AN INVESTIGATION OF THE PSYCHOMETRIC PROPERTIES OF THE GLOBAL ASSESSMENT OF SCHOOL FUNCTIONING

BY

JOSEPH D. PALAMARA

A DISSERTATION SUBMITTED TO THE FACULTY OF

ALFRED UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PSYCHOLOGY

IN

SCHOOL PSYCHOLOGY

DR. MARK FUGATE

ADVISOR

DIVISION OF SCHOOL PSYCHOLOGY

ALFRED, NY

APRIL, 2015

AN INITIAL INVESTIGATION OF THE PSYCHOMETRIC PROPERTIES OF A GLOBAL ASSESSMENT MEASURE OF STUDENT FUNCTIONING:

BY

JOSEPH D. PALAMARA

EASTERN MICHIGAN UNIVERSITY B.S. (1991)

EASTERN MICHIGAN UNIVERSITY M.S. (2000)

ALFRED UNIVERSITY M.A. (2010)

ALFRED UNIVERSITY C.A.S. (2015)

APPROVED BY: Mark Fugate, Ph.D.
Committee Chairperson

Nancy Evangelista, Ph.D.
Committee Member

Gordon Atlas, Ph.D.
Committee Member

Hannah Young, Psy.D.
Committee Member

Arthur Greil, Ph.D.
Committee Member

ACCEPTED BY Mark Fugate, Ph.D.
Chairperson, Division of School Psychology

ACCEPTED BY Nancy Evangelista, Ph.D.
Associate Provost and Dean, College of Professional Studies

ACCEPTED BY W. Richard Stephens, Jr., Ph.D.
Provost & Vice President for Academic Affairs

DEDICATION

It takes a village…

I borrow from this African proverb frequently; in fact, it is a running joke in our family and circle of friends regarding to the effort that goes into coaxing out the best in Joe. It takes a village – to raise a child…to be a parent…to affect change…to complete a dissertation. This work is dedicated to my family, the Palamaras and Cables, "passed", present, and future. I guess it's true – hard work is its own reward. Thank you all for your encouragement and belief in me and for trusting "my process"!

"Never let the fear of striking out get in your way."

Babe Ruth

ACKNOWLDEGMENTS

I would like to take this opportunity to thank the support staff, administrators, and teachers that comprise the Alfred University family. To the Counseling and School Psychology Department, I am especially grateful for your patience and understanding in dealing with my unique family and geographic considerations.

To my committee members, Drs. Atlas, Evangelista, and Young, you are truly the "Gold Standard" of all committee members. Your thoughtfulness and queries were both pointed and provocative, leading me to explore avenues that truly rounded out my manuscript. Dr. E., I hope you will allow yourself a little prideful smile as your instruction in Exceptionality and clinic is woven throughout the fabric of this dissertation.

To my honorary committee member, Dr. Greil, I can't thank you enough for your instruction, patience, and endurance! Not only are you my "Gold Standard" stats teacher, you are the Brooks Robinson of stats as far as I'm concerned (with the patience and endurance of Cal Ripkin, Jr.). It had to be exhausting, but please know you not only taught me applications, you gave me confidence to complete.

To my committee chairperson, Dr. Fugate, thank you so much for your time and "gentle nudging" to get me to complete this endeavor. I am so thankful for your ability to help me synthesize my big ideas into practical, researchable components. Your expertise with Response to Intervention was critical for crafting a methodology that helped bring utility to the GASF among school populations. What am I going to do with my Thursday mornings now?

I would also like to thank my colleague, partner, and friend Dr. Arthur Maerlender – this is all your fault! Without you this project truly does not exist. What a great idea! I can't thank you enough for permitting me to initiate research on the GASF. Who knew when we were

sitting across the table in consultation at the school that this is where the conversation would lead? I look forward to seeing how much farther this ball will roll.

Mom, I know you are proud, but please know that I am equally proud of you. You are a woman of many talents, and I know you take those talents for granted, but I don't. Your help with the house and the kids, and yes, the meals, were more than just pitching in, they were a gift. I am so thankful that you are sharing your "Golden Years" with us.

To my family, Dominic and Mary, you guys are rock stars! There aren't two kids on the planet that are more loved by their very proud parents. Thank you both for the very fun work breaks and for providing me with perspective throughout this process. Truth be told, you both made this process such a challenge – and I mean that in the best of ways! I so wanted to be hanging with you guys, watching you skate, watching you compete. You are my pride and joy.

Jen. How do you thank your best friend? I don't know what you saw, but I am sure glad you did. I don't know what the job description for a Saint looks like, but you're overqualified. You have afforded me (literally) the opportunity to tackle this program on my terms (and terms and terms and terms), and for that I am very grateful. I guess I will just get you a really big fruit basket!

Abstract

Schools are increasingly held accountable for student academic and behavioral performance, and showing efficacy of these treatment efforts. The primary metric for reporting academic progress, state endorsed standardized tests, does not take into account or effectively measure discrete skills or behavioral improvement. This necessitates the development of tools efficient in quantifying students' school-based behaviors. Mental health practitioners achieve this metric utilizing the Global Assessment of Functioning (GAF). The Global Assessment of School Functioning (GASF) is being developed to be an efficient scale used by teachers for similar means. The aim of the present study is to examine the utility of the GASF in capturing overall school functioning. This study was broken into two phases. Teacher consultants assessed content validity and validated vignettes that would be used to assess inter-rater reliability. School personnel then rated five vignettes using the GASF and responded to questions regarding their perceptions of the instrument. Correlational statistics suggested that school personnel were able to rate vignettes with substantial reliability (.877). Responses to questions relating to the raters competency and training and the raters overall impressions of the technical quality of the GASF were positive. The culminating analysis from the data presented in this study suggest that the GASF warrants further study to determine its technical properties and utility as a rating scale that school personnel can use to benchmark and progress monitor student behavior.

Tables and Figures

Page

Table of Contents

Chapter 1: Introduction

Policy changes in public education that emphasize greater school accountability in students' academic and behavioral performance have led to a paradigm shift in how schools provide instruction and intervention to students and have intensified procedures for how schools monitor and report progress to federal agencies. The signing of the No Child Left Behind Act of 2002 requires that schools demonstrate improvement through the use of state mandated assessments in reading, math, science, and social studies (United States Department of Education, 2002). As a result, school personnel – specifically teachers, find themselves more accountable than ever for the progress of their students.

In an effort to improve both academic and behavioral outcomes for children while attempting to reduce the number of students referred for special education services, many Local Education Agencies (LEAs) have implemented models based on prevention science espoused by mental health agencies. These models utilize assessment at the primary, secondary, and tertiary levels as a means of identifying students who may be "at-risk" of academic and/or behavioral problems and those who may require more intensive support. The current reauthorization of the Individuals with Disabilities Education Act provides for the use of these models within a Response to Intervention (RtI) framework (Bonner & Barnett, 2004; Gresham et al., 2004).

RtI is a model used to manage the performance of all children. RtI can be defined as a scientific process for identifying, operationalizing, and mitigating a student's academic and/or behavioral difficulties (Brown-Chidsey & Steege, 2005). Previous diagnostic test-and-place education models that sought to diagnose a student's school problems conceptualize student failure as a "within child" problem are giving way to a more systems based approach that focuses

heavily on the educational environment as the intervention lynchpin. Assessment technology appears to be moving toward screening and progress monitoring to not only evaluate student performance, but also to assess the efficacy of curriculum and instruction.

In addition to concerns over academic shortfalls, a great deal of attention has been focused on the increasing prevalence of mental health and behavior disorders among school-aged children (Levitt, Saka, Hunter Romanelli, & Hoagwood, 2007; Robert, Attkisson, & Abram, 1998). According to the United States Surgeon General, 10% of school-age children suffer from some form of mental illness that causes some level of impairment with estimates that only one in five of these children receives the needed treatment (*U.S. Public Health Service, Report of the Surgeon General's Conference on Children's Mental Health: A National Action Agenda*, 2000). Similarly the American Academy of Pediatrics, in its policy statement delivered by the Committee on School Health, estimates that more than 20% of school-age children have diagnosable mental health problems (Taras, 2004). This policy statement advocates for, among other things, the use of a three-tiered model, coordinated written protocols for use in mental health referrals, and the use of outcomes-based research on the efficacy of school-based mental health models designed to improve student outcomes (Taras, 2004).

Positive Behavior Interventions and Supports (PBIS) has become one of the most widely accepted approaches to improving school climate and student behavior. PBIS originated in response to the reauthorization of the Individuals with Disabilities Education Act (IDEA) which called for the use of functional behavioral assessment and positive behavior supports to be used with students identified with behavioral problems that interfered with learning (Sugai, 2007). Over the past decade, PBIS has been adapted and expanded for use in classrooms, schools, and districts, and has thus become generalized as “School Wide” PBIS (SWPBIS). PBIS mirrors the

three-tiered prevention model and thus has structural similarities to the RtI model. Grounded in this PBIS approach are the elements of universal, classroom based assessment and intervention, and the use of frequent monitoring to determine the efficacy of instruction, modeling and interventions.

It may be sensible to assume that improvements in behavior positively affect academic achievement. Fewer behavioral interruptions provide for increases in teacher instruction. Students who are not misbehaving may be more available to learn. This has led to extended research on the relationship between programs geared toward improving behavior and improvements in academic achievement (Kamps et al., 2003; Stewart, Benner, Martella, & Marchand-Martella, 2007). A recent meta-analysis conducted on the impact of social-emotional learning curricula on improving school outcomes indicated that intervention increased academic achievement by 11 percentile points (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). Indeed, the notion that a powerful positive relationship exists between behavior and academics has led to a national initiative – a marriage between Response to Intervention (RtI) and School Wide Positive Behavioral Supports (SWPBIS).

Many states have established a dowry of sorts to promote this marriage. For example, Michigan has created the Michigan Behavior and Learning Support Initiative (MiBLSi). The MiBLSi mission statement reads that it was developed to support and sustain implementation of data driven systems, utilizing a problem solving model in schools in order to help students become better readers with social skills necessary for success (Goodman, McGlinchey, & Schallmo, 2009). The model is presently used in 512 Michigan schools and with 45 collaborating intermediate school districts. Schools affiliated with the project utilize an evidence-based curriculum to promote reading and school-wide positive behavior supports to

reduce the number of behavioral referrals and to increase overall school climate. Member schools are required to report data to the state via outcome measures in reading and behavioral discipline referrals. While reading data is gathered from multiple sources, and utilizes screening, progress monitoring, and outcome measure data, it appears that major office referrals is the only source of data relating to behavior. If this is the case, the model may be neglecting an opportunity for a proactive assessment method that could be utilized to identify and reduce problem behavior through a more intensive screening approach.

Universal screening for emotional and behavior disorders is suggested among the best practices for identifying and serving students who may be experiencing distress; subsequently, providing opportunities to identify and intervene to reduce behaviors that may become more severe in the future (Renshaw et al., 2009). Despite this call, Romer and McIntosh estimate that only 2% of schools screen all students for mental health concerns (as cited in Renshaw et al., 2009). While teachers receive substantial training on curriculum development, instruction, and measurement, they receive comparatively little training on measuring and intervening with problematic student behaviors. This may account for the present shortcoming of schools to identify students needing behavioral intervention through consistent behavioral screening practice. Behavioral screening conducted at prescribed periods can be useful in not only identifying student in need of intervention, but also in triggering early intervention support before more extensive assessments and more restrictive action becomes warranted. The disparity between the current level of teacher training and the increasing identification of children in need of behavioral support proves to be an area of concern for today's schools.

Current Behavioral Screening Technology

Few tools are available to teachers to conduct evaluations of student behavior. Some assessments such as the Systematic Screening for Behavioral Disorders (Walker & Severson, 1992), the Social Skills Improvement System (Gresham, Elliott, Cook, Vance, & Kettler, 2010), and the BASC-2 Behavioral and Emotional Screening System (R. Kamphaus & Reynolds, 2007), have attempted to bridge this gap.

The Systematic Screening for Behavioral Disorders (SSBD), developed by Walker and Severson (1992), is a multi-gated tool that consists of three successive stages of assessment. Stage 1 requires teachers to rank order students according to the presence or absence of internalizing or externalizing behaviors. Stage 2 requires the teacher to complete a 56-question instrument on the top three students from both the internalizing and externalizing lists. Stage 3 requires independent observation of the students' behavior in academic and non-academic settings (Richardson, Caldarella, Young, Young, & Young, 2009; Walker & Severson, 1994).

The Social Skills Improvement System (SSIS) is a comprehensive social skills program that utilizes multi-tiered assessment and intervention at the classroom level (Elliott & Gresham, 2008). The screening component of the SSIS, the Performance Screening Guide (PSG), is a criterion related universal screening measure teachers use to assess all students within a setting focusing on observable behaviors in the domains of positive social behaviors, motivation to learn, reading skills, and math skills (Gresham, Elliott, Vance, & Cook, 2011).

The Behavior and Emotional Screening System (BESS) is a screening instrument used to identify emotional and behavioral strengths and weaknesses in students from preschool to high school that assesses both internalizing and externalizing problems, school-related difficulties, and adaptive skills (R. Kamphaus & Reynolds, 2007). Similar to the BASC – 2, the BESS

utilizes parent, teacher, and self-report forms – each comprised of 25 to 30 items. Informants rate items on a four-point frequency scale (i.e., never, sometimes, often, almost always) resulting in a single score that informs student risk level – normal, elevated, or extremely elevated. While the BESS is considered a new instrument that has not enjoyed extensive study, initial investigations of its psychometric properties indicate generally acceptable levels of reliability and validity (Renshaw et al., 2009).

While each of the screening measures is purported to be efficient and effective in screening for student risk levels, screening of all students can be both expensive and time consuming for teachers. BESS protocols can cost up to $3.00 per child if teacher, parent, and self-report forms are used. Scoring time can be reduced if districts choose to spend nearly $600.00 for scoring software. Furthermore, assessment professionals may be required for scoring and interpreting the BESS protocols. Similarly, the SSIS protocols cost more than $4.00 per child. While cost effective, the SSBD by virtue of its multi-gated approach, can be time consuming and challenging for teachers and/or other qualified staff members required to conduct stage three academic and non-academic student observations.

The need for teacher friendly assessment tools that are both cost effective and efficient, and that measure both academic and behavioral student functioning, is evident. In the fall of 2010, the School Psychology Review dedicated a special print series on behavioral assessment within problem solving models. Content included commentary on needs, limitations, and directions for future research. Universal screening and progress monitoring for school behavior problems have been identified as the new frontier (Evans & Sarno-Owens, 2010; Merrell, 2010). Creating tools that are reliable and efficient is key to the exploration of this new territory. These assessments must be feasible for school staff to administer and interpret. Assessment feasibility

in this regard is not limited to ease of use or brevity; rather, it must also be relevant to the context of school (Evans & Sarno-Owens, 2010). Merrell (2010), identifies three "big ideas" focused on creating more effective school-based behavioral assessment. He lists these as, "(1) universal screening or behavioral and mental health, (2) assessing student strengths, and (3) linking assessment to intervention" (Merrell, 2010, p. 423).

As federal accountability standards mandate schools to report on the efficacy of their treatment efforts, and as greater attention from the medical community focuses on schools as a venue for prevention, school districts and specifically teachers will likely need tools that are efficient for quantifying students' school-based behaviors. Mental health practitioners achieve this metric utilizing the Global Assessment of Functioning (GAF) as part of a multi-axial diagnostic procedure (*American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders*, 2000).

It may be worthwhile to investigate the utility of a global assessment scale of school functioning that may be used by both teachers and school psychologists as a screening tool for the identification of students at-risk of school failure. The Global Assessment of School Functioning (GASF) is purported to be an efficient, inexpensive scale that can be used by teachers who are trained to use the GASF for either screening or progress monitoring purposes (A.C. Maerlender, personal communication, February 5, 2009). The GASF models the Global Assessment of Functioning, which is used by mental health clinicians to quantify an individual's behavior over a period of time (*American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders*, 2000). It was adapted utilizing input from subject matter experts within public and private schools in Northern New England. These experts identified behaviors displayed by students that range from unremarkable to severe. The GASF is a uni-modal

measure encompassing six domains associated with school behavior (work completion, work quality, peer relationships, adult relationships, disruptive behavior, and attendance) that requires the rater, usually a teacher, to assign a numeric score that best describes a student's current functioning. While a brief, global measure may not be as accurate as multidimensional rating scales or direct observation reports in diagnosing problem behavior or its etiology, it would certainly be desirable for capturing the essence of student functioning in a brief, quantifiable manner. Furthermore, the GASF may fill a void in the present assessment technology as a screening and progress-monitoring instrument possessing relevance and feasibility to school staff seeking behavioral assessment options. Finally, the GASF represents an assessment measure that may possess unique transferability that will enhance communication between schools and mental health agencies regarding the functioning of students serviced in both the school and clinic setting.

