels were found for female students above the 90th percentile only when using CWS scoring metric.

One novel finding of this study, and possibly the most important, was discovered examining kappa coefficients for female students, as this was the first study to examine levels of

agreement between teachers' ratings and CBM-WE scores using kappa analyses to take chance into account. Across all three scoring metrics and both types of norms, there were poor levels of agreement for female students below the 10th percentile. As a result, it is possible that these female students are not being identified as needing support and thereby not receiving interventions that they likely need. Although levels of agreement tended to increase for female students above the 10th percentile, arguably, level of agreement for students whose writing skills fall below the 10th percentile might be the most important, as these may be the students most in need of additional writing support. Thus, it may be crucial for schools to use converging data sources (i.e., teacher referrals and CBM-WE universal screening) to accurately identify female students in need of additional support in the area of writing.

Levels of Agreement for Male Students

For male students, levels of agreement ranged between poor agreement and moderate agreement. Although levels of agreement tended to be lower for female students whose writing performances fall below the 10th percentile, opposite trends occurred for the male students in this study. Across all three scoring metric and with both types of norms, levels of agreement were fair to moderate for male students whose writing performances fell below the 10th percentile; whereas poor levels of agreement were found for male students whose skills fell at or above the 25th percentile when using TWW and above the 75th percentile for WSC. Previous research conducted by Reilly et al. (2019) found that over the past three decades of NAEP data being collected, female students were more likely to obtain proficient scores on writing assessments when compared to their male counterparts, and this gender difference was associated with an overall medium effect size ($d = .54$). Furthermore, teachers have been found to rate female students' writing abilities higher than male students' (Troia et al., 2012). Given that female

students typically outperform their male peers and teachers perceive gender differences in writing, it may be possible that teachers are paying more attention to male students' writing and thereby leading to higher agreement with CBM-WE assessments.

Overall, the results of this study suggest that levels of agreement of different assessments (i.e., teacher judgments and CBM-WE) differ as a function of gender, especially for female students with less than average writing skills. Conversely, for male students with less than average writing skills, there were fair to moderate levels of agreement; thus suggesting that teacher judgments and CBM-WE results are better aligned for male students when compared to female students. Therefore, it might be critical for schools to use converging data sources to identify female students in need of additional academic support, but may not be needed for male students.

Limitations

There are several limitations in this study that need to be considered when interpreting the results. One major limitation was its small sample size of 14 teachers. Not only might this limit the generalizability of the results to the general population, but also having a small sample size can limit statistical conclusion validity. In addition, given the limited sampling approach (i.e., convenience sample) used in the present study, it is unknown to what extent these findings would be replicated with more or different teachers. Another limitation is that these results relied on teacher report. Not only might teachers have different writing expectations based on gender, but teachers may also have other inherent biases that may impact their judgments of students' writing skills. Furthermore, this study examined writing skills based on CBM-WE data, which may be biased for some students. For example, the primary outcome measures assessed in this study focused on writing fluency, thereby favoring students who simply write more. No attention

was given to more qualitative aspects of the students' writing, such as whether students' writing pertained to the prompt or the use of diverse or complex vocabulary. Furthermore, scoring of the CBM-WE writing samples was based on AIMSWeb® (2018) scoring guidelines, which uses White mainstream English as its standard. It is possible that agreement levels between teachers and CBM-WE would be different in a language-inclusive setting.

Because the study relied upon a cross-sectional approach and correlational outcomes, additional limitations that should be considered. It is possible that the results of this study would be different if data were collected at a different time point in the academic year or if a longitudinal approach was used to measure students' writing development over the course of the school year. Furthermore, the design of this study did not allow for causal inferences to be made and it is unknown whether student gender causes levels of agreement to vary.

Finally, there are limits related to data collection that also need to be considered. Gender in this study was conceptualized in a binary way (i.e., female or male), and thereby was not inclusive or accurately represent the spectrum of genders. In addition, these data were provided by the school, which gathered this information from parent report. Therefore, it is possible that students' gender was not accurately portrayed in this study. Lastly, another limitation related to data collection is that because normative data were unavailable for CIWS, this study did not use CIWS as a scoring metric despite it having the highest criterion validity estimates compared to other CBM-WE scoring metrics (Romig et al., 2017).

Implications for Future Research and Practice

Although previous studies have investigated agreement levels, none to date used kappa analyses, despite the advantages of kappa taking chance into consideration. Given that poor levels of agreement were reported in the present study for female students with writing skills

below the 25th percentile, it appears that future research examining levels of agreement should consider using kappa to get a more accurate estimate of agreement levels. Future research investigating levels of agreement between teacher ratings and students' performances on CBM-WE should also consider using a longitudinal approach, which would allow researchers to not only investigate how levels of agreement might change over time, but also examine whether these agreement levels change as students develop their writing skills. For example, it may be possible that as students become more proficient in their writing skills, metrics measuring accuracy (i.e., CIWS) may better align with teacher judgments than fluency metrics (i.e., WSC and CWS). Future research should also investigate to whether developing local norms for universal school screening is needed. The results of this study suggest that there may be little to no benefit, as the local norms were not significantly different from the national population. However, it may be possible that in other settings in which students' performances do differ significantly from the national population, developing local norms is critical for adequate agreement levels.

In terms of school practice, there are several implications from this study. One major takeaway from this study is that it may be necessary for schools to use converging data sources (i.e., teacher referral and universal screening using CBM-WE) to accurately identify female students in need of academic support. One of the most novel and impactful findings of this study suggested that female students functioning below the 25th percentile were not being identified as experiencing writing difficulties based on teacher report. Furthermore, the results of this study suggest that schools may benefit from relying upon more time-intensive CBM-WE scoring metrics (e.g., WSC and CWS) given their convergence with teachers' perceptions of student writing skills.

Conclusions

Given that national assessments indicate that a majority of students are not demonstrating proficient writing skills (NCES, 2012; Persky et al., 2003), and writing skills impact performance in other academic areas (Ray et al., 2016) as well as limit students' access to higher education (Addison & McGee, 2010) and opportunities for jobs in the adult workforce (National Commission on Writing, 2005), it is imperative that educators are accurately identifying students in need of academic support.. Typically, the two primary ways schools identify students in need of writing support are through teacher referral and standardized assessments such as CBM-WE. However, to date, it is unknown to what extent these methods agree when chance is taken into account. In addition, due to previous literature suggesting gender differences in writing performances among students (Farrington et al., 2014; Jewell & Malecki, 2005; Reilly et al., 2019), as well as gender differences in teacher judgments and feedback of students' writing performances (Beard & Burrell, 2010; Stagg Peterson & Kennedy, 2006) this study examined how levels of agreement varied based on gender and how these levels of agreement varied based on gender. Finally, this study examined how levels of agreement varied based on scoring metric and type of norm used.

Overall, this study found that levels of agreement ranged from poor agreement to moderate agreement, with variations observed between scoring metrics and gender. More specifically, and possibly most alarming, there were poor levels of agreement for female students whose writing skills fell below the 10th percentile across all scoring metrics and using both national and local norms. This is alarming because it suggests that unless schools are using converging data sources, female students may not be identified and thereby not receive the support they need in the area of writing. The novelty of adding kappa analyses into levels of

agreement between CBM-WE and teacher judgments revealed that levels of agreement may not be as high as previously thought and warrants more investigation in future research.