Research Questions.

If the GASF is to become a useful tool for school personnel to utilize as part of a larger assessment process, its psychometric properties will need to be studied. Currently, no empirical evidence exists on whether this instrument is a reliable or valid measure of student behavior as assessed by teachers and school psychologists. Thus, the purpose of the current project is to answer the following questions:

1. Does the Global Assessment of School Functioning (GASF) possess adequate content validity as assessed by an expert panel?
2. Can school professionals, namely teachers and school psychologists, be adequately trained to utilize the GASF to quantify behavior?
3. Does the GASF demonstrate adequate reliability as measured by an examination of inter-rater reliability?

Chapter 2: Literature Review

Assessing Assessment

Educational reform has led to a paradigm shift in the way we view classroom success. To this end, federal policy has been overhauled and funding has been made available for states to implement prevention based, Response to Intervention (RtI) models as a national framework for delivering curriculum and assessing the quality of both the curriculum and its presentation. RtI is a structure that provides opportunities for accountability and documentation of progress and outcomes based on data and behavior. It does not define the content of instruction, except to recommend validated, best-practice content. Rather, children are screened in a specific domain such as reading, math, or behavior, and then identified for the purpose of managing their education in that domain. Thus, problems with students or curricula can be identified early and addressed more proactively. This is achieved utilizing three types of assessments: universal screeners, progress monitors, and outcome measures. Screeners allow for a rank-ordering of students within the cohort while progress monitors compare the child to himself in a specific skill such as reading fluency. Outcome measures are standardized, often based on national normative data, that compare the child to a national expectation.

In 2009, President Barack Obama, as part of the American Recovery and Reinvestment Act of 2009, launched an initiative to reform schools and empower states and local school districts to research and utilize evidence based best practices to improve educational outcomes for students. The "Race to the Top" initiative is a 4.5 billion dollar competitive funding program that encourages and rewards states that are implementing significant reforms addressing four primary areas: improving standards and assessments, improving data use and collection, building teacher effectiveness and achieving teacher equity distribution, and improving struggling schools (Education, 2010). The area of assessment has received considerable attention from the federal

government; consequently, requests for information have centered on the research and development of assessment technology standards (U.S. Department of Education, 2010). The federal government has identified assessment as a critical component to student progress and school accountability. The Race to the Top Assessment Program was created to,

> provide funding to consortia of States to develop assessments that are valid, support and inform instruction, provide accurate information about what students know and can do, and measure student achievement against standards designed to ensure that all students gain the knowledge and skills needed to succeed in college and the workplace. These assessments are intended to play a critical role in educational systems; provide administrators, educators, parents, and students with the data and information needed to continuously improve teaching and learning. (U. S. Department of Education, Race to the Top guidelines and FAQs, p. 3, 2010)

Assessment may be defined as the process of collecting data for the purpose of making decisions about the performance of individuals or groups on a given skill or competency set (Salvia & Ysseldyke, 2007). While its forms vary, assessment can be used for screening, progress monitoring, or summarizing purposes. Assessment begins at the dawn of every school year and continues through the dusk that precedes the close of school. It begins when the teacher places the first score next to her pupil's name to indicate a level of mastery on a given task.

Among the identified challenges that arise with assessment (grading) is the inevitable variability that exists in assessing skills. Furthermore, increased teacher responsibilities within the classroom to teach expanding content to increasingly diverse learners provides another hurdle. Efficient assessment tools are at a premium, and the ability to assess large populations

in relatively short order is a demand that has surfaced. This need has resulted in the research and development of curriculum based assessment to measure discrete skills in reading, writing, and math. In order to effectively use these curriculum based measures, Deno (1985), suggests that these tools must be reliable and valid, simple and efficient so that teachers can use them frequently for monitoring, easily understood and communicated, and inexpensive to be utilized repeatedly. As a result, commercially based curriculum based measures and monitoring systems such as the Dynamic Indicators of Basic Early Literacy Skills (Good & Kaminski, 2009), and Aimsweb (Shinn, 2005), have enjoyed popular acclaim for their ability to assess and monitor progress in reading and in the case of Aimsweb, in math, writing, and spelling. Additionally, Aimsweb offers the ability to track screening and progress monitoring results for students over behavioral domains using office discipline referrals or other commercially available assessments of student behavior (i.e., BASC 2 – BESS, or SSIS). Indications are that both the Dynamic Indicators of Basic Early Literacy Skills and the Aimsweb technology provide schools with a very good method of interpreting the academic data they gather (Good & Kaminski, 2009; Shinn, 2005).

Information on the Aimsweb behavior management program does not necessarily pertain to the focus of the present investigation as Aimsweb is simply a data management tool in regard to behavior. Instead, the entirety of this review will center on the measures that are presently employed in the screening and progress monitoring of students in schools. A brief discussion on the use of office discipline referrals (ODR) is provided in terms of their use in progress monitoring. Additionally, this paper reviews the use of global assessment scales used in the mental health arena, with the thought that global assessment scales may prove to be an assessment technique that could well serve schools in assessing student behavior.

Screening for Behavioral Risk in Students

Increased interest in screening and monitoring school-based behavior has led to the production of several assessments and techniques designed to quantify student behavior. Three of these tools, the Social Skills Improvement System (SSIS), the Systematic Screening of Behavior Disorders (SSBD), and the Behavioral and Emotional Screening System (BESS), have been identified as psychometrically sound measures intended to identify students who may be at-risk for behavioral difficulty. They were chosen for this review based on assertions that each of the measures may be used in part or whole as screening and/or progress monitoring tools as part of a proactive behavioral assessment program.

The Social Skills Improvement System (SSIS).

The SSIS, published in 2008, is a norm referenced assessment system purported to classify behaviors deemed important for school success (Gresham & Elliott). The SSIS marks a revision of the Social Skills Rating System (Gresham & Elliot, 1990), and includes changes to both content and normative data. The SSIS is comprised of four components, the Performance Screening Guide (PSG), the Social Skills Rating Scales (SSRS), a curriculum meant to strengthen student social skills in the general education classroom, and an intervention guide that utilizes SSRS data to inform social skill interventions. The PSG allows for universal screening of students across a class or an entire school. Data from the PSG may be used to develop class wide intervention programs or to evaluate the effects of interventions on academic and behavioral performance (Gresham & Elliott, 2008).

The PSG is presented as a booklet that allows teachers to rank order class rosters based on performance levels over four skill areas – prosocial behavior, motivation to learn, reading, and math. Three forms of the booklet are available and are stratified by preschool, elementary,

and secondary levels. Low rankings indicate areas of concern. Test authors estimate that a classroom can be ranked using the PSG in 30 minutes.

The PSG was field-tested using 138 teachers who had participated in the standardization studies on the SSRS. Teachers used the PSG to rate a total of 2,497 students from preschool through secondary grades. Survey results indicated that teachers agreed or strongly agreed that behaviors identified in the PSG are important indicators and that the assessment tool was easy to use. Test-retest reliability is described as moderate with correlations ranging from .53 to .62 for preschool teacher/student ratings and from the high .60s to low .70s for elementary and secondary teacher/student ratings. On average, the retest interval was completed in 74 days. Inter-rater reliability participants included 44 teachers and/or teaching assistants evaluating 434 total students. Individual teachers were paired with a team teacher, teaching assistant, or other staff member who had sufficient student contact to provide a rating. Ratings were calculated based on three school levels – preschool, elementary, and secondary over four skill areas – prosocial behavior, motivation to learn, early reading skills, and early math skills. Intraclass correlations were consistently established in the moderate range for all school levels and skill areas with the exception of the secondary prosocial behavior area (.37). Three of the four skill areas at the preschool level exceeded .70, which is described as a substantial correlation in the manual. In addition to data substantiating the reliability of the PSG, moderate correlations between the PSG skill area scores and subscale scores from the SSIS Rating Scales are offered as support for the criterion validity of the SSIS Performance Screening Guide (Elliott & Gresham, 2008).

The Systematic Screening for Behavior Disorders (SSBD).

The SSBD is a three stage, multiple-gating procedure, that leads to the eventual identification of students at-risk of behavioral difficulties (Walker & Severson, 1992). Stage one requires a teacher to rank order their student rosters in terms of the observance of student externalizing behaviors and again rank order their students in terms of the observance of internalizing behaviors. Students who occupy the first three rankings in either or both behavioral dimensions progress to the second stage. Stage two requires the teacher to complete two questionnaires – the Critical Events Index (CEI) and the Combined Frequency Index (CFI) on each of the students identified in stage one. The CEI is a 33-item checklist that indicates the presence or absence of particular target behaviors, while the CFI utilizes a five-point rating scale (never to frequently) to rank 12 adaptive and 11 maladaptive behaviors (Zlomke & Spies, 1998). Students who meet criteria on either the CEI or CFI are identified for the final evaluation stage. Stage three requires systematic student observation by school personnel trained in the use of observational coding. Students are observed in two different settings during four different 15 minute interval recording sessions (Zlomke & Spies, 1998). Students who "pass" through this third gate are identified for the referral process.

Technical data on the SSBD appears to exhibit good psychometric properties. While no data exists on sampling for stage one, stage two and stage three were tested on a national standardization sample (eight states, 18 school districts) of 4,463 and 1,275 students respectively from grades kindergarten through sixth (Zlomke & Spies, 1998). Stage 1 test-retest data is reported as .76 for externalizing behaviors and .74 for internalizing behaviors over a one-month span. Stage two test-retest reliability is listed at .88 for adaptive and .83 for maladaptive

behaviors as defined on the CFI. Stage three inter-rater reliability is identified between .80 and .90 (using 10 second interval recording).

Additional studies on the technical merits of the SSBD in identifying at-risk students have demonstrated that the SSBD accurately and efficiently identified students in need of special services. Walker and colleagues utilized first through fifth grade students (N = 1,468) and their teachers (N = 58) in three Utah elementary schools (Walker, Severson, Nicholson, & Kehle, 1994). Walker et al used videotaped instruction to train teachers on the use of the SSBD and on observation procedures. Eighty-four percent of students were correctly classified using the SSBD into internalizing, externalizing, and non-ranked subgroups. The authors also reported that teacher satisfaction surveys indicated that resource teachers and school psychologists view the SSBD favorably in its effectiveness in identifying externalizing and internalizing behaviors and rated the measure as helpful in identifying and screening children.

BASC – 2 Behavioral and Emotional Screening System.

The BASC – 2 Behavioral and Emotional Screening System (heretofore referred to as the BESS) is a screening tool designed to assess the strengths and weaknesses of students aged 3 – 18 (Furlong & O'Brennan, 2007). It is intended for use by schools, pediatric clinics, communities, mental health clinics and researchers to screen for a variety of emotional and behavioral concerns. Creation of the BESS stemmed from the need for an efficient, psychometrically sound instrument that could accurately identify children with varying risk levels in the emotional and behavioral domains. The test developers specify using the BESS as an efficient method to conduct systematic, early screening to identify students at-risk of behavioral difficulty in schools.

The BESS was developed using items that were drawn from the Behavior Assessment Scale for Children, Second Edition. Items that comprised the highest factor loadings from the BASC-2 composites were selected to create teacher, parent, and self-report (for grades 3-12) forms consisting of 25 to 30 items that take about 5 minutes per form to complete. A total T-score derived from raw scores is reported and accompanied by a qualitative descriptor of risk level. T-scores below 60 are considered normal, 61-70 are considered elevated, and scores 71 and above are considered extremely elevated.

Furlong, O'Brennan, and Johnson (2007), provided a supportive summary of the technical merits of the BESS. The normative sample was reported as diverse and commensurate with the U.S. census data in terms of race and ethnicity, geography, socioeconomic status, and special education classification. The nationwide sample was conducted with respondents from 40 states and included a sample of 3,300 students, 4,450 teachers, and 4,600 parents. Sufficient evidence of internal consistency, test-retest reliability, and interrater reliability is found in the BESS manual (Furlong & O'Brennan, 2007). Internal consistency, measured using split-half reliability coefficients is reported within a range of .90 -.97. Test-retest reliability coefficients are reported as ranging from .80 to .91. The interval between testing ranged from 0 – 88 days. Interrater reliability was assessed using paired ratings of a single child. Parent forms are reported to demonstrate slightly higher reliability (.83 and .82) than teacher forms (.80 and .71). Evidence for the concurrent and predictive validity of the BESS's use as a screener appears adequate (Furlong & O'Brennan, 2007).

Furlong and O'Brennan cite concurrent validity statistics from the BESS manual comparing the total score to scores taken from the BASC-2, Achenbach System of Empirically Based Assessment, the Behavior Rating Inventory of Executive Function (BRIEF), the Conners'

checklists, and the Vineland – II. The BESS correlated highly with the BASC-2 teacher (.94), parent (.90), and self-report (.86) global composite scores. Correlations with various forms of the Achenbach System of Empirically Based Assessment measures appear relatively strong ranging from .71 to .77. Similarly strong validity correlations were shared with the BRIEF global composite score (.78). Relatively high correlations were found between the BESS total score and measures from the teacher and parent forms of the Conners' (.78 and .62 respectively), and moderate validity correlations with the Conners' student forms (.52). Moderate concurrent validity correlations were achieved with comparison to the Vineland – II. Correlations with teacher measures of the Vineland are listed as -.39 for preschool age and -.66 for child/adolescent forms. Parent forms correlations are reported as -.46 and -.50 for preschool and child/adolescent forms respectively. In addition to concurrent validity with standardized measures, the BESS risk-level classifications have been shown in at least one study to demonstrate concurrent validity with related school-based outcomes (Renshaw et al., 2009).

Initial predictive validity studies at both the preschool and school-aged ranges suggest that the BESS may be used as a risk indicator to forecast future academic and or behavioral difficulty (DiStefano & Kamphaus, 2007; Kamphaus et al., 2007). A longitudinal study conducted on 423 kindergarten students in Georgia compared ratings from the initial teacher screener to measures of discipline referrals and academic performance and found that higher scores on the screening tool were related to weaknesses in students' behavioral and academic readiness (DiStefano & Kamphaus, 2007). Similarly, using the BASC – 2 Behavioral Symptoms Index as a correlate, the screener was reportedly efficient at identifying students' risk levels. The BESS demonstrated high sensitivity values (.94) in identifying children exhibiting significant

problematic behavior problems and good specificity (.74) in determining the absence of problems (Kamphaus, DiStefano, Dowdy, Eklund, & Dunn, 2010).

In reviewing the literature, the BESS appears to have good psychometric properties, and as a screener, it appears to be efficiently administered to students. Less clear is the perceived ease of use and maintenance using the system. While administration time was reported as requiring about five minutes per informant, there is no mention as to the amount of time it takes for staff to score and interpret the results of the screener. Scoring software exists to aid in the expediency of scoring and interpretation for roughly $600.00. Furthermore, the scoring and interpretation is recommended to be completed by professionals with experience in testing and affiliation with professional organizations such as the National Association of School Psychologists or the American Psychological Association (Pearson, 2011). Schools who are utilizing multiple informant ratings using the BESS may experience scoring fatigue or delays in gathering protocols, presenting scores, and creating intervention groups. This time challenge paired with the purchase price of rating forms, manuals, and software may be prohibitive.

Office Discipline Referrals

Office discipline referrals (ODR) have been endorsed nationally as a metric for managing and monitoring behavior deemed disruptive to schools (Clonan, McDougal, Clark, & Davison, 2007; Irvin et al., 2006; Irvin, Tobin, Sprague, Sugai, & Vincent, 2004; Sugai, Sprague, Horner, & Walker, 2000). ODR are an efficient measure that schools have used for many years—the difference is in how this information is analyzed. In schools utilizing principles consistent with PBIS, ODR represent a tool that is utilized to identify environmental factors that accompany referable behavioral infractions. They provide school based behavioral teams with information as to the efficacy of the present program as well as guidelines as to where, when, and what

behavior continues to be problematic (Sugai et al., 2000). In terms of usefulness to data managers and decision teams, computer databases appear to be helpful in interpreting ODR data. Using the School Wide Information System (SWIS), Irvin et al. (2006), found that ODR data can be useful in identifying early problem behavior, identifying specific behavior problems, developing interventions, and monitoring interventions at both the elementary and middle school level (2006). As a method of assessing systems based behavioral questions, such as where and when, ODR appear acceptable.

Use of ODR as a screening tool may not be as sensible. When assessing the convergent validity of ODR with the Teacher Rating Form of the Child Behavior Checklist, ODR failed to identify existing problems (Nelson, Benner, Reid, Epstein, & Currin, 2002). While the discrepancy was higher for the Internalizing scale, ODR was unable to adequately identify students on the Externalizing scale, the Delinquent Behavior subscale, or the Aggressive Behavior subscale. This result was surprising to the researchers, as these behaviors would most likely correlate to the behaviors associated with ODR.

In addition to the limitations of ODR as a screening tool, one must question the validity of the ODR's ability to measure behavior change. Are reductions in ODR a product of the intervention on student behavior or teacher behavior? Perhaps that is not a question that this form of measurement seeks to answer, but it is a question worth asking. Teachers evaluated on their referral frequency may give pause before sending a child to the office; thus, rendering the ODR metric flawed.

Global Assessment Scales: Reliability, Validity, and Utility of the GAF and CGAS

The use of global scales for measuring behavior appears to provide researchers and clinicians with an efficient, valid, and reliable quantitative measure of behavior that is sensitive

to change (Keraus, 1991). Global assessment permits the clinician to utilize her clinical judgment when providing estimates of current functioning while avoiding ambiguous descriptors. The numbering system of global assessments allows clinicians to provide numeric values that are linked to behaviorally descriptive anchor points that avoid the use of ambiguous, subjective wording such as *better or worse*. Mental health agencies that have utilized this model for several decades report their findings via a multi-axial diagnosis that is summarized by the Diagnostic and Statistical Manual, Fourth Edition (DSM-IV) Axis V -- Global Assessment of Functioning (GAF).

The Global Assessment of Functioning Scale.

The present form of the GAF was adapted from the Global Assessment Scale (Endicott, Spitzer, Fleiss, & Cohen, 1976) and the third edition of the Diagnostic and Statistical Manual (1987). The GAF is 100-point scale that is divided into ten ranges of functioning. The clinician is trained to assign a single, global number that represents his or her best judgment of the client's overall functioning. The GAF is frequently required by third-party payers and insurance companies at intake and exit and is a key element in tracking clinical progress in individuals (*American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders*, 2000).

Several studies exploring the psychometric properties of the GAF have demonstrated it is a reliable and valid measure (Hilsenroth et al., 2000; Rey, Starling, Wever, Dossetor, & Plapp, 1995; Schorre & Vandvik, 2004; Startup, Jackson, & Bendix, 2002). Hilsenroth and colleagues assessed the reliability and validity of the DSM-IV Axis V (heretofore referred to as the GAF) ratings of ten graduate students on 44 patients in an outpatient university-based community clinic (Hilsenroth et al., 2000). Raters consisted of ten advanced students in a clinical psychology

Ph.D. program who received training on scoring using three scales of interest. Inter-rater reliability for each scale was assessed using Intra-class correlation (ICC) of DSM-IV symptoms; relational, social, and occupational functioning; self-report measures; and Axis II pathology. In doing so, the investigators provided an additional measure of inter-rater reliability for the GAF. Additionally, Hilsenroth and colleagues examined a convergent and discriminant validity of the GAF using factor analysis. The scales included the Global Assessment of Functioning Scale, and two experimental scales -- the Global Assessment of Relational Functioning Scale (GARFS), and the Social and Occupational Functioning Assessment Scale (SOFAS). Using a one-way random effects model, results indicated reliability of the GAF (as well as the GARFS and the SOFAS) in the excellent range (ICC > .74). This suggests that ratings indicated significant agreement in scores across raters. Factor analysis revealed that both the GARFS and SOFAS are related to the GAF constructs (social and occupational functioning).

Additional studies have investigated the relationship between training and experience rater agreement on GAF ratings. Warsi and colleagues (2007), investigated differences between medical students', psychiatry residents', and staff psychiatrists' ratings using two clinical vignettes. The investigators also examined whether reviewing GAF scoring guidelines decreased differences in ratings between the groups. Using measures of central tendency, the researchers found that the ratings of medical students differed significantly from both the residents' and staff psychiatrists' assigned ratings on one of the two vignettes. However, when participants were given the chance to review the GAF scoring guidelines, and asked to re-rate the vignettes, agreement improved. Implications suggest that training and experience may lead to higher levels of agreement when assigning GAF ratings.

The use of a global functioning scale has been considered important for describing the

level of functioning of a person as part of the context for understanding his present condition, status, or diagnosis. Despite discrepancies over the reliability and validity of the GAF, mental health practitioners continue to utilize the GAF as a clinical monitoring tool.

The Children's Global Assessment Scale.

The CGAS was developed as a measure to assess child and adolescent global functioning (Shaffer et al., 1983). Like the GAF, the CGAS was adapted from the Global Assessment Scale developed by Endicott et al., (1976). The CGAS is intended to represent the lowest level of functioning of a child or adolescent during a determined time period. Scores are reported as a single number that ranges from 1-100 with scores above 70 indicating normal functioning. The CGAS contains behavioral descriptors at anchor points that are intended to express levels of functioning ranging from superior to extremely impaired. Psychometric properties are reported to be satisfactory. The initial study conducted by Shaffer et al. (1983) utilized 19 written case vignettes, rated by five second-year child psychiatry fellows to assess inter-rater reliability; ICC coefficient is reported as .84. Test-retest stability was assessed after a period of roughly six months using the same raters and same vignettes. ICC at the second time point was reported at .85, and the author noted that all but one of the five raters demonstrated consistent ratings over time. A measure of discriminant validity was provided by Shaffer et al. (1983) comparing inpatient CGAS ratings to the CGAS ratings of outpatient child clinic patients. The data suggests mean CGAS scores of 65 indicate caseness to receive outpatient services while mean scores of 46 and below were typically found for inpatient children.

In an effort to provide additional support to the reliability of the CGAS, Bird and colleagues (1987) conducted a pilot study at a clinic staffed by professionals at the University of Puerto Rico School of Medicine in San Juan. Four child psychiatrists working in teams of two

(one who served as primary interviewer, and one who based ratings on observations from videotaped interviews) provided paired ratings on a total of 91 patients. Inter-rater reliability was reported as high (.83). This study also examined the correlation of the current CGAS with the total problems scores of the parent version of the Child Behavior Checklist at the time of investigation and after six months. Pearson correlations were reported at -.65 and -.62 respectively.

Dryborg and colleagues found similar inter-rater reliability coefficients among practicing child and adolescent psychiatrists, clinical psychologists, and child psychiatry trainees using the CGAS to evaluate 145 patients seen in a child and adolescent psychiatric hospital setting (Dyrborg et al., 2000). Practicing child and adolescent psychiatrists evidenced the highest levels of agreement (.87). Combined ratings including raters from all levels of training demonstrated moderate agreement among all raters (.79). Of particular interest, Dyrborg, et al. suggests that level of training and sample size appear to be indicators of agreement. When experienced clinicians, the practicing psychiatrists and psychologist, rated the same 95 cases, agreement improved to .89.

Additional studies suggest at least moderate levels of inter-rater reliability when the CGAS is used in naturalistic settings (Lundh, Kowalski, Sundberg, Gumpert, & Landen, 2010). Lundh and colleagues compared the ratings of 703 mental health care workers to five experienced clinicians on five case vignettes and found agreement among health workers compared to the expert ratings at .73. Another study examining the inter-rater reliability of the GAF compared ratings to the CGAS and found raters to demonstrate moderate levels of agreement on the CGAS and the GAF. Four separate studies were conducted as part of the larger investigation published by Rey and colleagues (1994). Twenty trained professionals (four child

psychiatrists, two child psychiatry trainees, four psychiatrists in training, two clinical psychologists, and two social workers, as well as an additional six professionals from a separate clinic) provided ratings on 162 child patients in outpatient and inpatient clinical settings. Two separate professionals from the rater pool ranked each child, yielding 324 separate ratings. Training provided on the use of the CGAS and GAF was described as "minimal". The investigation consisted of four studies. Studies 1 and 3 utilized outpatient ratings (study 1 ratings were made using the GAF) Studies 2 and 4 utilized inpatient ratings (study 2 ratings were made using the GAF). Results indicated ICC ratings in the moderate range (.54 - .66). The authors note that the correlations are similar to ratings on previous versions of the GAF but were substantially lower than on previous studies using the CGAS (Bird, Canino, Rubio-Stipec, & Ribera, 1987; Shaffer et al., 1983).

While these studies using the CGAS and GAF indicate moderate to significant levels of reliability in clinical settings, there is no evidence to suggest that these results are transferable to public schools settings. Furthermore, the validity studies that have been conducted are focused on the global measures' ability to indicate caseness for diagnostic and treatment purposes. Used for these purposes, global assessment measures appear to be satisfactory tools for monitoring the functioning of individuals seen in clinical settings.

In sum, these global assessment measures were developed to be utilized by trained clinical professionals in mental health settings, and use in public school settings does not appear appropriate as those professionals are not typically employed by schools. Their popularity and technical qualities suggest that a measure geared toward schools is worth consideration. Practitioners identified a primary limitation of the GAF – it did not adequately assess the functioning of children; subsequently, the CGAS was created to mitigate that limitation.

Similarly, the CGAS represents a considerable limitation to education professionals who may seek to use a global assessment tool for screening and progress monitoring students – the CGAS is a clinical measure not intended specifically for school consumption; therefore, a school based global assessment tool would serve as a response to the stated challenge to using the CGAS in schools.

Before engaging in a full-blown study on the merits of using a global measure for school use, it should be noted that considerable debate regarding global assessment measures – specifically the GAF. The American Psychiatric Association, in its release of the Diagnostic and Statistical Manual for Mental Disorders, Fifth Edition (DSM-5), has eliminated the use of the GAF completely citing concerns ranging from a lack of conceptual clarity to questionable psychometrics (Gold, 2014). Despite these concerns and omission of the GAF from DSM-5, practitioners may continue using global measures based on the practitioners' familiarity with the measure and reservations regarding the DSM-5.

Summary

The "Race to the Top" is no doubt a marathon – not a sprint in regard to behavioral assessment and intervention. The finish line is well marked by way of office discipline referrals, but presently, the course is relatively void of check-points and fueling stations. As a result, the development of research based, psychometrically sound screening and progress-monitoring tools has received considerable attention. While a handful of assessments presently exist, their feasibility has not yet been proven on a large scale. School-based screening and progress monitoring of behavior on a grand scale is a relatively new enterprise and questions arise as to how assessment will be conducted, how data will be analyzed, and which tools represent the most feasible and relevant materials.

Based on this information, investigation of a global assessment scale for use in schools may warrant further consideration. The Global Assessment of School Functioning (GASF) is being developed to be an efficient, inexpensive scale that can be used by teachers for either screening for behavioral functioning or progress monitoring purposes (A.C. Maerlender, personal communication, February 22, 2009). The GASF (see Appendix A) is modeled after the Global Assessment of Functioning (GAF), used by many mental health providers to assess current levels of functioning and for progress monitoring of treatment outcomes. The rationale for modeling the GASF after the GAF was based on several factors. Along with assessing present levels of functioning and progress monitoring, the GAF has also demonstrated itself to be a valid and reliable measure that synthesizes information relating to the patient as a whole rather than the sum of his or her parts. An assessment that looks at the whole child rather than one or multiple fine-grained behavioral or academic measures may be of great value to teachers who seek to organize their perceptions about the student. Global measures like the GAF are intended to assess whether a patient is doing better or worse in quantifiable terms without the focus on the why or the how. Furthermore, the GAF provides a consistent language and metric that is understood by those professionals using the tool on a daily basis.

While a brief, global measure may not be as accurate as multidimensional rating scales or direct observation reports in diagnosing problem behavior or its etiology, it may be desirable for capturing the essence of student functioning in a brief, quantifiable manner. The GASF may fill a void in the present assessment technology as a screening and progress-monitoring instrument possessing relevance and feasibility to school staff seeking behavioral assessment options. Finally, the GASF represents an assessment measure that may possess unique transferability

between schools and mental health agencies when communicating the functioning of students serviced in both the school and clinic setting.

The purpose of this research project is to investigate the psychometric qualities of the GASF. This study assessed elements of content validity and inter-rater reliability (IRR) utilizing content matter experts (CMEs) to provide feedback regarding the structure of the GASF. CMEs also served to validate vignettes that were rated by school professionals to assess IRR. Finally, this study gathered data based on responses from school personnel regarding their perceptions of the GASF.

Chapter 3: Method and Results

The researcher examined the content validity and inter-rater reliability of the GASF. The GASF is a uni-modal measure encompassing five domains associated with school behavior (work completion, work quality, peer and adult relationships, disruptive behavior, and attendance) that requires the rater, a teacher or school psychologist, to assign a numeric score that best describes a student's current functioning. It was developed utilizing input from subject matter experts within public and private schools in Northern New England. These experts identified behaviors displayed by students that range from unremarkable to severe. The GASF is a 100-point rating scale of global student functioning, modeled after the Global Assessment of Functioning of the DSM-IV, was designed to be completed by knowledgeable teachers and school psychologists. The GASF is partitioned into ten-point increments that are expected to indicate levels of global school functioning; scores in the 91 – 100 range indicate the highest level of school functioning while lower numbers indicate progressively poorer school functioning. Each band has a description of typical student behavior for that functional level. The frequency, intensity, and duration of the behaviors were considered when clustering and developing the anchor points (the ten point bands). Student behavior was operationalized to consist of six dimensions – work completion, work quality, peer relationships, adult relationships, disruptive behavior, and attendance.

A two-phased approach was utilized to examine the GASF's psychometric properties. Phase one consisted of three tasks to explore aspects of validity. First, a simple measure of content validity was used to determine the GASF's structure and anchoring system. Next, 15 vignettes were rated by content matter experts and identified as fitting into a given range within the GASF. This provided another measure of content validity. Finally, 10 of these vignettes

were selected for the larger reliability study. This was achieved by requesting expert judgment from experienced teacher consultants who possess knowledge in assessment, achievement, behavior, intervention, and progress monitoring. Phase two consisted of a reliability study that required school professionals, namely teachers and school psychologists to use the GASF to practice rating five validated vignettes and then make final ratings for another five vignettes, in addition to assessing school professionals' perceptions of the GASF. It should be noted that the researcher obtained permission from the test author to utilize the GASF. Furthermore, the researcher collaborated with the test author to improve the GASF based on the findings of the present study.

Phase 1a: Content validity and the GASF

Two elements of validity for the GASF, face and content validity, were assessed through solicitation of feedback provided by teacher consultants serving as content matter experts (CME). In addition, for the purpose of the present study, vignettes created by the principle investigator were validated by the CME. The panel was asked if the vignettes created represented typical school based behaviors and possessed sufficient information for making ratings.

Critical elements of content validity for the GASF include the notion that the measure contains an adequate content sample of student behaviors, the behaviors are defined in global terms, and these behaviors are organized in a manner that reflects the incremental severity of the behavioral groupings. In other words, does the GASF reflect real-world characteristics or behaviors that are demonstrated by students within the school environment? It was hypothesized that the factors identified within the GASF (attendance, work completion, work quality, peer and adult interactions, and behavior disruptions), account for much of what is deemed "school based

behavior". One acceptable procedure for judging a measure's content is by seeking feedback from individuals who can provide intelligent judgment regarding the adequacy of an instrument (Fraenkel & Wallen, 2000).

Participants

Teacher consultants were selected as an expert panel to provide judgment regarding the adequacy of the GASF based on the assumption that in their roles, they possess broad skills in assessment, measurement, student and teacher behavior, classroom dynamics, and learning problems. Among their responsibilities, teacher consultants frequently provide both direct and indirect services to students identified as needing academic support; administer, score, and interpret academic achievement assessments; assist general education teachers in the modification of the general education curriculum for special education students; serve as multi-disciplinary education team (MET) members; and work with teachers and students to implement interventions and accommodations.

Four teacher consultants from northwestern Michigan, with a minimum five years experience working with both general and special education students were selected by the researcher to serve as content matter experts. Three of the individuals were recommended to the primary researcher by his internship supervisor. The other teacher consultant worked closely with the primary researcher as part of his assessment team. The teacher consultants selected demonstrated good understanding of assessment and intervention and were experienced applying these skills with students at the elementary school level. Based on their education and experiences, these individuals have evidenced the ability to follow best-practice methods and dynamic assessment to obtain positive outcomes for the students they serve.