Table 1*Demographic Data for Student Participants*

Characteristics	<i>n</i>	%
Gender		
Female	141	53.6
Male	122	46.4
Race and Ethnicity		
Asian	0	0
Black or African American	2	0.8
Latinx	0	0
Multiracial	4	1.5
Native American	40	15.2
White	217	82.5
Grade		
First	81	30.8
Second	97	36.9
Third	85	32.3

Table 2*Original and Modified Story Prompts*

Story Prompt	Original	Modified
Story Stem #1	I was walking home when I found a \$100 bill on the sidewalk and....	A boy was walking home when he found a \$100 bill on the sidewalk. The boy...
Story Stem #2	One day I went to school but nobody was there except me, so I...	One day, Lily went to school but nobody was there except her. Lily...
Story Stem #3	I was walking down the street when I saw....	My friend was walking down the street when he saw something. My friend...

Table 3*Descriptive Results of Students' Performance on CBM-WE*

	TWW			WSC			CWS		
	<i>M (SD)</i>	Skewness	Kurtosis	<i>M (SD)</i>	Skewness	Kurtosis	<i>M (SD)</i>	Skewness	Kurtosis
Grade 1	15.39 (7.31)	0.90	0.77	12.62 (6.67)	0.91	0.72	8.88 (5.63)	0.78	-0.19
Females	19 (7.13)	0.84	0.33	15.75 (6.6)	0.8	0.26	11.14 (5.7)	0.40	-0.85
Males	11.09 (4.8)	0.77	0.80	8.89 (4.55)	0.94	1.41	6.2 (4.26)	1.47	3.30
Grade 2	30.68 (12.18)	0.40	0.60	28.15 (11.77)	0.55	0.68	23.42 (12.1)	0.83	0.54
Females	35.4 (11.95)	0.66	0.51	33.03 (11.84)	0.67	-0.01	28.07 (12.85)	0.67	-0.35
Males	25.66 (10.37)	-0.17	0.17	22.97 (9.31)	-0.17	0.10	18.48 (9.01)	0.44	0.80
Grade 3	39.08 (11.68)	-0.25	-0.24	36.32 (11.52)	-0.11	-0.13	31.71 (12.53)	0.50	0.60
Females	42.16 (12.16)	-0.63	0.26	39.54 (12.16)	-0.4	0.19	35.26 (13.5)	0.34	0.37
Males	35.26 (9.69)	-0.17	-0.41	32.34 (9.39)	-0.28	-0.6	27.32 (9.71)	0.07	-0.30

Note. TWW = Total Words Written; WSC = Words Spelled Correctly; CWS = Correct Writing Sequences.

Table 4*National and Local Norms of Students' CBM-WE Performances*

	National Norms			Local Norms		
	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
	TWW					
10 th	9	16	23	7.2	16	22.6
25 th	14	24	30	10	23	31.5
50 th	20	32	39	13	31	41
75 th	28	41	49	21.5	37	48.5
90 th	35	49	59	24.8	48.6	54
	WSC					
10 th	5	11	16	5.2	14	20.6
25 th	10	17	23	8	20.5	29.5
50 th	16	24	33	11	28	38
75 th	23	33	43	16.5	35	44.5
90 th	31	42	54	22.8	43.2	49.8
	CWS					
10 th	2	8	13	3	10.9	15.2
25 th	5	13	21	5	13.5	22
50 th	11	21	30	7	22	31
75 th	20	30	43	12.5	29.5	39.5
90 th	26	39	56	18	42.2	46.8

Note. TWW = Total Words Written; WSC = Words Spelled Correctly; CWS = Correct Writing Sequences.

Table 5*Frequencies in Risk Categories by CBM-WE Results and Teacher Ratings*

	Grade 1		Grade 2		Grade 3	
	Females (<i>n</i> = 44)	Males (<i>n</i> = 37)	Females (<i>n</i> = 50)	Males (<i>n</i> = 47)	Females (<i>n</i> = 47)	Males (<i>n</i> = 38)
TWW ^a						
Frequencies						
Below 10 th Percentile	1 (0)	16 (6)	0 (1)	8 (9)	5 (5)	3 (3)
10 th – 24 th Percentile	11 (1)	13 (10)	7 (6)	11 (10)	2 (2)	6 (9)
25 th – 75 th Percentile	27 (24)	8 (19)	31 (24)	25 (22)	22 (22)	26 (23)
76 th – 90 th Percentile	2 (11)	0 (2)	4 (11)	2 (5)	16 (10)	3 (2)
Above 90 th Percentile	3 (8)	0 (0)	8 (8)	1 (1)	2 (8)	0 (1)
WSC ^b						
Frequencies						
Below 10 th Percentile	0 (0)	4 (4)	0 (0)	4 (10)	3 (4)	3 (4)
10 th – 24 th Percentile	6 (2)	19 (11)	3 (7)	8 (7)	2 (4)	3 (7)
25 th – 75 th Percentile	33 (23)	14 (19)	24 (22)	30 (25)	21 (20)	27 (25)
76 th – 90 th Percentile	4 (20)	0 (2)	12 (12)	3 (5)	14 (11)	5 (2)
Above 90 th Percentile	1 (9)	0 (1)	11 (9)	2 (0)	7 (8)	0 (0)
CWS ^c						
Frequencies						
Below 10 th Percentile	4 (1)	4 (4)	0 (1)	4 (8)	2 (3)	4 (5)
10 th – 24 th Percentile	0 (3)	10 (10)	4 (6)	11 (10)	5 (6)	4 (8)
25 th – 75 th Percentile	32 (23)	22 (20)	27 (24)	27 (24)	27 (22)	27 (20)
76 th – 90 th Percentile	8 (9)	1 (2)	7 (11)	4 (4)	8 (8)	3 (4)
Above 90 th Percentile	0 (8)	0 (1)	12 (8)	1 (1)	5 (8)	0 (1)
Teacher Ratings						
Frequencies						
Below 10 th Percentile	2	7	3	8	3	7
10 th – 24 th Percentile	3	11	8	17	10	14
25 th – 75 th Percentile	12	14	16	14	15	12
76 th – 90 th Percentile	19	4	11	6	14	3
Above 90 th Percentile	8	1	12	2	5	2

Note. Students' median scores for the three CBM-WE assessments were used for the analyses. Frequencies using local norms are presented in parentheses. ^aTWW = Total Words Written; ^bWSC = Words Spelled Correctly; ^cCWS = Correct Writing Sequences; ^dSD = Standard Deviation.