Table 1

Content Matter Experts (CME) Experience and Qualifications

CME	Years as Teacher Consultant	Teaching Experience	Grade levels	Education/Endorsements
CME1	8	15	K-8	BA Special Education MS Special Education Systems Coach
CME2	8	6	K-5	BA Special Education MA Elementary Education
CME3	15	22	3-6	BA Elementary and Special Education (MR) Multicategorical Teacher
CME4	5	12	K-8	BS Special Education, CI MS General Education

Note. Abbreviations for professional specialization endorsements: MR = Mentally Retarded, CI = Cognitively Impaired.

Instruments/Measures

The Global Assessment of School Functioning (GASF).

This version of the GASF was provided by the test author as part of a collegial collaboration.

Content Validity Protocol.

Content validity was assessed by teacher consultants via a validation tool constructed for use in this study. In order to produce the content validity protocol, the GASF was modified by randomly ordering the descriptor groupings and adding corresponding blanks adjacent to the groupings for CME's to record their ordering of the descriptors (Appendix B).

Procedure

Following university institutional review board and approval from the agency employing the teacher consultants, the researcher utilized the following procedure in conducting phase one of the proposed study:

Step 1. The principal investigator contacted potential teacher consultants from the employing intermediate school district via email to solicit their participation as CME in this research study. The email consisted of a cover letter introducing the recipients to the purpose of the study (Appendix D) and an attachment that included a university approved informed consent document (Appendix E). The informed consent document addressed the general purposes of the study, the expected experimental requirements for the teacher participants, the confidentiality of their responses, the adherence to ethical principles in the planning and conduct of the study, and the opportunity to receive a summary of the results at the conclusion of the project.

Step 2. Upon receipt of the signed informed consent documents from all four CMEs, the principal investigator arranged a meeting to conduct the validation procedure. The investigator presented the raters with a copy of the validation protocol to be rated, read aloud the directions, and fielded procedural questions during the group meeting. The CMEs completed the content validity protocol and the completed ratings were placed in a sealed envelope.

Content Validity Analysis and Results

Elements of content validity for the GASF were assessed utilizing feedback provided by content matter experts (CME) during a meeting that took place at the ISD offices. Each of the CME's was given a *Content Validity Protocol* that was constructed for the purpose of assessing the behavioral hierarchy of the GASF (Appendix B). CME rank ordered the behavioral groupings from highest (10) to lowest (1) level of functioning. For example, each of the raters

scored the "Meets all expectations, …superior functioning day in and day out", statement a 10, "Completes work with no reminders…", a 9, and so on. The four raters independently completed the Content Validity Protocol, and upon completion, the researcher tallied results and asked the CME if they experienced any challenges with the task. CME were able to order the deciles with 95% accuracy (38 of 40 possible agreement points). The results satisfied the 100% agreement criteria (each of the four respondents within plus or minus one ranking of each other on each anchor point). The singular discrepancy came from one rater who reversed items five and six when compared to the three other CME. Discussion indicated that the term "moderate" was not readily identified by the rater. Table 2 represents a summary of raters' scores on their Content Validity Protocol. The top row of Table 2 indicates where on the Content Validity Protocol each item was presented. Subsequent rows indicate the rater and the rating given to each item.

Table 2

Content validity and the GASF: Content Matter Experts Item Agreement Results

Item/	1	2	3	4	5	6	7	8	9	10
*Expected ranking	2	9	3	1	10	7	5	6	8	4
Rater 1	2	9	3	1	10	7	5	6	8	4
Rater 2	2	9	3	1	10	7	6	5	8	4
Rater 3	2	9	3	1	10	7	5	6	8	4
Rater 4	2	9	3	1	10	7	5	6	8	4

Note. The expected ranking is the decile order of the item taken from the GASF.

At the conclusion of the rating, the four raters were asked to share their thoughts relating to the wording of the descriptors, the grouping of the items, and the clarity of the statements.

Specifically, CME's were asked to answer the following five questions relating to properties and usage of the GASF: Does the GASF contain an adequate content sample of student behaviors? Are the behaviors defined in global terms (e.g., are the terms broad enough to allow the rater to consider a variety of behaviors representative of each of the anchor points)? Do the groups as you ranked them appear to comprise a hierarchy of behavior (e.g. do behaviors reflect the incremental severity of the behavioral groupings)? Is there anything you would add, change, delete, etc.? Do you feel that when trained, teachers can utilize the GASF reliably as a universal screening tool for students? Based on discussion and input from the CME, it was agreed upon by the group and the researcher to amend the GASF to include a statement about attendance/truancy throughout the first seven levels of the GASF. It was also decided that the word "significant" would be added to the 31-40 range to indicate the level of intervention. Lastly, the group agreed that mention of special education status in areas of the GASF was warranted. At first glance, two members of the CME panel expressed concern that school personnel may feel inclined to rate special education students lower as the first mention of special education services came at the 41-50 range. This was addressed by adding a statement in the 71-80 range (if a special education student, is nearing exit based on remediation of skill deficits). Table 3 represents a summary of CME responses to the five questions. The researcher collected the information from the CME and sealed the information in an envelope for analysis.

Table 3

Content Matter Experts Responses to Questions Regarding the Properties of the GASF

	Rater 1	Rater 2	Rater 3	Rater 4
Student Behavior Content	Yes	Yes	No. At Grade level doesn't hand in work	Yes
Global Terms	Yes	Yes	Yes	Yes

Hierarchy	Yes	Yes. Would they understand the severity in terms of disability v. not in special education, making progress toward goals v. not progressing, etc.	Yes. I feel putting in Tiers for MTSS/RtI is helpful in describing the behavior.	Yes
Changes	for number four, qualify intervention as "significant" (…requires significant intervention…)	Perhaps defining what is considered an intervention (behavioral and academic.	Does behavior include executive functioning / students' motivation	Perhaps include a comment about absences/truancy/ tardies in all areas. #s 4 and 5 special ed. V. Tier II / Tier III
Teacher Use	Yes	Yes	Yes	Yes

Phase 1b: Vignette Reliability Pilot Study

In addition to assessing face and content validity of the GASF, the same four CMEs were asked to rate 15 case vignettes using the GASF in order to identify a group of ten vignettes to be used in the phase 2 reliability study. The final set of vignettes depicted representative levels of functioning as measured by the GASF. Vignettes that did not fall within the 16-point range were eliminated from further study. The vignette reliability study took place at the conclusion of the content validation exercise and utilized the same informed consent document.

Instruments/Measures

The Global Assessment Measure for Schools (GASF).

The original GASF was used for this phase of the study.

Case Vignettes.

The principal investigator utilized teacher nominated child study referrals as the primary source to create fifteen case vignettes. During child study team meetings, the principal investigator asked questions related to the students' work quality, work completion, attendance, peer and teacher relations, and discipline concerns. The vignettes were modified from this case material to ensure anonymity but retained relevant available information required for making ratings (Appendix C). The principal investigator then used the GASF to quantify the students' functioning based on the information provided. The researcher used this information to write the vignettes and then ordered them into three groups (below 30, 31-70, and 71-100). Two vignettes were developed and scored to represent behavior expected to score above 80 on the GASF, two vignettes were developed and scored to represent behavior expected to score below 30, and the remaining vignettes were developed and scored to represent behavior expected to score between 30 and 80. Vignettes that received ratings from each CME that were within 16 points of one another were selected for use in the pilot study. Scores at the higher extreme (above 90) represent fictitious cases based on professional experience, as students representing these scores were not encountered during child study meetings. Names of school and student were changed to ensure confidentiality and anonymity. Word count for vignettes ranged from 68 to 360 with a mean word count of 157.

Vignette Reliability: Procedure, Analysis, and Results

Procedure

After the CME's completed the content validity protocol (and after a short break), the principal investigator introduced the vignette scoring exercise and briefly explained how the GASF was developed and how it may be used as a screening and progress monitoring tool by

school personnel. The principal investigator then provided brief instruction on the use of global assessment measures to quantify current functioning using wording consistent with the GAF scoring directions set forth by the American Psychological Association's Diagnostic and Statistical Manual, Fourth Edition (*American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders*, 2000) as a means of training the participants. Participants scored each of the 15 vignettes using the directions and GASF protocol in the presence of the principal investigator who was available to answer procedural questions.

Results

Descriptive statistics including the mean and range of scores generated by the CME were calculated using Microsoft Excel for Mac. Vignettes that were rated within 16 points of each other by all four raters met inclusion criteria for use in the study. Scores were recorded into a four-by-15 matrix to be evaluated for agreement. The goal of obtaining ten useable vignettes representing varying levels of student functioning was met. Twelve of the fifteen vignettes were scored within a range of 16 points by each of the four CME. Of those meeting criteria, two vignettes classified with mean scores in the below 30 range, six in the 31-70 range, and two in the above 70 range were selected to represent the various levels of functioning captured by the GASF. Table 4 contains the scores and descriptive data based on the CME ratings of the vignettes used in the study.

Table 4

Content Matter Expert GASF Scores for Vignettes Qualifying for Study

Vignette	R1	R2	R3	R4	M	Range*	How Used
Allison (95)	99	100	100	100	99.75	99-100	Practice
Alyssa (70)	71	78	71	75	73.75	71-78	Practice

Annie (58)	66	55	61	35	54.25	35-66	Eliminated
Braden (39)	31	42	35	45	38.25	31-45	Study
Danny (74)	75	78	80	80	78.25	75-80	Study
Hailey (52)	55	52	51	55	53.25	51-55	Practice
Issac (15)	22	21	21	10	18.50	10-22	Study
Jake (35)	31	32	31	26	30.00	26-32	Not Selected
James (45)	60	60	51	48	54.75	48-60	Not Selected
Kenny (72)	75	79	70	70	73.50	70-79	Study
Nico (95)	99	100	100	100	99.75	99-100	Study
Patty (56)	69	65	65	50	62.25	50-69	Eliminated
Steven (53)	50	58	50	50	52.00	50-58	Practice
Tiffany (42)	46	52	50	35	45.75	35-52	Eliminated
Tommy (24)	30	30	30	21	27.75	21-30	Practice

Note. Score in parenthesis next to student name represents researcher's assigned GASF score.

Phase 2: Inter-rater Reliability

Presently, no research has been conducted to assess the level of agreement among raters who use the GASF. The purpose of phase 2 of the present study was to assess the inter-rater reliability of raters rating short vignettes that were crafted based on actual students that the principal researcher treated while completing his internship. First through fifth grade teachers working in public schools in northwestern Michigan, and the school psychologists serving those schools, were contacted via email and asked to be participants in the study. After multiple contacts and attempts to secure the intended sample from this group, it was necessary for the researcher to extend the invitation for participation beyond northwest Michigan. As a result, the

researcher personally contacted individuals from his graduate cohort for assistance, asking them to complete study requirements. The study was ultimately populated with participants from southern Michigan, Upstate New York, and Pennsylvania.

Participants

Sixty-four school professionals were recruited to serve as GASF raters. The rater group was comprised of general education teachers (n = 36), special education teachers (n = 10), school psychologists (n = 15), and those who identified as *other* (n = 3). Of the three who identified as "other", one identified as a counselor/behavior specialist, one identified as a Response to Intervention coordinator, and one as an elementary school principal. Participants were primarily female (85.9%). 22% of respondents indicated age affiliation in the range 31-35 years old (more than half of the participants were under 40 years old). In terms of location, 92% of these professionals reported that they reside and work in the state of Michigan (n = 59). Two respondents were from New York, and three were from Pennsylvania. Of the 59 Michigan respondents, 54.7% work in a single school district in northwest Michigan. The largest level of experience in years, 32.8% of the professionals completing the study, indicated that they have 6 - 10 years experience (n = 21). Another 21% reported 3 – 5 years experience while 20% marked over 20 years experience. Table 5 provides a summary of the demographic data collected based on responses provided by the participating raters.

The sample size was determined based on the logic that this particular study is analogous to a multiple regression utilizing one independent variable (Greil, 2010). Using the sample size calculator, an alpha level of 0.05, an effect size of 0.15, and moderate power (0.8) were used to calculate the sample (Soper, 2011).

Table 5

Summary of Responses to Demographic Variables Provided by Raters

Variable	**Number**	**Percent**
Sex		
Female	55	85.9
Male	9	14.1
Age		
25-30	13	20.3
31-35	14	21.9
36-40	7	10.9
41-45	8	12.5
46-50	12	18.8
51-55	3	4.7
56-60	5	7.8
Over 60	2	3.1
State		
* Michigan	59	92.2
New York	2	3.1
Pennsylvania	3	4.7
Occupation		
Gen. Education	36	56.3
Sp. Education	10	15.6
Psychologists	15	23.4
Other	3	4.7

Experience		
3-5 years	14	21.9
6-10 years	21	32.8
11-15 years	8	12.5
16-20 years	8	12.5
Over 20 years	13	20.3

Note. 35 of the 59 raters from Michigan were represented by one school district in northwest Michigan.

Instruments/Measures

The Global Assessment of School Functioning (GASF).

The GASF was modified to reflect changes and suggestions based on information gathered during phase 1 of the present study and was used in this phase of the study (see Appendix F.)

Case Vignettes.

The ten vignettes drawn from the previous reliability pilot conducted by content matter experts were utilized. The vignettes were grouped according to levels of functioning (mean score below 30, 31-70, and 71-100). One vignette from the each of the upper and lower functioning ranges was randomly selected, as were three from the middle range, to be utilized in the practice/training section. Similarly five remaining vignettes were selected and served as the subjects of the main study using surveymonkey.com (Appendix G).

Procedures

Inter-rater reliability was measured using five practice and five rated vignettes that were the same for all participants. Following university institutional review board approval, the researcher utilized the following procedure in conducting phase two of the proposed study:

The researcher recruited participants via email. The email included an introductory letter that discussed the purpose of the study and a link to a training website which was located on the intermediate school district server. The training website contained a brief welcome statement, a description and rationale section, a downloadable copy of the GASF protocol, directions on the using the GASF, an example using the GASF, the training section, and a link to the survey. This training website was presented using the Moodle platform. Moodle is a free course management software system that educators use to create effective learning solutions. The Moodle site and technical assistance were provided by the intermediate school district where the principal investigator conducted his school psychology internship.

Section two of the training website contained the informed consent document (Appendix E). Participants were required to electronically sign a statement of informed consent that paralleled the language of the recruitment letter and satisfied the requirements of the human subjects committee at Alfred University.

Participants were instructed to read a brief introduction about the importance of teacher training on using the GASF to rate student behavior, how the GASF was developed, and how student behavior is operationalized within the GASF (i.e., the domains of attendance, academic quality, work completion, social functioning/peer relationships, and disruptiveness).

The participants were asked to open attachments formatted in Microsoft Word (or in .pdf) that included directions for using the GASF and the actual GASF protocol. Participants

exercised the option to either print a hard copy of the measurement tool (GASF) or access the measurement tool on the computer desktop using Microsoft Word or .pdf reader.

Next, participants were instructed to complete the training quiz found on the Moodle site. Teachers and school psychologists were asked to read and score each of the five practice vignettes using the GASF protocol. Directions for using the GASF accompanied each vignette as a reminder. Participants received immediate feedback on each score. Feedback was presented in terms of "correct" or "incorrect". Correct responses were accompanied by the mean (as determined by CME ratings) and range (plus or minus eight points from the mean) of acceptable scores. Incorrect responses were accompanied by a reminder for respondents to carefully read the directions and vignette. Participants were given a second chance to make a correct response that fell within the acceptable range. Second chance scores that fell within the acceptable range were marked as correct. Scores that were incorrect after the second try were marked as incorrect. In both cases, the participant moved to the next question or the next section of the study in the case of the final question. The decision to use a range of +/- eight stemmed from the researcher's graduate course and clinic experience. The researcher was exposed to training using the GAF and was expected to rate cases +/- ten points from the instructor's target score (N. Evangelista, personal communication, February 23, 2015).

Finally, participants were asked to open the link to the survey located on surveymonkey.com (2011). Surveymonkey is a web based software program that allows for the development, distribution, and analysis of survey results in a format that meets the standards set forth by Institutional Review Boards.

The survey was comprised of a "rater questionnaire" section (e.g., age, gender, job description, years of service, years of training/degree earned), the vignettes section, where

participants provided a rating to five vignettes using the GASF, and a feedback section where participants were asked to provide feedback on the instructions, the measure's vocabulary, ease of use, efficiency, and whether this brief global measure, if reliable, would be useful in helping them quantify student behavior. Eligible raters who completed the requirements of the study were given the option to be entered into a drawing to win a new Apple iPad.

Analysis

Data was collected and entered into the Statistical Package for Social Sciences (SPSS v.22) software and analyzed using intra-class correlation (ICC) to assess the inter-rater reliability. The guidelines for choosing the appropriate form of the ICC suggest that the researcher consider whether a one-way or two-way analysis of variance (ANOVA) is appropriate for the analysis of the reliability study, if the judges mean ratings are relevant to the reliability of interest, and if the researcher will use the mean of several ratings or treat each rating individually (Shrout & Fleiss, 1979). The variants and means of interpretation are based on four major factors as determined by the study methodology based on the guidelines set forth by McGraw and Wong (1996). These four factors are summarized by the following four statements, Hallgren (2012):

1. A one-way or two-way model for the ICC is selected based on the way coders are selected for the study.

2. The researcher must specify whether consistency or absolute agreement characterize good inter-rater reliability (IRR).

3. The researcher must declare the unit of analysis (either consistency or average ratings) that is to be interpreted for the ICC.

4. Coders selected for the study are determined to be either fixed or random effects based on their selection and whether the results may be generalized to a larger population.

The researcher utilized a two-way model based on the determination that all raters would score each of the five vignettes. When considering the appropriate means for establishing good IRR, the researcher primarily focused on absolute agreement. Agreement is thought to be more appropriate in terms of this study because ultimately in practice, the absolute value of the rating made by an individual is expected to represent the true score of the student. Additionally, the researcher included the two-way consistency model in the instance that rank ordering may be considered useful. The unit of analysis most relevant to this study, the single measures ICC, was used for interpretative purposes for each of the calculated ICCs. The GASF has been developed to be a tool that can be used by a single teacher, psychologist, or case manager to screen (baseline/benchmark) and progress monitor students. While there is potential for the GASF score to be the product of a group rating, its predecessors in the mental health arena, the CGAS and GAF are predominately scored individually. Finally, the random effects model was employed based on the researcher's decision to select a random sample of raters from a larger population of raters while assuming that the ratings from the sample generalize to the larger population of potential raters. To summarize, IRR was assessed using a two-way random, absolute agreement, single measures ICC (McGraw & Wong, 1996). The shorthand associated with this model is represented as ICC (A,1), whereas the "A" represents "agreement" and the "1" indicates that single measures is used for interpretation (McGraw & Wong, 1996). This is synonymous to the ICC (2,1) model presented by Shrout and Fleiss (1979), whereas the "2" indicates that each rater rates each vignette and is considered representative of a larger population, and the "1" indicates that reliability is calculated based on single measures.

Results

Descriptive statistics for the ratings given by the 64 raters were calculated using SPSS v.22, based on responses for each of the five vignettes. The mean, standard deviation, and range for each vignette can be found in Table 6. In addition to these measures, Table 4 includes the "Target Score" which was based on the principal investigator's gold standard rating of the five vignettes prior to the study.

Table 6

Means, Standard Deviations, Ranges, and Relationships of Ratings to the Target Score for Each Vignette of the 64 Raters

Vignette	Target Score	*M*	*SD*	Range	+/- 5 Percent	+/- 10 from	>10 target
Kenny	72	76.80	5.449	56-90	53	86	14
Braden	39	45.42	10.202	29-80	42	70	30
Isaac	15	15.44	7.645	1-36	59	89	11
Danny	74	72.50	7.167	51-85	64	89	11
Nico	95	98.03	2.330	90-100	100	100	0

Note. The target score was derived from the principal investigator's ratings of the subjects.

The researcher was also interested in any differences between subgroups of raters. School personnel ratings were categorized into four groups – general education teachers, special education teachers, school psychologists, and "other", which was a small group of three raters comprised of two elementary school principals and another who identified as a behavior specialist. Table 7 represents the resulting descriptive statistics for the mean, standard deviation and range based on the subgroups of raters affiliated by occupation. Furthermore, this table compares the means, standard deviations, and ranges based on ratings provided for each of the five vignettes. The researcher was primarily concerned with identifying any irregularities

regarding ratings based on occupation and to identify if one subgroup of raters scored consistently higher or lower than other subgroups in regard to a particular vignette.

Table 7

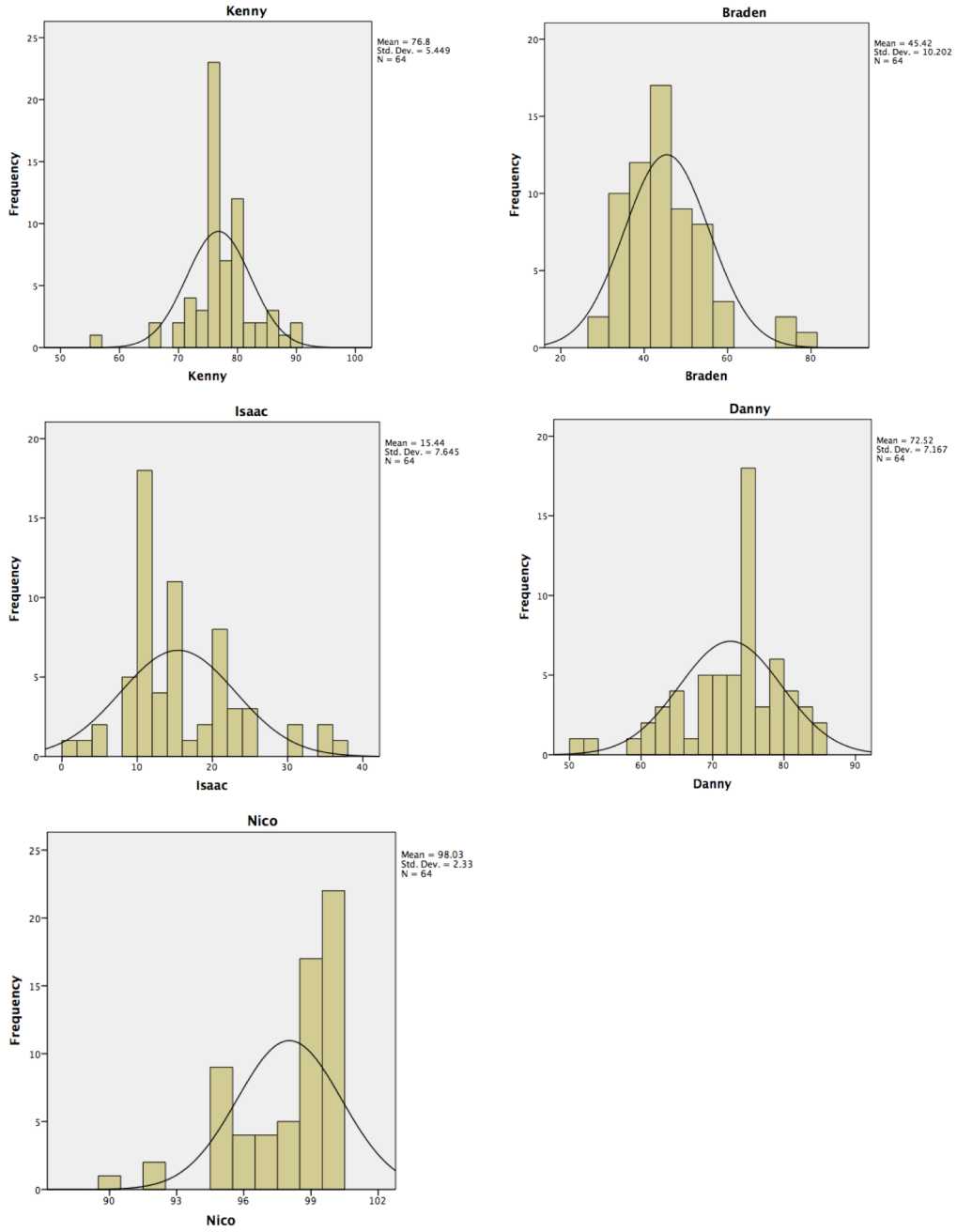
Means, Standard Deviations, Standard Errors, Confidence Intervals, and Ranges for Five Vignettes Based on Subgroup Occupation

						95% CI. for $\overline{X}$			
		N	$\overline{X}$	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Min	Max
Kenny	Gen. Ed.	36	76.19	5.285	.881	74.41	77.98	56	89
	Sp. Ed.	10	80.40	5.719	1.809	76.31	84.49	72	90
	Psych.	15	75.20	4.902	1.266	72.49	77.91	65	80
	Other	3	80.00	5.000	2.887	67.58	92.42	75	85
	Total	64	76.80	5.449	.681	75.44	78.16	56	90
Braden	Gen. Ed.	36	43.78	9.100	1.517	40.70	46.86	32	76
	Sp. Ed.	10	52.30	14.863	4.700	41.67	62.93	38	80
	Psych.	15	45.93	7.941	2.050	41.54	50.33	29	60
	Other	3	39.67	7.572	4.372	20.86	58.48	31	45
	Total	64	45.42	10.202	1.275	42.87	47.97	29	80
Isaac	Gen. Ed.	36	13.97	6.381	1.063	11.81	16.13	1	37
	Sp. Ed.	10	17.00	8.380	2.650	11.01	22.99	5	35
	Psych.	15	19.73	8.293	2.141	15.14	24.33	8	35
	Other	3	6.33	4.163	2.404	-4.01	16.68	3	11
	Total	64	15.44	7.645	.956	13.53	17.35	1	37

						95% CI. for $\overline{X}$			
		N	$\overline{X}$	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Min	Max
Danny	Gen. Ed.	36	73.67	7.155	1.193	71.25	76.09	52	85
	Sp. Ed.	10	70.70	8.957	2.833	64.29	77.11	51	79
	Psych.	15	70.00	4.175	1.078	67.69	72.31	62	75
	Other	3	77.33	10.786	6.227	50.54	104.13	65	85
	Total	64	72.52	7.167	.896	70.73	74.31	51	85
Nico	Gen. Ed.	36	97.78	2.631	.438	96.89	98.67	90	100
	Sp. Ed.	10	99.00	1.491	.471	97.93	100.07	95	100
	Psych.	15	98.20	2.042	.527	97.07	99.33	95	100
	Other	3	97.00	1.732	1.000	92.70	101.30	95	98
	Total	64	98.03	2.330	.291	97.45	98.61	90	100

Figure 1 represents histograms based on the 64 - rater sample. Each vignette includes the frequencies and the bell curve of the distribution.

Figure 1. Histograms Raters GASF Scores for Each Vignette



Inter-rater Reliability and the GASF

Inter-rater reliability (IRR) was assessed using a two-way random, absolute agreement, single measures intraclass correlation (ICC A,1). A two-way random, consistency, single measures intraclass correlation (ICC C,1) is included as well. The inclusion of the ICC (C,1) is presented in the event that systematic differences exist between subgroups. High ICC values indicate a strong level of IRR with 1 equal to perfect agreement and 0 indicating random agreement. Negative ICC nearing -1 are indicative of systematic disagreement (Hallgren, 2012). Each of the ICC tables below include intraclass correlations representing both the single and average measures indexes. Single measures ICCs represent an index of each vignette's score as determined by a single rater. The average measures ICCs represent an index of scores derived from the average of multiple raters scores.

Table 8 summarizes the calculated ICC for the two-way absolute agreement, single measures model (.877) was in the substantial range (Shrout, 1998), indicating that raters had a high degree of agreement and that 87.7% of the observed variance of a single rater is true variance (Landers, 2011).

Table 8

Intraclass Correlation From School Personnel Sample Using a Two-Way Random, Absolute Agreement Definition

	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.877	.715	.983	478.001	4	252	.000
Average Measures	.998	.994	1.000	478.001	4	252	.000

Note. Two-way random effects model where both people effects and measures effects are random.

Similarly, the calculated ICC for the two-way random, consistency, single measures model, 0.882, was in the substantial range (Shrout, 1998), which implies that coders mean ratings had a high degree of agreement and suggests that vignettes were rated similarly across raters (Table 9). The high ICC suggests a minimal amount of measurement error was introduced as a result of the independent coders; therefore, statistical power for subsequent analyses is not substantially reduced. Therefore, the GASF ratings for the purpose of testing the reliability of this measure appear suitable for this experiment.

Table 9

Intraclass Correlation From School Personnel Sample Using a Two-Way Random Consistency, Single Measures Definition

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.882[a]	.724	.984	478.001	4	252	.000
Average Measures	.998	.994	1.000	478.001	4	252	.000

Note. Two-way random effects model where both people effects and measures effects are random.
a. The estimator is the same, whether the interaction effect is present or not.
b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

The researcher was also interested in the IRR of those representing the occupational subgroup. Separate ICC (2,1) was calculated and interpreted using absolute agreement as a means to determine how similar coders ratings were in absolute value. Single measures ICCs for each of the occupational subgroups were in the excellent range (> .90)

Table 10.

Intraclass Correlation Coefficients by Subgroup Occupation Using a Two-Way Random, Absolute Agreement, Single Measures Model.

		95% Confidence Interval		F Test with True Value 0			
	Intraclass Correlation	Lower Bound	Upper Bound	Value	df1	df2	Sig
General Education							
n = 36							
Single Measures	.963[a]	.900	.995	1002.636	4	140	.000
Average Measures	.999	.997	1.000	1002.636	4	140	.000
Special Education							
n =10							
Single Measures	.922[a]	.790	.990	133.031	4	36	.000
Average Measures	.992	.974	.999	133.031	4	36	.000
Psychologists							
n = 15							
Single Measures	.963[a]	.898	.995	424.885	4	56	.000
Average Measures	.997	.992	1.000	424.885	4	56	.000
Other							
n = 3							
Single Measures	.968[a]	.815	.996	167.361	4	8	.000
Average Measures	.989	.930	.999	167.361	4	8	.000

Note. Two-way random effects model where both people effects and measures effects are random.

After comparing the reliability of the ratings provided by school personnel, the researcher questioned the relationship between the ratings of raters based on occupation for each of the five vignettes and whether any significant differences existed between groups based on any of the vignettes. A one-way analysis of variance (ANOVA) was conducted using SPSS 22. Findings indicated a statistically significant difference between groups as determined by a one-way ANOVA (F (3,60) = 4.106, p = 0.010) for the vignette "Isaac". A Tukey post hoc test revealed that ratings based on occupation were indeed significant for the subgroup "psychologists" (19.73

± 8.3) compared to the subgroup "Other" (6.33 ± 4.16) $p = 0.022$. Psychologists tended to rate Isaac higher than the subgroup "Other". Furthermore, post-hoc analysis indicated that general education teachers tended to score the vignettes lower than special education teachers with the exception of the vignette labeled "Danny". Table 11 represents the findings based on the one-way ANOVA comparing ratings of subgroup occupations for each vignette. Appendix I contains results of the Tukey post-hoc analysis for subgroups based on the vignette "Isaac" and accompanying vignette/group comparisons.

Table 11.

One-Way ANOVA Results for All Vignettes Based on Occupation

		Sum of Squares	df	Mean Square	F	Sig.
Kenny	Between Groups	211.920	3	70.640	2.556	.064
	Within Groups	1658.439	60	27.641		
	Total	1870.359	63			
Braden	Between Groups	673.687	3	224.562	2.290	.087
	Within Groups	5883.922	60	98.065		
	Total	6557.609	63			
Isaac	Between Groups	627.178	3	209.059	4.106	*.010
	Within Groups	3054.572	60	50.910		
	Total	3681.750	63			
Danny	Between Groups	245.218	3	81.739	1.640	.190
	Within Groups	2990.767	60	49.846		
	Total	3235.984	63			
Nico	Between Groups	15.315	3	5.105	.938	.428
	Within Groups	326.622	60	5.444		
	Total	341.938	63			

Note. Sig. = Significance $p < .05$

School Personnel Perceptions of the GASF

While the primary focus of this investigation was to gather data regarding the technical properties of the GASF, the researcher was also interested in the perceptions of the individuals

who would ultimately be using the instrument on a daily basis. Participants were asked to respond to the following five statements regarding usability of the GASF:

Statement 1. The GASF is worded clearly.

Statement 2. The hierarchy of levels is a fair representation of behavior in global terms.

Statement 3. A sufficient range of behavioral descriptors is provided within and between levels.

Statement 4. Items within levels are appropriately placed in terms of intensity and severity.

Statement 5. My training and experience have provided me with necessary skills to utilize this tool.

Participants responded by selecting from four choices – strongly disagree, disagree, agree, strongly agree. Over 90% of the participants endorsed that they either "strongly agreed" or "agreed" with each of the five statements. Table 6 represents the percentages and frequencies of the responses supplied by the participants for each of the five statements.