Table 6

Kappa Coefficients between Teacher Ratings of Students' Writing Skills and CBM-WE Performances

Risk Category	National Norms			Local Norms		
	Overall (n=263)	Females Only (n = 141)	Males Only (n = 122)	Overall (n = 263)	Females Only (n = 141)	Males Only (n = 122)
TWW						
Below 10 th	0.19	-0.05	0.21	0.22	-0.05	0.28
10 th -24 th	0.09	0.00	0.1	0.21	0.12	0.2
25 th – 75 th	-0.03	-0.01	-0.05	0.01	0.05	-0.03
76 th – 90 th	0.00	-0.03	-0.06	0.08	0.04	0.00
Above 90 th	0.17	0.16	-0.01	0.2	0.18	-0.02
WSC						
Below 10 th	0.26	-0.03	0.35	0.23	-0.04	0.28
10 th -24 th	0.19	0.09	0.18	0.17	0.2	0.1
25 th – 75 th	0.07	0.09	0.06	0.07	0.09	0.05
76 th – 90 th	0.04	0.02	-0.09	0.15	0.17	-0.1
Above 90 th	0.24	0.19	0.27	0.35	0.35	-0.01
CWS						
Below 10 th	0.36	-0.02	0.46	0.32	-0.05	0.42
10 th -24 th	0.26	0.27	0.22	0.3	0.24	0.29
25 th – 75 th	0.17	0.18	0.16	0.24	0.23	0.26
76 th – 90 th	0.11	0.07	0.12	0.17	0.16	0.09
Above 90 th	0.45	0.44	0.32	0.46	0.43	0.48

Note. Interpretation of kappa: < 0 = poor agreement; 0.01 to 0.2 = slight agreement; 0.21 to 0.4 = fair agreement; 0.41 to 0.6 = moderate agreement; 0.61 to 0.8 = substantial agreement; 0.81 to 0.99 = almost perfect agreement (Landis and Koch, 1977).

Table 7*Correlations Between Teacher Ratings and CBM-WE Performances for Overall Sample*

Variable	1	2	3	4	5	6	7
1. Teacher Rating	—	—	—	—	—	—	—
2. TWW with national norms	.36**	—	—	—	—	—	—
3. WSC with national norms	.43**	.85**	—	—	—	—	—
4. CWS with national norms	.52**	.63**	.75**	—	—	—	—
5. TWW with local norms	.42**	.86**	.79**	.63**	—	—	—
6. WSC with local norms	.54**	.80**	.81**	.75**	.88**	—	—
7. CWS with local norms	.62**	.66**	.71**	.91**	.69**	.81**	—

Note. ** $p < .01$.

Table 8

Correlations Between Teacher Ratings and CBM-WE Performances for Female and Male Students

Variable	1	2	3	4	5	6	7
1. Teacher Rating	—	(.24**)	(.38**)	(.49**)	(.29**)	(.46**)	(.59**)
2. TWW with national norms	.25**	—	(.86**)	(.56**)	(.84**)	(.75**)	(.55**)
3. WSC with national norms	.28**	.81**	—	(.71**)	(.78**)	(.77**)	(.65**)
4. CWS with national norms	.44**	.60**	.74**	—	(.56**)	(.70**)	(.90**)
5. TWW with local norms	.31**	.83**	.73**	.60**	—	(.82**)	(.58**)
6. WSC with local norms	.44**	.79**	.79**	.74**	.88**	—	(.77**)
7. CWS with local norms	.53**	.64**	.69**	.89**	.68**	.80**	—

Note. Data using male student information is presented in parentheses. ** $p < .01$.

Table 9

Results of Fisher z -Transformations to Examine Relationships between CBM-WE Scores and Teachers' Ratings Based on Gender

Scoring Metric	z -score	p
TWW		
National norms	0.16	0.873
Local norms	0.25	0.803
WSC		
National norms	-0.88	-0.27
Local norms	0.379	0.787
CWS		
National norms	0.51	-0.77
Local norms	0.610	0.441

List of Appendices

Appendix A: CBM-WE Probe

Appendix B: Teacher Judgments of Students' Writing Skills

Appendix C: Procedural Script for Session 1

Appendix D: Procedural Script for Session 2

Appendix B

Teacher Judgments of Students' Writing Skills

Please rank each student's overall writing skills relative to their classmates' overall writing skills:

Student 1	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 2	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 3	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 4	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 5	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 6	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 7	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 8	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 9	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 10	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 11	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 12	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 13	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 14	Much worse	Somewhat worse	About the same	Somewhat better	Much better
Student 15	Much worse	Somewhat worse	About the same	Somewhat better	Much better

Appendix D

Procedural Script for Session 1

18.	<p>State to the students:</p> <p><i>“You are going to write a story. First, I will read a sentence, and then you will write a story about what happens next. You will have 1 minute to think about what you will write, and 3 minutes to write your story. Remember to do your best work. If you don’t know how to spell a word, you should guess. If you make a mistake, just cross it out and keep writing. Are there any questions? (Pause). Put your pencils down and listen.</i></p> <p><i>“For the next minute, think about: <u>A boy was walking home when he found a \$100 bill on the sidewalk. The boy...</u>”</i></p>	
19.	<p>Start your stopwatch. During the planning period, ensure students are thinking, not writing.</p>	
20.	<p>At 30 seconds, state to the students:</p> <p><i>“You should be thinking about: <u>A boy was walking home when he found a \$100 bill on the sidewalk. The boy...</u>”</i></p>	
21.	<p>At 1 minute, state to the students:</p> <p><i>“Now begin writing.”</i></p>	
22.	<p>Start your stopwatch. Monitor students' participation. If individual students pause for about 10 seconds or say they are done before the test is finished, move close to them and say, <i>“Keep writing the best story you can.”</i> This prompt can be repeated to students should they pause again.</p>	
23.	<p>At 1:30 minutes (90 seconds), state to the students:</p> <p><i>“You should be writing about: <u>A boy was walking home when he found a \$100 bill on the sidewalk. The boy...</u>”</i></p>	
24.	<p>At 3 minutes, state to the students:</p> <p><i>“Stop writing and turn to the next page of your packet that says page #7 at the top corner.”</i> (Ensure students have stopped writing and are on the right page.)</p>	

Appendix E

Procedural Script for Session 2

6.	<p>State to the students:</p> <p><i>“You are going to write a story. First, I will read a sentence, and then you will write a story about what happens next. You will have 1 minute to think about what you will write, and 3 minutes to write your story. Remember to do your best work. If you don’t know how to spell a word, you should guess. If you make a mistake, just cross it out and keep writing. Are there any questions? (Pause). Put your pencils down and listen.</i></p> <p><i>“For the next minute, think about: <u>One day, Lily went to school but nobody was there except her. Lily...</u>”</i></p>	
7.	<p>Start your stopwatch. During the planning period, ensure students are thinking, not writing.</p>	
8.	<p>At 30 seconds, state to the students:</p> <p><i>“You should be thinking about: <u>One day, Lily went to school but nobody was there except her. Lily...</u>”</i></p>	
9.	<p>At 1 minute, state to the students:</p> <p><i>“Now begin writing.”</i></p>	
10.	<p>Start your stopwatch. Monitor students' participation. If individual students pause for about 10 seconds or say they are done before the test is finished, move close to them and say, <i>“Keep writing the best story you can.”</i> This prompt can be repeated to students should they pause again.</p>	
11.	<p>At 1:30 minutes (90 seconds), state to the students:</p> <p><i>“You should be writing about <u>One day, Lily went to school but nobody was there except her. Lily...</u>”</i></p>	
12.	<p>At 3 minutes, state to the students:</p> <p><i>“Stop writing and turn to the next page of your packet that says page #3 at the top corner.”</i> (Ensure students have stopped writing and are on the right page.)</p>	
19.	<p>State to the students:</p> <p><i>“You are going to write another story. Put your pencils down and listen. For the next minute, think about: <u>My friend was walking down the street when he saw something. My friend...</u>”</i></p>	