Based on this sample, one may speculate that school personnel possess solid baseline levels of comfort and understanding of the GASF. It appears that this group endorses a level of confidence in the measure that may suggest a willingness to utilize the GASF in school settings as a component to a larger problem-solving model.

Table 6.

School Personnel Responses to Questions About the Usability of the GASF

Questions/Responses	Frequency	Percent	Cumulative Percent
Wording			
Strongly Disagree	0	0	0
Disagree	6	9.4	9.4
Agree	38	59.4	68.8
Strongly Agree	20	31.3	100.0
Total	64	100.0	

Hierarchy			
Strongly Disagree	1	1.6	1.6
Disagree	2	3.1	4.7
Agree	41	64.1	68.8
Strongly Agree	20	31.3	100.0
Total	64	100.0	
Descriptors			
Strongly Disagree	0	0	0
Disagree	6	9.4	9.4
Agree	41	64.1	73.4
Strongly Agree	17	26.6	100.0
Total	64	100.0	
Levels			
Strongly Disagree	0	0	0
Disagree	4	6.2	6.2
Agree	40	62.5	68.8
Strongly Agree	20	31.3	100.0
Total	64	100.0	
Training			
Strongly Disagree	1	1.6	1.6
Disagree	3	4.7	6.3
Agree	44	68.8	75.0
Strongly Agree	16	25.0	100.0
Total	64	100.0	

Chapter 4: Discussion

This study represents the initial psychometric testing performed on a new measure, the Global Assessment of School Functioning. The purpose of this study was to investigate three stated research questions: 1. Does the GASF demonstrate elements of content validity, 2. Can school personnel be trained to utilize the GASF to accurately quantify student behavior, and 3. Does the GASF demonstrate inter-rater reliability based on school professionals GASF scores. Based on responses from Content Matter Experts and analysis of data relating to this particular study, there is evidence that affirmatively supports each of these questions.

Question 1: Does the Global Assessment of School Functioning (GASF) possess adequate content validity as assessed by an expert panel?

In addition to paper pencil response to the above question, the principal researcher facilitated discussion based on responses provided by the expert panel. While care was taken to limit researcher bias, the researcher did interact with the panel. Phase 1 examined whether the GASF appeared to possess adequate content validity as assessed by an expert panel comprised of teacher consultants who were employed by an intermediate school district in Northern Michigan. The results of this portion of the study suggest that the GASF possesses adequate face and content validity. The structure of the instrument closely resembles the Global Assessment of Functioning, Axis V of the Diagnostic and Statistical Manual for Mental Disorders, Fourth Edition, Text Revision (2000).

Responses to questions regarding the properties of the GASF were predominately positive. Each of the Content Matter Experts (CME) responded *Yes* to questions relating to whether the behaviors are defined in global terms and whether the GASF behavioral descriptors comprised a hierarchy of behavior and represented incremental change in behavioral severity.

Teacher consultant responses to the content validity protocol strongly endorsed the hierarchal structure of the GASF. When asked to rank order the ten descriptive anchors, the teacher consultants were able to accurately order these descriptors perfectly with the exception of one reversal from one of the four raters. As expected, the anchors at the extremes of the scale were more easily ranked than those representing the descriptive anchors falling in the middle range. The one reversal came via one rater reversing the fifth and sixth descriptive anchors.

In addition to numerically ranking the descriptors to form the behavioral hierarchy, the teacher consultants were asked to provide feedback as to the utility of the GASF with school personnel. One rater expressed concern that teachers may not clearly understand the severity of behavior. Namely, this rater wondered if teachers would be able to recognize progress towards goals, especially for students identified with a disability v. those without an identified disability. This question is particularly important to the researcher as results may suggest that individuals may unwittingly be weighting aspects of behavior captured within the GASF or perhaps more importantly, that some school personnel may exhibit bias in their scores depending on the eligibility/identification status of a child. Specifically, might school personnel rate students identified with a special education disability differently from those not receiving special education services? This question will require further study and should be extended to include analysis of ratings provided to students of different race, ethnicity, and socioeconomic status. Similarly, another teacher consultant suggested placing the descriptors into tiers for Response to Intervention models and that doing so would be helpful in describing behavior. As GASF research is extended, this suggestion may be addressed with predictive validity studies that identify anchor points within the GASF that correspond to the intervention tiers associated with response to intervention models.

In terms of whether the instrument contains an adequate content sample of student behaviors, three of the four experts agreed that the GASF contained an adequate sample. The fourth questioned how to score students who are at grade level academically but do not hand in work. The question prompted the group to the re-examine the GASF. After review, the group concluded that a student who is at grade level but is not turning in work would likely fall between 51 – 60 or 61 – 70 depending on the accompanying information. This dialogue seemed to satisfy the teacher consultant who originally posed the question. Furthermore, the panel as a whole agreed that reminding individuals who score students using the GASF to carefully consider reading below the actual scoring anchor point might be important in future training of school personnel. Recognizing the global nature of the instrument and that multiple facets of behavior contribute to the overall functioning score of the student is important to providing accurate GASF scores. Following the directions and focusing on finding the appropriate range, reading down to ensure that the rater is indeed at the lowest level of behavioral functioning represents the key to obtaining not only accurate results, but also consistent reliable scores.

Question 2: Can school professionals, namely teachers and school psychologists, be adequately trained to utilize the GASF to quantify behavior?

Despite anecdotal information related to whether school personnel can be trained to use the GASF, at this time the researcher cannot endorse that this is the case. The results of this study do indicate that there is potential for this training to prove effective. Each of the four Content Matter Experts agreed that school personnel can be trained to use the GASF to score student behavior. Not only did CME endorse teachers' ability to utilize the GASF, comments were made that teachers would indeed use the GASF based on its ease of use, convenience, and

the expectation that scores can be ascertained and recorded quickly. Anecdotal information from a handful of teachers who completed the study indicate that the GASF may be useful for progress monitoring students identified in a multi-tiered system or those students identified with special education disabilities. Based on information provided by content matter experts, accompanied by robust statistical results related to inter-rater reliability, it seems that school personnel can indeed be trained to utilize the GASF to quantify student behavior. This statement must however be interpreted cautiously. While data provided indicates that there is potential for training school personnel, presently no training procedures and protocols exist.

GASF training modules should be explored as a means of ensuring school personnel gain competencies necessary to make accurate, informed student ratings. Standardized procedures that closely follow the instructions for making GASF ratings represent the basic prerequisite for utilizing the measure. Initial formats for training may mirror the researcher's graduate school training received in-class. As a course requirement, students were required to read vignettes weekly and provide GAF scores as part of a greater multi-axial diagnosis. Diagnoses and scores were discussed and compared in a professor led, seminar style course. Student scores were compared to the professor's target scores. Student scores that extended beyond a range of +/- 10 points of the professor's target score were considered incorrect. While the researcher cannot recall the number of attempts per vignette, or the number of vignette exposures, it is hypothesized that practice and discussion represent integral components to a GASF training program.

A simple search of training methodologies for making GAF ratings turned up very little information. Two studies, one from Africa and one from Norway acknowledged training led to more accurate GAF ratings based on test-retest data. The study conducted in Uganda, Africa

found that a one-hour training yielded improvement in GAF scores for a group of medical assistants as compared to a gold standard score (Abbo, Okello, & Nakku, 2013). While the study indicated improvement in ICC correlations, it is unclear what constituted training for the medical assistants. Alternatively, the study conducted in Norway, a web based training, references an expanded manual that included information on rating guidelines, as well as additional information regarding symptom and function scales (Valen et al., 2015). The web based training provided immediate feedback to the rater based on the ratings they provided. Furthermore, the technology provided for analysis of whether the rater was *too strict or too kind* in their ratings (Valen et al., 2015). This is similar to what the present study tried to accomplish with the training component that was presented to the raters using GASF. Both the Africa and Norway study indicate that practice and feedback result in improved reliability ratings, and more importantly, rating accuracy.

Sensible venues for practice and discussion include undergraduate and graduate training programs, professional development in-service training programs, off-campus breakout training sessions, conference sessions, or through web-based instruction where groups interact cooperatively. At minimum, the need for a training manual consisting of multiple practice vignettes appears warranted. Consideration to decision trees for making ratings may be explored. Further consideration should be paid to questioning and data gathering.

Question 3: Does the GASF demonstrate adequate reliability as measured by an examination of inter-rater reliability?

The calculated intraclass correlation (.877 single measures using the absolute agreement model) would suggest that the GASF possesses excellent inter-rater reliability. While the

researcher was primarily interested in the reliability of the sample in rating the vignettes, it is also important to explore the reliability of single raters if the GASF is at times to be used by individuals. Similarly, the single measures correlation (.882 utilizing consistency model) is excellent. This would suggest that individuals can score vignettes reliably with over 95% of the variability in scores captured by the measure itself.

Despite the strength of the correlational statistic, this metric should be interpreted cautiously. The limited number of vignettes (only five were scored) appears to contribute to the high ICC. Considering the extensive range obtained for four of the five vignettes, it is necessary to consider the circumstances for the variability in these ratings. The fifth vignette (Nico) was based on a fictitious case that would likely not be encountered for intervention or support. Despite the substantial range, mean scores are consistent with what one would expect for each of the five vignettes. Sample raters on average scored vignettes slightly higher (4 points) than projected scores provided by the principal investigator. Furthermore, standard deviations for four of the five vignettes were very close to the target scores established by the researcher (< +/- 5 points of target score).

Questions regarding ratings at the extreme ends of the ranges for vignettes (Kenny, Braden, Isaac, and Danny) were not easily explained. Scores indicated at the tails of the histograms were not associated with individual raters; rather, the extreme scores were single items from six separate raters. For example, rater 36 scored Braden, Isaac, Danny, and Nico within five points of the mean, but scored Kenny (56) twenty points below the mean. Similarly, rater 24 scored Kenny, Braden, Isaac, and Nico within the acceptable range, but scored Danny twenty-one points below the mean (51). The only exception came from rater 20 who scored both Kenny (90) and Braden (80) higher than would be expected.

If the reader will recall, phase 2 of this study, was constructed to mitigate the chances of error attributed to the vignettes intended for scoring. However, the vignette coded *Braden* represented the source of the largest score variability as evidenced by a range of 51; 30% of participants scored Braden more than 10 points over the target score. Furthermore, the standard deviation for this particular vignette was larger than other study vignettes. It may be possible that Braden represents a "bad vignette". The wording for this particular vignette may have been confusing or otherwise unclear to the reader. It should be noted, however, that teacher consultants scored this particular vignette within +/- seven points (M = 38.25, SD = 6.397) of the researcher's expected score. This level of agreement qualified the Braden vignette for inclusion in the study.

It is possible that extreme scores were the product of data entry error on the part of the raters. Reversals of numbers (e.g., 56 instead of 65) or accidental key strokes are not uncommon in some coding activities, and without careful attention to the task, the possibility for these types of errors exist. For example, simple data entry errors were encountered during data analysis which were easily corrected by the researcher. On two occasions, participants used the letter "O" instead of the zero for their numeric representations. Letters to numbers were explainable and represented sensible corrections. The same cannot be said for digit coding that may or may not represent errors. Subsequent studies may be better served to employ the use of data collection assistants who can personally, and without bias, aid in recording school personnel responses as opposed to relying on responses that are recorded via a large, impersonal database that may not intuitively sense participant error.

After analyzing the response patterns, the researcher accessed the "time spent" metric found in the respondent information section found in the individual response section of the

survey on surveymonkey.com. While the time taken on the entire survey was minimal, time on task did not appear to be a factor or determinant of the extreme scores. The average time spent taking the survey (based on the time stamp for 55 of the participants used in the survey) was just over 12 ½ minutes with a minimum of four minutes and a maximum of 35 minutes. Raters who scored vignettes at the extreme low and high ranges were similar (rater 20, 10 minutes; rater 24, 14 minutes; rater 36, 8 minutes). It is also possible that the extreme ratings were the product of fatigue or distraction (or a lack of focus and/or effort) as these ratings were most frequently scored in evenings after the dinner hour.

Regardless of the reasons for the ratings found at the outer limits, the GASF ratings from the sample as a whole would suggest that there is potential for school personnel to accurately utilize the GASF to score students with additional training. This study suggests that one of the benefits of using the GASF is that it takes very little time for school personnel to score student behavior. While this may indeed be the case the results from this study may suggest that research correlating time-on-task to scoring accuracy may be warranted. Is there a critical mass for how much time can be spent rating students before fatigue leads to score degradation? If a teacher is rating 30 students and each student takes the teacher 5 minutes to score, does fatigue set in at the two to two and a half-hour mark?

Observation regarding the ranges was primary to the researcher's decision to generate the one-way ANOVA statistic. Despite concerns over the range specific to the Braden vignette, analysis of the means regarding this particular vignette did not yield findings of statistical significance. However, the one-way ANOVA did identify statistically significant differences in mean ratings for the *Isaac* vignette, namely between the *psychologists* and *other* subgroups. This phenomenon may be of particular interest when considering the occupation and level of

behavioral disturbance represented by this particular vignette. It is difficult to speculate why statistical significance exists between these two groups. While the researcher cannot say with certainty, it may be hypothesized that school administrators (the primary respondents from the *other* group) may be more likely to score students with externalizing behaviors lower than psychologists. This may be due to the likelihood that school administrators are to a higher degree concerned with the safety of the school population at-large than psychologists who may be more likely to be more empathic and tolerant to individual student behavior. It is also possible that the scoring differences are the product of how school administrators and psychologists weight labeling diagnoses. It is possible that diagnostic labels (emotionally impaired, conduct disordered, major depression, etc.), and the manner in which they are interpreted, may affect scoring.

Regardless of the factors influencing vignette (or more importantly student) ratings, the researcher hypothesizes that in some cases raters may not be scoring vignettes accurately. It is imperative that raters recognize that the directions for the GASF require the rater to begin at the top of the scale and continue moving down the scale until the best descriptive range for the student is found. This level indicates the student's behavioral severity OR the level of functioning over the past month. The rater is then reminded to consider the range beneath the previously determined range to ensure against prematurely stopping. This range should be deemed too severe both in terms of severity AND functioning. If this range is indeed too severe, the previously determined range is accurate. The key is that raters are identifying the lowest rating over the past month for the student.

School Personnel Perceptions of the GASF

While the primary focus of this investigation was to gather data regarding the technical properties of the GASF, the researcher was also interested in the perceptions of the individuals who would ultimately be using the instrument on a daily basis. Participants were asked to respond to statements about the GASF's wording, hierarchical structure, behavioral ranges and descriptors, placement of behaviors in levels in terms of intensity and severity, and raters' training and experience in terms of using the GASF.

Participants responded by selecting from four choices – *strongly disagree, disagree, agree, strongly agree*. Over 90% of the participants endorsed that they either strongly agreed or agreed with each of the five statements. Based on this sample, one may speculate that school personnel possess solid baseline levels of comfort and understanding of the GASF. It appears that this group endorses a level of confidence in the measure that may suggest a willingness to utilize the GASF in school settings as a component to a larger problem solving model.

Limitations

The purpose of this research was to embark on an initial investigation of the GASF. In doing so, this study was able to provide valuable information related to the technical properties of the measure. While this study was successful in gathering and sharing data on the reliability and content validity, there are several limitations to the present research.

The proposed method for this study indicated that 64 school professionals from a northern Michigan would be sought to complete the study utilizing surveymonkey.com. Despite multiple attempts and contacts via email and telephone contacts to participating teachers, school psychologists, and principals, it was necessary to cast a wider net to obtain 64 participants who completed the study with acceptable integrity. As a result, it was necessary to extend to other

areas in Michigan, and ultimately to members of the principal investigator's graduate cohort practicing beyond the state of Michigan. This sample should be considered neither random nor representative of a greater population of raters.

In terms of generalizability, the results of the present study should be interpreted with caution. In order for the GASF to be deemed a useable measurement tool, elements of external validity will need to be examined more rigorously. The majority of the participants included in the results were drawn from rural and suburban school districts in Northwest Michigan. This geographic area tends to be racially and ethnically homogenous, and while there is economic diversity, this area is not considered otherwise diverse or representative of the potential global population likely to use the GASF. Despite these limitations, the sample that was collected appears to be representative of school populations in terms of age, gender, experience, and occupation. Furthermore, the sample reporting from the initially proposed Northern Michigan region schools, participants' sex, age, experience, and occupation comprises a good representative sample of the demographic.