20.	Start your stopwatch. During the planning period, ensure students are thinking, not writing.	
21.	At 30 seconds, state to the students: <i><u>"You should be thinking about: <u>My friend was walking down the street when he saw something. My friend..."</u></u></i>	
22.	At 1 minute, state to the students: <i>"Now begin writing."</i>	
23.	Start your stopwatch. Monitor students' participation. If individual students pause for about 10 seconds or say they are done before the test is finished, move close to them and say, " <i>Keep writing the best story you can.</i> " This prompt can be repeated to students should they pause again.	
24.	At 1:30 minutes (90 seconds), state to the students: <i><u>"You should be writing about: <u>My friend was walking down the street when he saw something. My friend..."</u></u></i>	
25.	At 3 minutes, state to the students: <i>"Stop writing and turn to the next page of your packet that says page #7 at the top corner."</i> (Ensure students have stopped writing and are on the right page.)	

References

- Abbott, R. D., & Berninger, V. W. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary- and intermediate-grade writers. *Journal of Educational Psychology, 85*, 478–508. <https://doi-org.libezproxy2.syr.edu/10.1037/0022-0663.85.3.478>
- Addison, J., & McGee, S.J. (2010). Writing in high school/ writing in college: Research trends and future directions. *College Composition and Communication, 62*(1), 147-179.
- AIMSweb® (2018). *AIMSweb(R) National Norms Table*.
<https://aimswetraining.pearson.com/Report.cfm?inc=rReport18.cfm&ClearSchool=>
- Beard, R., & Burrell, A. (2010). Writing attainment in 9- to 11-year-olds: Some differences between females and males in two genres. *Language and Education, 24*(6), 495–515. <https://doi.org/10.1080/09500782.2010.502968>.
- Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' judgments of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly, 23*(1), 43–55. <https://doi.org/10.1037/1045-3830.23.1.43>
- Berninger, V. W. (2000). Development of language by hand and its connections with language by ear, mouth, and eye. *Topics in Language Disorders, 20*(4), 65–84. <https://doi-org.libezproxy2.syr.edu/10.1097/00011363-200020040-00007>
- Berninger, V. W., Vaughan, K. B., & Abbott, R. D. (1998). Early intervention for spelling problems: teaching functional spelling units of varying size with a multiple-connections framework. *Journal of Educational Psychology, 90*(4), 587–605. <https://doi-org.libezproxy2.syr.edu/10.1037/0022-0663.90.4.587>.

- Berninger, V. W., & Winn, W. D. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research*. (pp. 96–114). Guilford Press.
- Codding, R. S., Petscher, Y., & Truckenmiller, A. (2015). CBM reading, mathematics, and written expression at the secondary level: Examining latent composite relations among indices and unique predictions with a state achievement test. *Journal of Educational Psychology, 107*(2), 437–450. <https://doi-org.libezproxy2.syr.edu/10.1037/a0037520>
- Common Core State Standards Initiative. (2018). *Standards in Your State*.
<http://www.corestandards.org/standards-in-your-state/>
- Delacre, M., Leys, C., Mora, Y.L., & Laken, D. (2019). Taking parametric assumptions seriously: Arguments for the use of Welch’s F-test instead of the classical F-test in one-way ANOVA. *International Review of Social Psychology, 32*(1).
<https://doi.org/10.5334/irsp.198>
- Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*(3), 184–192. <https://doi-org.libezproxy2.syr.edu/10.1177/00224669030370030801>.
- Deno, S. L. (2015). Problem-solving assessment. In R. Brown-Chidsey & K. J. Andren (Eds.), *Assessment for intervention: A problem-solving approach., 2nd ed.* (pp. 10–36). Guilford Press.
- Eckert, T. L., & Arbolino, L. A. (2005). The role of teacher perspectives in diagnostic and program evaluation decision making. In R. Brown-Chidsey (Ed.), *Assessment for intervention: A problem-solving approach.* (pp. 65–81). Guilford Press.

- Eckert, T. L., Dunn, E. K., & Coddling, R. S. (2006). Assessment of mathematics and reading performance: An examination of the correspondence between direct assessment of student performance and teacher report. *Psychology in the Schools, 43*(3), 247–265. <https://doi-org.libezproxy2.syr.edu/10.1002/pits.20147>.
- Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology, 69*, 275-298. <https://doi.org/10.1146/annurev-psych-122216-011719>
- Espin, C., Shin, J., & Deno, S. L. (2000). Identifying indicators of written expression proficiency for middle school students. *Journal of Special Education, 34*(3), 140–153. <https://doi-org.libezproxy2.syr.edu/10.1177/002246690003400303>
- Farrington, J. Y., Parker, P. D., Kidder-Ashley, P., Gagnon, S. G., McCane-Bowling, S., & Sorrell, C. A. (2014). Gender differences in written expression curriculum-based measurement in third- through eighth-grade students. *Psychology in the Schools, 51*(1), 85–96. <https://doi.org/10.1002/pits.21733>
- Fuchs, L. S., & Deno, S. L. (1982). Developing goals and objectives for educational programs. Institute for Research in Learning Disabilities.
- Gansle, K. A., Noell, G. H., VanDerHeyden, A. M., Naquin, G. M., & Slider, N. J. (2002). Moving beyond total words written: The reliability, criterion validity, and time cost of alternate measures for curriculum-based measurement in writing. *School Psychology Review, 31*(4), 477–497. <https://doi.org/10.1080/02796015.2002.12086169>.
- Habedank-Stewart, L., & Kaminski, R. (2002). Best Practices in Developing Local Norms for Academic Problem Solving. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV, 1*(2), 737–752. National Association of School Psychologists.

- Harris, K. R., & Graham, S. (1992). Self-regulated strategy development: A part of the writing process. In M. Pressley, K. R. Harris, & J. T. Guthrie (Eds.), *Promoting Academic Competence and Literacy in School*. (pp. 277–309). Academic Press.
- Hartman, J. M., & Fuller, M. L. (1997). The development of curriculum-based measurement norms in literature-based classrooms. *Journal of School Psychology, 35*, 377–389.
[https://doi-org.libezproxy2.syr.edu/10.1016/S0022-4405\(97\)00013-7](https://doi-org.libezproxy2.syr.edu/10.1016/S0022-4405(97)00013-7)
- Hayes, J. R., & Flower, L. S. (1980). Writing as problem solving. *Visible Language, 14*(4), 388–399.
- Hayes, J. R., & Flower, L. S. (1986). Writing research and the writer. *American Psychologist, 41*(10), 1106–1113. <https://doi-org.libezproxy2.syr.edu/10.1037/0003-066X.41.10.1106>
- Hier, B. O., Datchuk, S. M., & Watts, E. A. (2021). Accounting for Levels of Language in Narrative and Expository Writing: A Skills Analysis of Second-Grade Student Writing. *The Elementary School Journal, 121*(4), 541-560.
- Hier, B. O., Eckert, T. L., & Viney, E. A. (2019). Generalization and maintenance effects of writing fluency intervention strategies for elementary-age students: A randomized controlled trial. *School Psychology Review, 48*(4), 377–382. <https://doi.org/10.17105/spr-2017-0123.v48-4>
- IBM Corp. (2017). IBM SPSS Statistics for Windows. Armonk, NY: IBM Corp. Retrieved from <https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-26>
- Jewell, J., & Malecki, C. (2005). The utility of CBM written language indices: An investigation of production-dependent, production-independent, and accurate-production scores. *School Psychology Review, 34*, 27–44. <https://doi.org/10.1080/02796015.2005.12086273>

- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80(4), 437–447.
<https://doi.org/10.1037/0022-0663.80.4.437>
- Keller-Margulis, M. A., Mercer, S. H., Payan, A., & McGee, W. (2015). Measuring Annual Growth Using Written Expression Curriculum-Based Measurement: An Examination of Seasonal and Gender Differences. *School Psychology Quarterly*, 30(2), 276–288.
<https://doi.org/10.1037/spq0000086>
- Kim, Y.-S., Otaiba, S. A., Puranik, C., Folsom, J. S., Greulich, L., & Wagner, R. K. (2011). Componential skills of beginning writing: An exploratory study. *Learning and Individual Differences*, 21(5), 517–525. <https://doi-org.libezproxy2.syr.edu/10.1016/j.lindif.2011.06.004>.
- Kim, Y-S., Otaiba, S. A., Wanzek, J., & Gatlin, B. (2015). Toward an understanding of dimensions, predictors, and the gender gap in written composition. *Journal of Educational Psychology*, 107(1), 79–95. <https://doi-org.libezproxy2.syr.edu/10.1037/a0037210>.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Lowry, R. (2001). *Significance of the difference between two correlation coefficients* [Online calculator]. <http://vassarstats.net/rdiff.html>.
- Malecki, C. K., & Jewell, J. (2003). Developmental, gender, and practical considerations in scoring curriculum-based measurement writing probes. *Psychology in the Schools*, 40(4), 379–390. <https://doi.org/10.1002/pits.10096>

- McCutchen, D. (1986). Domain knowledge and linguistic knowledge in the development of writing ability. *Journal of Memory and Language*, 25(4), 431–444. [https://doi-org.libezproxy2.syr.edu/10.1016/0749-596X\(86\)90036-7](https://doi-org.libezproxy2.syr.edu/10.1016/0749-596X(86)90036-7).
- McFarland, M., & Wallace, D. L. (2008). Assessment of written expression by curriculum-based measurement and teachers' rating of identified critical quality writing components. *International Journal of Psychology: A Biopsychosocial Approach / Tarptautinis Psichologijos Žurnalas: Biopsichosocialinis Požiūris*, 2, 99–119.
- McMaster, K. L., Du, X., & Petursdottir, A.-L. (2009). Technical features of curriculum-based measures for beginning writers. *Journal of Learning Disabilities*, 42(1), 41–60. <https://doi.org/10.1177/0022219408326212>
- McMaster, K.L., Shin, J., Espin, C.A., Jung, P.G., Wayman, M.M., & Deno, S.L. (2017). Monitoring elementary students' writing progress using curriculum-based measures: Grade and gender differences. *Reading and Writing: An Interdisciplinary Journal*, 30(9), 2069-2091. <https://doi.org/10.1007/s11145-017-9766-9>
- McMaster, K. L., Wayman, M. M., Deno, S. L., Espin, C. A., & Yeo, S. (2010). *Examining technical features of progress monitoring measures across grade levels in writing*. Research Institute on Progress Monitoring. <https://www.progressmonitoring.org/pdf/tr38pmswriting.pdf>
- Mowrey, S. C., & Farran, D. C. (2016). Performance and preparation: Alignment between student achievement, teacher ratings, and parent perceptions in urban middle-grades mathematics classrooms. *Journal of Urban Learning, Teaching, and Research*, 12, 61–74.

- National Center for Education Statistics (2012). *The nation's report card: Writing 2011 (NCES 2012–470)*. Institute of Education Sciences, U.S. Department of Education.
<https://nces.ed.gov/nationsreportcard/pdf/main2011/2012470.pdf>
- National Commission on Writing. (2004). *Writing: A Ticket to work... Or a ticket out: A Survey of business leaders*. College Board. https://nwp.org/cs/public/download/nwp_file/21479/writing-a-ticker-work-or-a-ticket-out.pdf?x-r=pcfile_d
- National Commission on Writing. (2005). *Writing: A powerful message from state government*. College Board. http://www.collegeboard.com/prod_downloads/writingcom/powerful-message-from-state.pdf.
- Olinghouse, N.G., & Leaird, J.T. (2008) The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading and Writing*, 2, 545–565 (2009). <https://doi.org/10.1007/s11145-008-9124-z>
- Persky, H. R., Daane, M. C., & Jin, Y. (2003). *The Nation's report card writing 2002 (NCES 2003– 529)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
<https://nces.ed.gov/nationsreportcard/pdf/main2002/2003529.pdf>
- Peterson, S. S., & Kennedy, K. (2006). Sixth-grade teachers' written comments on student writing: Genre and gender influences. *Written Communication*, 23(1), 36–62. <https://doi.org/10.1177/0741088305282762>
- Powell-Smith, K. A. & Shinn, M. R. (2004). *Administration and Scoring of Written Expression Curriculum-Based Measurement (WE-CBM) for Use in General Outcome Measurement: Training Handbook*. Pearson.

- Randall, A. (2013). Effects of teacher efficacy and student's gender and ethnicity on special education referral and response to intervention. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 73(9).
- Ray, A., Graham, S., Houston, J., & Harris, K. (2016). Teachers use of writing to support students' learning in middle school: A national survey in the United States. *Reading and Writing*, 29 (5), 1039-1068. <https://doi.org/10.1007/s11145-015-9602-z>
- Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist*, 74(4), 445–458. <https://doi-org.libezproxy2.syr.edu/10.1037/amp0000356.supp>
- Ritchey, K. D., & Coker, D. L. (2013). An investigation of the validity and utility of two curriculum-based measurement writing tasks. *Reading & Writing Quarterly*, 29(1), 89–119. <https://doi-org.libezproxy2.syr.edu/10.1080/10573569.2013.741957>
- Romig, J. E., Therrien, W. J., & Lloyd, J. W. (2017). Meta-analysis of criterion validity for curriculum-based measurement in written language. *Journal of Special Education*, 51(2), 72–82. <https://doi.org/10.1177/0022466916670637>
- Romig, J. E., Miller, A. A., Therrien, W. J., & Lloyd, J. W. (2020). Meta-analysis of prompt and duration for curriculum-based measurement of written language. *Exceptionality*. <https://doi-org.libezproxy2.syr.edu/10.1080/09362835.2020.1743706>
- Sandberg Patton, K. L., Reschly, A. L., & Appleton, J. (2014). Curriculum-based measurement as a predictor of performance on a state assessment: Diagnostic efficiency of local norms. *Educational Assessment*, 19(4), 284–301. <https://doi.org/10.1080/10627197.2014.964117>