Training v. exposure: While the researcher attempted to provide an element of training to the participants rating the vignettes, it may be more accurate to identify that which was purported to be "training" as "exposure". Participants were instructed on the use of the GASF and provided opportunities for practice; however, the participants were not given an opportunity to ask questions or receive clarifying statements regarding the vignettes. Furthermore, answers that were first scored as "wrong" were not accompanied with explanation or further instruction. Instead, the participant was prompted to simply try again. After a second try, the participant was permitted to move to the next vignette whether the second response was correct or not. This limitation was primarily related to the nature of the research method which was meant to allow

the participants the ability to complete the tasks at their pace and convenience, assuming that time on task would be a prohibitive factor in collecting a useable sample of respondents. A handful of responses were extremely inconsistent. For example, the raters tended to score vignettes Kenny and Danny similarly. However, two raters evidenced a > 20 point difference in their ratings on these vignettes. Another rater scored Kenny 56, three standard deviations below the mean, but scored all other vignettes within normal limits.

A more comprehensive training program that included face-to-face or webinar style interaction may have been preferable insofar as it would have permitted the researcher and participant to engage in constructive dialog that would not only have permitted the participant to make more accurate ratings, but also for the researcher to assess the level of participant understanding. To this end, the author is very interested in using *Smart Board* or other interactive technology to facilitate training. The researcher hypothesizes that training methods utilizing a facilitator who can easily check for understanding based on ratings provided would be able to remind individuals to follow directions when making inaccurate ratings. Identifying clusters of inaccurate ratings would also allow the facilitator to discuss the nuance of making ratings (e.g. remind the rater to consider the five facets of behavior that are being considered). Furthermore, the opportunity to assess and generate "class reports" could provide valuable information relating to difficulties training participants may be experiencing.

As research is continued, increasing the overall sample of vignettes or actual study cases – students who the GASF was designed to assess is important to generalizing results. At the onset of this study, the researcher identified the need to have a large sample of teachers, school psychologists, and other school personnel to provide ratings for the vignettes. What the researcher failed to identify was the need for a larger sample of vignettes for the rater to score.

In retrospect, the method could have been modified to require multiple groups of raters to rate multiple sets of vignettes. For example, rater group A rates vignette group A, rater group B rates vignette group B, etc. or the researcher could have allowed for some randomization of vignettes and raters. The present design was selected to limit the level of variability and to control for any rater/vignette interaction.

Furthermore, the final group of vignettes contained only males. This was simply the result of chance selection that took place when the vignettes were taken from the larger 15-item sample of vignettes. Based on the disproportionate referral rate of males to females in the public schools where the material for the vignettes was created, males comprised the majority of vignettes that were created. As a result, the GASF appears to possess excellent reliability relating to males, but the researcher cannot make any inferences as to the GASF's reliability measuring females.

Implications for Practice

Despite its limitations, the GASF represents a potentially useful measure of student functioning. Based on the present study's findings, school professionals can use the GASF to reliably rate student behavior. Its structure allows school personnel to quantify student behavior efficiently without the need for technically cumbersome scoring procedures.

In some instances, namely in intensive alternative education behavior programs, the GASF may have the ability to function as a universal screening and benchmarking tool insofar as it can provide a baseline rating and measure of change. As research on the GASF is extended, score thresholds may be identified that inform placement decisions. For example, scores below 30 may indicate a student requires placement into a highly restrictive environment. Similarly, students who are placed in a highly restrictive environment and demonstrate growth and gains

(scores above 40 perhaps) will be considered for reintegration into less restrictive environments. While universal screening tools have historically been used to assess targeted skills within a curricular domain (e.g. fluency in reading), screening may be used in alternative education setting with behavioral foci. Jenkins (2003), suggests universal screeners posses elements of sensitivity, specificity, practicality, and consequential validity. As research extends to criterion related validity, sensitivity and specificity may be addressed and indicate that particular anchor points may be indicative of continued school problems and at other levels exit from alternative programs to a less restrictive environment. Furthermore, the rating may be useful in manifestation determination meetings for students being considered for placement in more restrictive environments. In terms of practicality and consequential validity (the measure does not harm the student), the GASF appears promising.

Since the genesis of the GASF is rooted in mental health progress monitoring and so closely conforms to the framework of the GAF, the GASF may represent an opportunity for schools and mental health care providers to communicate student progress between channels in a manner that is meaningful to both entities. This would be especially true if future research indicates positive correlations between GASF and GAF scores. While it may not be necessary for the scores to match completely, if it is identified that both GASF and GAF scores increase or decrease similarly on case basis, the GASF could certainly become a useful method of communicating progress in a manner that could strengthen the relationship and quality of care between schools and clinics.

For over 30 years, mental health professionals used the Global Assessment of Functioning (GAF) score, Axis V of the Diagnostic and Statistical Manual of Mental Disorders to quantifiably establish a benchmark score and subsequently use the GAF as a progress

monitoring metric for patients. With the growing attention given to standardized, high stakes testing, and the pressure for school personnel to demonstrate student improvement, especially in the areas of reading and math, teachers have expressed frustration and concern over their inability to practice the "art of teaching". As a global screening tool, the GASF may serve to bring some balance to student measurement with the potential to serve as an evidence-based, quantitative measure of the whole child. Ideally, the GASF represents a method of assessing student progress as a part of greater Response to Intervention process. Initial ratings could be recorded during child study meetings or as part of a similar problem solving structure. Students receiving special education services or those identified as Tier II or Tier III needs based students may be assessed during benchmark periods or as deemed necessary by the established progress-monitoring schedule.

In the public health arenas, assessment measures have been used successfully report changes in patient symptomology. As research is extended, the GASF may indeed be a useful tool for similar use in the school setting. Research and practice in problem solving in reading and math, and to a lesser extent behavior, has become more established. Those administering curriculum based measures can demonstrate growth academic skills like reading fluency, digits correct, and reductions in office discipline referrals, but there is not a quantifiable method that simply and accurately states whether the whole child is getting better or worse. In instances where reading, math, and/or discipline interventions are not progressing at a rate that was expected, is there value in identifying that overall, a student is "doing better" within the school environment? The GASF may possess the potential to demonstrate that students are indeed making gains in being students. If future research on the GASF can demonstrate that it possesses additional elements of validity, the GASF may indeed fill a gap in student assessment.

At very least, the GASF may represent an evidence-based measure that quantifies student progress that is not readily identified using standardized achievement tests or even curriculum-based measures. It is a holistic measure insofar as it incorporates broad-stroked, observational assessment. The GASF appears to have the ability for teachers to quantify whether a student is generally doing better or worse. For students identified with multiple disabilities, challenges, or risk factors, teachers might discover that while standardized scores in a particular content area are not meeting the prescribed "rate of improvement" or growth estimate, teachers may find that students are indeed improving in other areas, and those areas indeed may reflect in a GASF score. For example, one general education teacher reported to the researcher frustration over her inability to demonstrate student improvement that is not captured by district and state assessments. This teacher stated that she has worked with students in the past who had made modest academic gains, but despite using multiple interventions, some students do not meet grade level expectations or improve at a rate that is commensurate with established trend lines. The teacher expressed frustration based on the fact that students do indeed evidence growth, but that growth is not adequately reported. The teacher stated that presently her only means of reporting student growth for these students is through anecdotal notes that she shares on quarterly report cards.

It is important to note that the latest version of the Diagnostic and Statistical Manual, DSM-5, eliminated the multi-axial diagnosis; consequently, clinician use of the GAF appears to be extinguished. The DSM-5 Task Force cited among its reasons for eliminating the multi-axial diagnosis from the present DSM a desire to better align with the World Health Organization Disability Assessment Schedule (WHO DAS 2.0) and the International Statistical Classification of Diseases and Related Health Problems, Tenth Edition (ICD-10), a lack of conceptual clarity,

and questionable psychometrics and have instead suggested use of (American Psychiatric Association, 2013). Of particular concern to professionals treating children is the acknowledgement that the WHODAS 2.0 does not presently recognize or identify a classification system to be used with children and adolescents (World Health Organization, 2015). Additionally, WHODAS 2.0 may provide challenges to practitioners insofar as it requires either a self-administration completed by the patient or a rater administration (short form 12 items, long form 36 items). What is gained by this method of administration, is negated by its time intensive nature, which poses a threat to its use as a consistently utilized progress monitoring tool (Gold, 2014). While a single score cannot adequately address the multiple domains of functioning, identifying the incremental/decremental changes of patients (or students) still seems valuable. The elimination of the GAF may result in problems that parallel those presently being experienced by educators. The changes in assessment strategies (for both mental health and education professionals) represent challenges insofar as scoring and interpretation of results stresses an already stressed workload. The elimination of the GAF, and similarly, the inattention to progress of students not easily defined in schools, represents the loss of a metric that provides valuable information as to whether patients (or students) are getting "better or worse".

Furthermore, despite changes to the DSM, a brief and very informal survey of local social workers, psychologists, and psychiatrists suggest that these professionals continue to use the multi-axial diagnostic system to quantify patient progress. Furthermore, these professionals report that present electronic medical records (EMR) require clinicians to provide the multi-axial diagnosis in order to receive reimbursement from third party payers (Moses, 2014).

Implications for Future Research

Given the limited number of participants drawn from a small regional area, additional examination of the inter-rater reliability of the GASF is warranted. Investigating the responses from school personnel representing various geographic, ethnographic, and socioeconomic populations is needed to generalize results to a larger potential population of raters. Furthermore, subsequent studies are needed to extend what is known about those raters making ratings and the agreement between and within various subgroups who are expected to use the GASF in their schools. Comparing ratings of general education teachers, special education teachers, and school psychologists may indicate that one subgroup is more accurate in ratings, providing narrower target bands. Additionally, considering the various response to intervention implementation stages educators find themselves, it may be useful to determine if a particular level of experience and education appears to provide more accurate rating. Assuming that problem solving models are being taught in teacher education preparation, and school psychology programs, researchers may find that teachers and school psychologists trained in response to intervention may be more likely to endorse the use of the GASF as a benchmarking and/or progress monitoring tool.

Perhaps more importantly, the number of subjects rated must increase substantially. For the present study, the researcher used five vignettes. Future studies designed to study additional vignettes would aid in the generalizability of results. The five vignettes most certainly do not represent an exhaustive list of the potential constellation of behaviors, academic or otherwise, school personnel encounter in their buildings. In addition to the inclusion of more representative training vignettes for the purpose of adding to the strength of the GASF's reliability, these vignettes would aid professional development and training assuming that reliability remains robust. Furthermore, while the inclusion of additional vignettes is important to extending the

statistical reliability of the GASF, the rating of students in vivo, both independently by individual professionals and as part of a small group, is essential to validating the utility of this measure in real time. The GASF may be useful in child study team meetings or other student-focused problem solving venues that include input from school professionals representing multidisciplinary roles within the school environment. If findings suggest that team members that may include general education teachers, special education teachers, school psychologists, and school administrators demonstrate reliable agreement in their ratings of students, their GASF ratings may indeed represent a quantifiable metric for decision making to be utilized as part of a greater decision making process.

Additional studies may seek to assess GASF reliability for use with adolescents. Furthermore, studies designed to assess a more equal representation of the student population in terms of gender are much needed. It may be beneficial for future researchers to consider designs that attempt to initiate a form of "gender neutrality" within the vignettes. Eliminating names and modifying pronouns to mitigate gender bias may provide further information that will allow researchers to draw conclusions regarding the technical quality of the GASF in terms of its ability to equitably measure functioning regardless of student sex.

While the initial findings regarding inter-rater reliability are promising, data regarding various forms of validity do not yet exist for the GASF. Future studies that explore whether the GASF possesses adequate construct validity. To date, only one other study has examined the technical properties of the GASF. The study examined the utility of the GASF as a measure of overall school functioning, comparing the GASF to total composite scores from three established behavioral assessments (Condiracci, Holcomb, Lichtenstein, Erdodi, & Maerlender, 2014). This

study suggested that the GASF significantly correlated with WISC-IV FSIQ, mean Achenbach Total Problems Index (TRF total) and the BRIEF (Teacher Report) Global Executive Composite.

A key question that would boost the practical utility is whether the GASF demonstrates concurrent validity with varying levels of school performance. For example, is there a threshold score/range within the GASF indicating risk for school failure? Is there a threshold that may be utilized for placement decisions for those students being considered for more/less restrictive placement? Does the GASF demonstrate sensitivity to change (e.g. to what extent does the GASF capture subtle changes in student behavior)? Similarly, additional studies are needed to study the consistency of raters using GAF scores. Do raters assign similar scores to vignettes as a measure of test-retest reliability?

Studies comparing the GASF to the GAF may be useful for both clinical practitioners and teachers and pending results may provide a "common language" for describing behavior progress between school and clinic. Despite the removal of the GAF from the DSM-V, practitioners may still be using the GAF, and the CGAS for children, for screening and progress monitoring. The CGAS represents a method for synthesizing the overall *mental wellness* and global functioning of children in clinical terms, but its language is indeed clinical when compared to the GASF (see Appendix J). In terms of *academic wellness*, the GASF may represent a method for synthesizing information from multiple aspects of student behavior – namely work quality, work completion, attendance, social interactions, and rule compliance. It is a quantifiable rating that school professionals can use to indicate whether a student is doing better or worse using their unique expertise and experience with kids to assess whether students are trending toward a more positive trajectory in their learning. This may be a critical metric, especially for those students

who are working hard but not experiencing the gains they hoped to achieve. This is indeed the case in at least one behavioral health clinic known to the author.

Questions aside, the findings from the present study provide valuable information as to the psychometric properties that of the GASF that were examined. The GASF proved to be a quick, reliable, easily understood measure of overall student behavior. Both teachers and school psychologists confirmed that the GASF represented a useful and valid assessment in terms of content validity.

Conclusion

This study represents the first investigation of the Global Assessment of School Functioning's reliability and content validity. The results indicate that the GASF demonstrates excellent reliability when rated by school professionals. School professionals also indicate that the GASF appears to capture school based behaviors, and that these professionals can use their present levels of education and training rate student performance levels using this measure. Present initiatives are heavily focused on our schools' ability to measure and quantify results of student academic performance. Education reform and accompanying legislation challenges teachers, principals, superintendents, state boards of education, and others to demonstrate student improvement. Education has adopted a culture of data driven decision-making. As a result, it appears that our schools have adopted a system of "Educational Sabermetrics". Teachers strive to improve their students' NWF, ORF, DCPM, and ODRs.

In an effort to assist students; teachers, school psychologists, and administrators are partnering to interpret data and make informed decisions on curriculum, identify students at risk of educational failure, and select evidence based interventions designed to remediate weaknesses and permit students to enjoy school success. Much of the research has focused on improving

discrete skills related to reading and math, and to a lesser extent, behavior and writing. Focus on these skills has resulted in success for many students and are readily seen in student trend lines and reports from case studies. For others, the results may be less evident.

What appears to be missing is a quantifiable way to measure students' overall functioning. The GASF may represent a method for synthesizing information from multiple aspects of student behavior – namely work quality, work completion, attendance, social interactions, and rule compliance. It is a quantifiable rating that school professionals can use to indicate whether a student is doing better or worse using their unique expertise and experience with kids to assess whether students are trending toward a more positive trajectory in their learning. This may be a critical metric, especially for those students who are working hard but not experiencing the gains they hoped to achieve. Based on the data collected from the present study, it appears that the GASF may indeed be a Global Assessment of School Functioning.

References

Abbo, C., Okello, E., & Nakku, J. (2013). Effect of brief training on reliability and applicability of Global Assessment of functioning scale by Psychiatric clinical officers in Uganda. *African Health Science, 13*(1), 4. doi: doi:10.4314/ahs.v13i1.11

American Psychiatric Association. (2013) *Diagnostic and statistical manual of mental disorders* (5TH ed.). Washington, D.C.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*. (Fourth Edition, Text Revision). Washington, DC:

Bird, H. R., Canino, G. J., Rubio-Stipec, M., & Ribera, J. C. (1987). Further measures of the psychometric properties of the Children's Global Assessment Scale. *Archives Of General Psychiatry, 44*(9), 821-824.

Bonner, M., & Barnett, D. W. (2004). Intervention-based school psychology services: Training for child-level accountability; preparing for program-level accountability. *Journal of School Psychology, 42*(1), 23-43.

Brown-Chidsey, R., & Steege, M. W. (2005). *Response to intervention : principles and strategies for effective practice*. New York: Guilford Press.

Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment,* 6 (4), 284-290.