- Shanahan, T. (2015). Common core state standards. *Elementary School Journal*, 115(4), 464–479. <https://doi-org.libezproxy2.syr.edu/10.1086/681130>
- Shinn, M. R. (2005). Identifying and validating academic problems in a problem-solving model. In Brown-Chidsey, R. (Ed). *Assessment for Intervention: A Problem-Solving Approach*. Guilford Publications.
- Südkamp, A., Kaiser, J., & Moller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>
- Swain, K. D., & Hagaman, J. L. (2020). Elementary special education teachers' use of CBM data: A 20-year follow-up. *Preventing School Failure: Alternative Education for Children and Youth*, 64(1), 48-54. <https://doi.org/10.1080/1045988x.2019.1678009>
- Swanson, H. L., & Berninger, V. W. (1996). Individual differences in children's working memory and writing skill. *Journal of Experimental Child Psychology*, 63, 358–385. <https://doi-org.libezproxy2.syr.edu/10.1006/jecp.1996.0054>
- Swartz, C. W., Hooper, S. R., & Montgomery, J. W. (1999). Using generalizability theory to estimate the reliability of writing scores derived from holistic and analytical scoring methods. *Educational & Psychological Measurement*, 59(3), 492–506. <https://doi-org.libezproxy2.syr.edu/10.1177/00131649921970008>.
- Troia, G. A., Harbaugh, A. G., Shankland, R. K., Wolbers, K. A., & Lawrence, A. M. (2013). Relationships between writing motivation, writing activity, and writing performance: Effects of grade, sex, and ability. *Reading and Writing*, 26(1), 17-44.

- Videen, J., Deno, S., Marston, D., & Institute for Research on Learning Disabilities University of Minnesota. (1982). *Correct Word Sequences: A Valid Indicator of Proficiency in Written Expression*. <https://files.eric.ed.gov/fulltext/ED225112.pdf>
- Vierra, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, *37*(5), 360-363.
- Wagner, D. L., Hammerschmidt, S. S. M., Espin, C. A., Seifert, K., & McMaster, K. L. (2017). Pre-service teachers' interpretation of CBM progress monitoring data. *Learning Disabilities Research & Practice*, *32*(1), 22–31. <https://doi-org.libezproxy2.syr.edu/10.1111/ldrp.12125>
- Wang, E., Matsumura, L.C., & Correnti, R. (2018) Student writing accepted as high-quality responses to analytic text-based writing tasks. *The Elementary School Journal*, *118*(3), 357-383. <https://doi.org/10.1086/696097>
- Watkins, M. W., & Pacheco, M. (2000). Interobserver agreement in behavioral research: importance and calculation. *Journal of Behavioral Education*, *10*(4), 205–212.
- Weiss, S. L., Brinkley, J., & Bock, J. (2019). Patterns of growth: Cluster analysis of written expression curriculum-based measurement. *Education and Treatment of Children*, *42*(2), 161–184. <https://doi.org/10.1353/etc.2019.0008>
- Ysseldyke, J. E., Algozzine, B., Richey, L., & Graden, J. (1982). Declaring students eligible for learning disability services: Why bother with the data? *Learning Disability Quarterly*, *5*(1), 37–44. <https://doi-org.libezproxy2.syr.edu/10.2307/1510614>.

Emily Watts, MS

230 W Willow Street Apt 104 Syracuse, NY 13202
(716) 472-0912
eawatts@syr.edu

EDUCATION

Syracuse University

PhD, School Psychology: Anticipated Graduation Date of May 2024

University at Buffalo, State University of New York

PhD Program, Counseling Psychology and School Psychology: Invited to transfer to Syracuse University

Rochester Institute of Technology

MS, School Psychology: Graduated May 2010

University at Buffalo, State University of New York

BA, Psychology: Graduated May 2005

PROFESSIONAL EXPERIENCE

Psychological Services Center

08/22- present

Title: Doctoral Student Clinician

Supervisor: Afton Kapuscinski, Ph.D.

- Completed treatment intakes and assessments in both virtual and in-person settings
- Provided individualized counseling for children, adolescents, and adults
- Developed intake/assessment reports
- Developed case conceptualizations for clients
- Completed necessary paperwork to ensure billing is completed in a timely manner
- Engaged in individual and group supervision
- Provided parents with consultation and psychoeducation
- Facilitated psychoeducation groups for parents

Access to Psychiatric Care Program

08/16-04/19

Title: Behavioral Consultant

Supervisor: Janelle Van Cleve, MS

- Completed face to face intake and family assessments
- Developed functional behavioral assessments
- Developed and implemented behavioral support plans
- Provided the individual and family with training on the behavioral support plans
- Provided individual and families with short term counseling utilizing solution focused and cognitive behavioral therapies
- Collaborated with the providers of psychiatric care to ensure appropriate program delivery

People Inc.

08/10-02/16

Title: Community Based Behavioral Health Specialist

Supervisor: Shannon Bianco, MS

- Completed face to face intake and family assessments
- Developed functional behavioral assessments
- Developed and implemented behavioral support plans
- Provided the individual and family with training on the behavioral support plan

- Provided individual and group with counseling utilizing solution focused and cognitive behavioral therapies
- Supervised and trained behavior technicians on behavioral support plans
- Collaborated with schools, programs, and respite to ensure consistency for individuals
- Completed necessary paperwork to ensure billing is completed in a timely manner
- Provided family education trainings
- Facilitated psycho-educational groups for parents such as Incredible Years parent training and skills enhancement groups
- Facilitated play therapy groups for children utilizing a developmental individual-difference relationship-based model
- Completed training and implemented several therapeutic interventions such as progressive counting, schema therapy, cognitive processing therapy
- Completed cognitive, academic, adaptive, social/emotional, and autism assessments for eligibility

FIELD EXPERIENCE

Meachum Elementary School, Syracuse NY 10/21 – 06/22

Title: Doctoral-level Practicum Student

Supervisor: Michael Gilbert, Psy.D

- Conducted psychological evaluations for triannual reviews
- Conducted functional behavioral assessment
- Developed and trained others on behavior intervention plans
- Presented and participated in Committee on Special Education meetings
- Consulted with teachers on tier-one, tier-two, and tier-three interventions
- Provided tier-two writing intervention sessions
- Provided tier-two social-emotional interventions directly to students

Liverpool Central School District, Liverpool NY 12/06/2019

Title: Professional Development Seminar Coach

Supervisor: Bridget O. Hier, PhD

- Participated in planning district teacher assistants' professional development on behavioral strategies
- Provided implementation planning consultation to teacher assistants during seminar
- Collected and entered feedback data from teacher assistants

Cleveland Hill Middle School, Cheektowaga, NY 09/18 – 05/19

Title: Doctoral-level Practicum Student

Supervisor: Lisa Hillyard, MA, NCSP

- Conducted psychological evaluations for triannual reviews at the middle school level
- Participated in student support meetings for behavioral referrals at the tier 2 and 3 level
- Provided individual counseling to students
- Facilitated skills-based counseling group to students
- Provided school-wide positive behavioral interventions and supports to students
- Developed functional behavioral assessments
- Provided consultation to teachers regarding behavioral strategies
- Participated in team building activities for staff
- Participated in trauma informed care trainings for staff
- Co-facilitated bullying awareness programs for students
- Participated in Committee on Special Education meetings

Williamsville Central School District, Williamsville, NY 08/17 – 11/17

Title: Doctoral Level Practicum Student for Universal Pre-Kindergarten Screening,

Supervisors: Rebecca Vujnovic, Ph.D., Linda Dugan, M.A.

- Received formal training in administration and scoring of the Brigance Inventory of Early Development III
- Administered and scored selected subtests of the Brigance Inventory of Early Development III assessment across all sites for all students enrolled in the district's UPK program

Lockport City School District, Lockport, NY

09/09 – 06/10

Title: Master-level School Psychology Intern

Supervisor: Lynn Witt, MA- DeWitt Clinton Elementary School

Supervisor: Michele Westgate, MA- Lockport Opportunity Projects

- Conducted psychological evaluation for initial and re-evaluation referrals at elementary, middle and high school levels
- Responsible for all psychological evaluations at Niagara County Jail for the Incarcerated Youth Program
- Independently facilitated counseling groups for kindergarten, fifth grade, and sixth grade students
- Co-facilitated counseling groups for first, third, and fourth grade students
- Participated and presented results at Committee on Special Education meetings
- Provided individual counseling to four high school students
- Provided crisis and grief counseling to elementary and high school students
- Developed functional behavioral assessments
- Developed and implemented behavior intervention plans
- Provided consultation to teachers and parents regarding behavioral strategies and plans
- Provided push-in lessons for general education and self-contained classrooms at the elementary school level
- Participated in School Based Support Teams
- Consulted with a Behavior Specialist from Aspire Services to effectively work with students with Autism

Kenmore East Senior High School, Tonawanda, NY

12/08 – 05/09

Title: Master-level School Psychology Practicum Student

Supervisor: Jeffery Spriegel, MA

- Conducted psychological evaluation for re-evaluation referrals
- Wrote psychological reports for evaluations
- Participated in Data Solving Team
- Attended Committee on Special Education meetings
- Provided consultation to teachers and parents regarding behavioral strategies and plans
- Developed functional behavioral assessments

Sweet Home Middle School, Amherst, NY

03/08 – 11/08

Title: Master-level School Psychology Practicum Student

Supervisor: Robin Raphael, MA

- Conducted psychological evaluations for re-evaluation referrals
- Wrote psychological reports for evaluations
- Co-facilitated a social skills group for seventh-grade girls
- Performed curriculum-based assessment in reading and math for students referred for academic difficulties
- Participated in Data Solving Team
- Provided individual counseling to seventh-grade student

Paul Road Elementary School, Gates-Chili, NY

09/07 – 02/08

Title: Master-level School Psychology Practicum Student

Supervisor, Emily Yates, MA

- Performed record reviews
- Conducted classroom observations

- Developed functional behavioral assessments
- Attended Student Improvement Meetings
- Co-facilitated a social skills group for third-grade students

TEACHING EXPERIENCE

Syracuse University, Syracuse, NY 08/20- 12/21

Title: Instructor

- Independently taught PSY205: Foundations of Human Behavior and PSY209: Foundations of Human Behavior- Honors Section
- Provided teaching in a hybrid, online-only, and in-person formats
- Developed course syllabus, course goals, and course structure
- Developed and provided hybrid lectures
- Developed and graded weekly activities, assignments, and quizzes

Syracuse University Teaching Mentor Program, Syracuse, NY 05/20 – 08/20

Title: Teaching Mentor Alternate

Advisors: Shawn Loner and Glenn Wright, PhD

- Developed and collaborated with team members on training sessions for all incoming Teaching Assistants
- Developed pre- and post-session assessments for incoming Teaching Assistants
- Developed online trainings for all incoming Teaching Assistants

Syracuse University, Syracuse, NY 05/20- 06/20

Title: Instructor

- Independently taught PSY336: Psychology of Adolescence asynchronous online course
- Developed course syllabus, course goals, and course structure
- Developed and provided online lectures
- Developed and graded weekly activities, assignments, and quizzes

Syracuse University, Syracuse, NY 08/19- 06-20, 01/22-

05/22

Title: Teaching Assistant

Advisors: Shannon Houck, PhD and Meredith Martin, PhD

- Assisted with PSY205: Foundations of Human Behavior
- Developed and taught recitations to undergraduate students
- Attended weekly teaching development meetings
- Developed and graded weekly quizzes, and graded multiple choice and essay quizzes administered to students
- Proctored lecture exams
- Provided weekly office hours to undergraduate students
- Engaged in peer reviews of other teaching assistants
- Led peer review group as an experienced teaching assistant

University at Buffalo, Buffalo, NY 08/18- 05/19

Title: Teaching Assistant

Advisor: Rebecca Vujnovic, PhD

- Assisted with the following courses: CEP590: Psychological Measurement and Intellectual Assessment; CEP590B: Adult Cognitive Assessment
- Taught lectures to graduate level students
- Provided training and supervision to graduate level students on assessment measures
- Helped to coordinate universal preschool assessments for local school district
- Provided direct observation and feedback to graduate students performing preschool assessments

- Reviewed and graded assessment protocols submitted by graduate level students

University at Buffalo, Buffalo, NY

08/17-05/18

Title: Teaching Assistant

Advisor: Michele Shanahan, PhD

- Assisted with the following courses: CEP400: Educational Psychology; CEP500: Psychological Foundations of Education; CEP560: Psychological Learning and Instruction
- Graded tests and homework submitted by students

RESEARCH EXPERIENCE

Syracuse University CARE Research Lab, Syracuse, NY

03/22- present

Title: Graduate Research Assistant

Advisor: Natalie Russo, PhD

Worked on the following projects:

- Conducted cognitive assessments for research participants using the Stanford-Binet Intelligence Scale- Fifth Edition
- Wrote cognitive and adaptive functioning sections for autism diagnostic reports

Syracuse University RITES Research Lab, Syracuse, NY

08/19- 02/22

Title: Graduate Research Assistant and Scoring Coordinator

Advisor: Bridget Hier, PhD

Worked on the following projects: Writing Quality Systematic Review, Effects of Choice on Second-Grade Students' Writing Fluency, and Reliability and Validity of Direct Behavior Ratings for Writing Quality

- Conducted systematic literature reviews
- Assisted with the development of research manuals for coding writing elements
- Coordinated tasks for graduate and undergraduate research assistants
- Provided weekly feedback meetings to graduate and undergraduate research assistants to gain proficiency and reliability in scoring data with writing measures
- Scored data with writing CBM and standardized measures
- Provided training to research assistants to gain proficiency in fidelity measures

University at Buffalo RITES Research Lab, Buffalo, NY

08/17- 05/19

Title: Graduate Research Assistant

Advisor: Bridget Hier, PhD

Worked on the following projects: Examining Sufficient Exemplars, Identifying Cognitive Mechanisms Behind a Successful Performance Feedback Intervention, Writing Errors Project, Examining Sequence of Fluency, An Examination of the Relationship Between Teachers' Perceptions of Student Writing and Students' Performances on Curriculum-Based Measurement-Written Expression and Generalization Interventions for Writing

- Performed literature reviews
- Assisted with the development of research manuals for coding writing elements
- Coordinated tasks for undergraduate research assistants
- Scored, entered, and checked data with writing CBM and standardized measures
- Provided research reliable intervention delivery
- Provided training to research assistants to gain proficiency in fidelity measures
- Provided training to research assistants on behavior strategies
- Assisted with writing and editing manuscript currently in review

R.I.T. School Psychology Graduate Office, Rochester, NY)

03/08- 05/09

Title: Research Assistant

Advisor: Vincent Pandolfi, PhD

- Conducted literature reviews on school psychologists' training for working with students with autism
- Organized, coded, and entered data

University at Buffalo Research Lab, Buffalo, NY

04/07- 08/09

Title: Research Assistant

Advisor: Craig Colder, PhD

Worked on the Adolescent and Family Development Project

- Conducted structured interviews with parents and adolescents
- Screened and scheduled participants
- Ran inhibitory tasks with participants
- Conducted basic achievement tests with adolescents

University at Buffalo Research Lab, Buffalo, NY

01/04-05/06

Title: Senior Research Assistant

Advisor: Jennifer, Read PhD

Worked on the following projects: Effects of Mood Manipulation on Alcohol Expectancies, Relations Among Gender, Affect, Alcohol Expectancies and Alcohol Consumption, Life Experiences and Alcohol Beliefs in College Students, and Stress Effects on Alcohol Expectancy and Urge

- Conducted physiological and questionnaire studies with participants
- Conducted timeline follow back interview sessions with participants assessing drinking patterns over previous three months
- Conducted mood manipulation on participants

GRADUATE ASSISTANT EXPERIENCE

R.I.T. School Psychology Graduate Office, Rochester, NY

09/08- 05/09

Title: Protocol Checker

Supervisor: Jennifer Lukomski, PhD

- Review psycho-educational protocols submitted by first-year graduate students
- Assisted first-year graduate students in administering and scoring tests

PUBLICATIONS

Datchuk, S. M., Hier, B. O., & **Watts, E. A.** (2021). Accounting for levels of language in narrative and expository writing: A skills analysis of second-grade student writing. *The Elementary School Journal*, 121(4), 541- 560. <https://doi.org/10.1086/714051>.

CONFERENCE PRESENTATIONS

Hier, B. O., Eckert, T. L., Hitchings, T. J., Circe, J. J., **Watts, E.** Finelli, C. C., Maguire, S., & Nelson, K. (2022, February). *Writing quality assessment: Preliminary results of a systematic review*. Poster presented at the Annual Conference of the National Association of School Psychologists, Boston, MA.

Hier, B. O., Datchuk, S. M., Eckert, T., **Watts, E.**, Circe, J. J., Hitchings, T. J., Goldstein, A., Finelli, C. C., Maguire, S., & Nelson, K. (2021, February). *Validity and classification accuracy of curriculum-based measurement of sentence construction*. Poster presented at the Annual Convention of the National Association of School Psychologists; Virtual Conference.

Hitchings, T. J., Hier, B. O., **Watts, E.**, & Finelli, C. C. (2021, February). *The Relationship Between Second-Grade Students' Off-Task Behavior and Text Generation*. Poster presented at the Annual Convention of the National Association of School Psychologists; Virtual Convention.

- Eckert, T. L., Hier, B. O., Williams, N. L., Circe, J. J., Hitchings, T. J., **Watts, E.**, & Finelli, C. C., & Ardoin, S. P. (2021, February 23-26). *Advance in writing research: Assessment, Intervention, and Acceptability*. [Paper presentation]. National Association of School Psychologists, Virtual Convention.
- Hier, B.O., Hitchings, T.J., Ardoin, S.P., Goldstein, A.R., & **Watts, E.** (2020, February). *The effect of topic choice on second-grade students' writing performances*. Poster presented at the Annual Conference of the National Association of School Psychologists, Baltimore, MD.
- Watts, E.**, Hier, B.O., Kesselring, S.A., Dolan, L., Tymchak, K., Finelli, C., Krajnik, J., & DeMarco, A. (2019, April). *Teachers' perceptions of student writing*. Poster Presented at the annual symposium of Graduate School of Education, Buffalo, NY.
- Watts, E.**, Finelli, C., Farinelli, A., Tymchak, K., & Fink, N. (April, 2018). *The role of gender and race in elementary students' acceptability ratings of a writing intervention*. Presentation at the annual symposium of Graduate School of Education, Buffalo, NY.
- Watts, E.**, Hier, B.O., Kesselring, S.A., Dolan, L., Tymchak, K., Finelli, C., Krajnik, J., & DeMarco, A. (2019, February). *Teachers' perceptions of student writing*. Poster Presented at the Annual Conference of the National Association of School Psychologists, Atlanta, GA.
- Hier, B.O., **Watts, E.**, & Tymchak, K. (2019, February). *An examination of common sentence-level writing errors among second-grade students*. Paper Presented at the Annual Conference of the National Association of School Psychologists, Atlanta, GA.
- Tymchak, K., & **Watts, E.** (April, 2018). *Examination of writing errors made by elementary students*. Presentation at the annual symposium of Graduate School of Education, Buffalo, NY.
- Watts, E.** (October, 2012). *Alternative career pathways for school psychologists*. Panel presentation at the annual meeting of the New York Association of School Psychologists, Niagara Falls, NY.

GUEST LECTURES

- Watts, E.** (April 2022). *Psychopathy*. Lecture delivered to PSY474 Forensic Psychology class at Syracuse University, Syracuse, NY.
- Watts, E.**, Stanford, S., & Hitchings, T. (April 2019). *Intelligence testing and neuropsychological assessment*. Lecture delivered to CEP503 Tests and Measurement class at University at Buffalo, Buffalo, NY.
- Watts, E.** (March 2019). *Introduction to Academic Achievement Testing*. Lecture delivered to CEP592 Psychological Assessment to Children and Adolescents class at University at Buffalo, Buffalo, NY.
- Watts, E.**, Button, A., & Ward, L.E. (April 2018). *Intelligence testing and neuropsychological assessment*. Lecture delivered to CEP503 Tests and Measurement class at University at Buffalo, Buffalo, NY.

PROFESSIONAL AWARDS AND ACCOMPLISHMENTS

Article of the Year from the Iowa Academy of Education	2021
Outstanding Teaching Assistant Award at Syracuse University	2021
Future Professoriate Program Certificate	2020
Outstanding Poster Award at the UB Graduate School of Education Symposium	2019

PROFESSIONAL MEMBERSHIP

Student member of Association for Positive Behavior Support	2019- present
Student member of Women in Science and Engineering	2019- 2021
Student member of American Psychological Association Division 16	2018- present
Student member of National Association of School Psychologists	2017- present

PROFESSIONAL SERVICE

Peer Facilitator of Women in Science and Engineering	2021- 2023
Student member of Professional Development Committee	2021- 2022
Co-President of Psychology Action Committee	2020- 2021
Student member of the Diversity, Equity, and Inclusion Committee	2020- 2022
Student member of Diversity and Communications Committee	2019- 2020
Member of Graduate School of Education 2019 Symposium Planning Committee	2019
Student member of Graduate School of Education Executive Committee	2018- 2019
Student leader of Committee for Social Justice and Inclusion	2017- 2019
Student member of Refugee Committee	2017- 2018