Clonan, S. M., McDougal, J. L., Clark, K., & Davison, S. (2007). Use of office discipline referrals in school-wide decision making: A practical example. *Psychology in the Schools, 44*(1), 19-27. doi: 10.1002/pits.20202

Condiracci, C., Holcomb, M., Lichtenstein, J., Erdodi, L., & Maerlender, A. C. (2014). *The Global Assessment of School Functioning Scale (GASF):*

A Measure of Cognition, Executive Functioning, and Behavior. Poster Presentation. Dartmouth College.

DiStefano, C. A., & Kamphaus, R. W. (2007). Development and validation of a behavioral screener for preschool-age children. *Journal of Emotional and Behavioral Disorders, 15*(2), 93-102. doi: 10.1177/10634266070150020401

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*(1), 405-432. doi: 10.1111/j.1467-8624.2010.01564.x

Dyrborg, J., Warborg Larsen, F., Nielsen, S., Byman, J., Buhl Nielsen, B., & Gautr√®-Delay, F. (2000). The Children's Global Assessment Scale (CGAS) and Global Assessment of Psychosocial Disability (GAPD) in clinical practice - substance and reliability as judged by intraclass correlations. *European Child & Adolescent Psychiatry, 9*(3), 195.

Education, U. S. D. O. (2010). *Race to the top program guidance and frequently asked questions*. Washington, D.C.: Retrieved from http://www2.ed.gov/programs/racetothetop/faq.html.

Elliott, S., & Gresham, F. (2008). *Social skills improvement system performance screening guide*. San Antonio, TX: Pearson.

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. *Arch Gen Psychiatry, 33*(6), 766-771.

Evans, S., & Sarno-Owens, J. (2010). Behavioral assessment within problem-solving models: Finding relevance and expanding feasibility. *School Psychology Review, 39*(3), 477-430.

Fraenkel, J. R., & Wallen, N. E. (2000). *How to design and evaluate research in education*. Boston: McGraw-Hill.

Furlong, M., & O'Brennan, L. (2007). BASC-2 Behavioral and Emotional Screening System. In J. F. C. In R.A. Spies, & J.F Geisinger (Eds.) (Ed.), *The eighteenth mental measurements yearbook and tests in print* (18 ed.).

Gold, L. (2014). DSM-5 and the assessment of functioning: The world health organizatioin disability assessment schedule 2.0 (WHODAS 2.0). *The Journal of the American Academy of Psychiatry and the Law. 42 (2), 173-181.*

Goodman, S., McGlinchey, M., & Schallmo, K. (2009). Michigan's Integrated Behavior & Learning Support Initiative: Regional Leadership Launching Change. http://miblsi.cenmi.org/About.aspx

Greil, A. Personal communication, (2010).

Gresham, F. M., & Elliott, S. N. (2008). Social Skills Improvement System Rating Scales: Pearson, 19500 Bulverde Road, San Antonio, TX 78259; Telephone: 800-627-7271; FAX: 800-632-9011; E-mail: pearsonassessments@pearson.com; Web: http://www.pearsonassessments.com.

Gresham, F. M., Elliott, S. N., Cook, C. R., Vance, M. J., & Kettler, R. (2010). Cross-Informant Agreement for Ratings for Social Skill and Problem Behavior Ratings: An Investigation of the Social Skills Improvement System--Rating Scales. *Psychological Assessment, 22*(1), 157-166.

Gresham, F. M., Elliott, S. N., Vance, M. J., & Cook, C. R. (2011). Comparability of the Social Skills Rating System to the Social Skills Improvement System: Content and

psychometric comparisons across elementary and secondary age levels. *School Psychology Quarterly, 26*(1), 27-44. doi: 10.1037/a0022662

Gresham, F. M., McIntyre, L. L., Olson-Tinker, H., Dolstra, L., McLaughlin, V., & Van, M. (2004). Relevance of functional behavioral assessment research for school-based interventions and positive behavioral support. *Research in Developmental Disabilities, 25*(1), 19-37.

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol, 8*(1), 23-34.

Hilsenroth, M. J., Ackerman, S. J., Blagys, M. D., Baumann, B. D., Baity, M. R., Smith, S. R., . . . Holdwick Jr, D. J. (2000). Reliability and validity of DSM-IV Axis V. *American Journal of Psychiatry, 157*(11), 1858 - 1863.

Irvin, L. K., Horner, R. H., Ingram, K., Todd, A. W., Sugai, G., Sampson, N. K., & Boland, J. B. (2006). Using Office Discipline Referral Data for Decision Making About Student Behavior in Elementary and Middle Schools: An Empirical Evaluation of Validity. *Journal of Positive Behavior Interventions, 8*(1), 10-23.

Irvin, L. K., Tobin, T. J., Sprague, J. R., Sugai, G., & Vincent, C. G. (2004) (Vol. 6, pp. 131-147): Sage Publications Inc.

Kamphaus, R., & Reynolds, C. (2007). *BASC -2 Behavioral and emotional screening system manual*. Circle Pines, MN: Pearson.

Kamphaus, R. W., DiStefano, C., Dowdy, E., Eklund, K., & Dunn, A. R. (2010). Determining the presence of a problem: Comparing two approaches for detecting youth behavioral risk. *School Psychology Review, 39*(3), 395-407.

Kamphaus, R. W., Thorpe, J. S., Winsor, A. P., Kroncke, A. P., Dowdy, E. T., & Vandeventer, M. C. (2007). Development and Predictive Validity of a Teacher Screener for Child Behavioral and Emotional Problems at School. *Educational & Psychological Measurement, 67*(2), 342-356.

Kamps, D. M., Wills, H. P., Greenwood, C. R., Thorne, S., Lazo, J. F., Crockett, J. L., . . . Swaggart, B. L. (2003). Curriculum influences on growth in early reading fluency for students with academic and behavioral risks: A descriptive study. *Journal of Emotional and Behavioral Disorders, 11*(4), 211-224. doi: 10.1177/10634266030110040301

Keraus, J. W. (1991). *Relative validity and reliability of four global assessment scales for children.* (51), University of Arkansas - Fayetteville, US. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1991-58359-001&site=ehost-live Available from EBSCOhost psyh database.

Landers, R. (2011). Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *neoacademic.* 2014, from http://neoacademic.com/2011/11/16/computing-intraclass-correlations-icc-as-estimates-of-interrater-reliability-in-spss/

Levitt, J. M., Saka, N., Hunter Romanelli, L., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology, 45*(2), 163-191.

Lundh, A., Kowalski, J., Sundberg, C., Gumpert, C., & Landen, M. (2010). Children's global assessment scale (CGAS) in a naturalistic clinical setting: Inter-rater reliability and comparison with expert ratings. *Psychiatry Research, 177*, 206-210.

Maerlender, A. C. (2009). *A global assessment for schools*. Unpublished instrument.

Maerlender, A. C. & Palamara, J. (2011). *The global assessment of school functioning.* Unpublished instrument.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30-46. doi: 10.1037/1082-989X.1.1.30

Merrell, K. (2010). Better methods, better solutions: developments in school-based behavioral assessment. *School Psychology Review, 39*(3), 422-426.

Moses, K. (2014). Personal communication.

Nelson, J. R., Benner, G. J., Reid, R. C., Epstein, M. H., & Currin, D. (2002). The convergent validity of office discipline referrals with the CBCL-TRF. *Journal of Emotional & Behavioral Disorders, 10*(3), 181.

Pearson. BASC - 2 Behavioral and Emotional Screening System. Retrieved May 15, 2011, from https://http://www.pearsonassessments.com/hai/RetailExtensions/newqual.aspx?MenuID=Panel1_3

Renshaw, T., Eklund, K., Dowdy, E., Jimerson, S., Hart, S., James Earhart, J., & Jones, C. (2009). Examining the relationship between scores on the behavioral and emotional screening system and student academic, behavioral, and engagement outcomes: an investigation of concurrent validity in elementary school. *The California School Psychologist, 14*, 81-88.

Rey, J. M., Starling, J., Wever, C., Dossetor, D. R., & Plapp, J. M. (1995). Inter-rater reliability of global assessment of functioning in a clinical setting. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 36*(5), 787-792. doi: 10.1111/1469-7610.ep11966285

Richardson, M. J., Caldarella, P., Young, B. J., Young, E. L., & Young, K. R. (2009). Further validation of the systematic screening for behavior disorders in middle and junior high school. *Psychology in the Schools, 46*(7), 605-615. doi: 10.1002/pits.20401

Robert, E. R., Attkisson, C. C., & Abram, R. (1998). Prevalence of psychopathology among children and adolescents. *The American Journal of Psychiatry, 155*(6), 715.

Salvia, J., & Ysseldyke, J. (2007). *Assessment in special and inclusive education* (Tenth ed.). Boston: Houghton Mifflin Company.

Schorre, B. r. E. H., & Vandvik, I. H. (2004). Global assessment of psychosocial functioning in child and adolescent psychiatry. *European Child & Adolescent Psychiatry, 13*(5), 273-286. doi: 10.1007/s00787-004-0390-2

Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., & Aluwahlia, S. (1983). A children's global assessment scale (CGAS). *Archives Of General Psychiatry, 40*(11), 1228-1231.

Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods In Medical Research, 7*(3), 301-317.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428. doi: 10.1037/0033-2909.86.2.420

Soper, D. S. (2011). The free statistical calculators website. Retrieved April 11, 2011, from http://www.danielsoper.com/statcalc/calc01.aspx

Startup, M., Jackson, M. C., & Bendix, S. (2002). The concurrent validity of the Global Assessment of Functioning (GAF). *British Journal of Clinical Psychology, 41*(4), 417.

Stewart, R. M., Benner, G. J., Martella, R. C., & Marchand-Martella, N. E. (2007). Three-tier models of reading and behavior: A research review. *Journal of Positive Behavior Interventions, 9*(4), 239-253. doi: 10.1177/10983007070090040601

Sugai, G. (2007). Promoting behavioral competence in schools: A commentary on exemplary practices. *Psychology in the Schools, 44*(1), 113-118. doi: 10.1002/pits.20210

Sugai, G., Sprague, J. R., Horner, R. H., & Walker, H. M. (2000). Preventing School Violence: The Use of Office Discipline Referrals to Assess and Monitor School-Wide Discipline Interventions. *Journal of Emotional & Behavioral Disorders, 8*(2), 94.

Taras, H. L. (2004). School-based mental health services. *Pediatrics, 113*(6), 1839-1845.

. *U.S. Public Health Service, Report of the Surgeon General's Conference on Children's Mental Health: A National Action Agenda*. (2000). Washington, DC.

Valen, S., Ryum, J., Pedersen, T., Pripp, G., Are, J., & Karterud, S. (2015). Does a Web-Based Feedback Training Program Result in Improved Reliability in Clinicians' Ratings of the Global Assessment of Functioning (GAF) Scale? *Psychological Assessment*, 9. doi: doi:10.1037/pas0000086

Varley, C. (2013). *Overview of DSM-5 Changes*. Presentation. Seattle Children's Hospital, University of Washington. Retrieved from http://www.omh.ny.gov/omhweb/resources/providers/dsm-5-overview.pdf

Walker, H. M., & Severson, H. H. (1992). Systematic Screening for Behavior Disorders (SSBD). Second Edition: Sopris West, 1140 Boston Ave., Longmont, CO 80501 ($195 for kit).

Walker, H. M., & Severson, H. H. (1994). Replication of the Systematic Screening for Behavior Disorders (SSBD) procedure for the. *Journal of Emotional & Behavioral Disorders, 2*(2), 66.

Walker, H. M., Severson, H. H., Nicholson, F., & Kehle, T. (1994). Replication of the Systematic Screening of Behavior Disorders (SSBD) procedure for the identification of at-risk children. *Journal of Emotional and Behavioral Disorders, 2*(2), 66-77. doi: 10.1177/106342669400200201

World Health Organization. (2015). *WHO disability assessment schedule 2.0.* http://www.who.int/classifications/icf/whodasii/en/index6.html. Retrieved April 22, 2015.

Zlomke, L., & Spies, R. (1998). Test review of the Systematic Screening for Behavior Disorders. http://www.unl.edu/buros

Appendix A

Global Assessment of School Functioning (GASF)

Instructions: Rate student over the past month; identify numeric range that captures his/her functioning, and estimate within the range to assign a single numeric rating; read descriptions above and below to verify placement.

100-91
Meets all academic and social expectations, a model student, superior functioning day in and day out.
90-81
Completes work with no reminders, quality of work is good, does not get upset when making mistakes, takes correction easily, and meets most social expectations; OR meets most academic expectations and all social expectations (is polite, raises hand, considerate of others); participates in wide range of activities.
80-71
Some occasional difficulties in schoolwork or behavioral regulation (may be due to psychosocial stressors); occasionally falls behind in schoolwork; demonstrates ability to make and maintain positive peer relationships typical for age; Participates in some activities.
70-61
Mild academic difficulties (occasional truancy, gets in some trouble, poor grades in one or two classes), but produces adequate academic work; OR behavior generally appropriate with occasional difficulty (may have to leave room or be disciplined once a quarter at most).
60-51
Moderate academic difficulty and at risk for educational failure – could be failing several classes but never identified for special education classes; passing most classes only with support OR few friends; conflicts with peers; behavior may require some form of intervention due to weekly behavioral disturbances. Rare school-activity participation (may play on a sports team). Attendance problems may be affecting ability to learn.
50-41
Academic performance is more than one grade level behind current grade level placement in more than one subject area; if identified with a special education disability, is making modest gains toward goals. OR Social, behavioral, academic difficulties primarily attributed to poor attendance. Demonstrates difficulty making and maintaining positive peer relationships. AND/OR Attendance severely impacting school performance. Is at-risk for retention based on truancy or absences.
40-31
Requires intervention for academics (1:1) AND behavior; behaviorally has good days and bad, with academic skills very fragile, slow progress; OR frequent behavioral outbursts requiring out of classroom time or in-class discipline (several times a week) AND dropping grades. OR Demonstrates weekly absences or more than 12 absences in a semester (7 to 8 in a trimester).
30-21
Severe academic difficulty. Identified with a disability (receiving special education services) but services and interventions having no positive impact; failing in several academic subjects despite interventions AND behavioral problems - at serious risk of being placed out of district due to behavior; multiple behavior problems per week.
20-11
Inability to function in school; educational needs cannot be met due to significant handicaps, severe impairments, or behavior that is out of control; impairment renders child unresponsive to interventions in present setting.
10-1
Danger to self and/or others OR unable to maintain appropriate hygiene OR gross impairment in communication; requires institutional placement, residential setting.

Appendix B

Global Assessment of School Functioning: Content Validity Protocol

Directions: Please read each of the descriptor categories and RANK them by placing a "10" on the line that corresponds with the cluster that represents the HIGHEST level of functioning, a "9" on the line that corresponds with the cluster that represents the next HIGHEST level of functioning, and so on. The number "1" should represent the LOWEST level of student functioning.

Inability to function in school; educational needs cannot be met due to significant handicaps, severe impairments, or behavior that is out of control; impairment renders child unresponsive to interventions in present setting. _____

Completes work with no reminders, quality of work is good, does not get upset when making mistakes, takes correction easily, and meets most social expectations; OR meets most academic expectations and all social expectations (is polite, raises hand, considerate of others); participates in wide range of activities. _____

Severe academic difficulty. Identified with a disability (receiving special education services) but services and interventions having no positive impact; failing in several academic subjects despite interventions AND behavioral problems - at serious risk of being placed out of district due to behavior; multiple behavior problems per week. _____

Danger to self and/or others OR unable to maintain appropriate hygiene OR gross impairment in communication; requires institutional placement, residential setting. _____

Meets all academic and social expectations, a model student, superior functioning day in and day out. _____

Mild academic difficulties (occasional truancy, gets in some trouble, poor grades in one or two classes), but produces adequate academic work; OR behavior generally appropriate with occasional difficulty (may have to leave room or be disciplined once a quarter at most). _____

Academic performance is more than one grade level behind current grade level placement in more than one subject area; if identified with a special education disability, is making modest gains toward goals. OR Social, behavioral, academic difficulties primarily attributed to poor attendance. Demonstrates difficulty making and maintaining positive peer relationships. AND/OR At-risk for retention based on truancy or absences _____

Moderate academic difficulty and at risk (RTI Tier II) for educational failure – could be failing several classes but never identified for special education classes; passing most classes only with support OR few friends; conflicts with peers; behavior may require some form of intervention due to weekly behavioral disturbances. Rare school-activity participation (may play on a sports team). _____

Some occasional difficulties in schoolwork or behavioral regulation (may be due to psychosocial stressors); temporarily falling behind in schoolwork. Participates in some activities. _____

Requires intervention for academics (1:1) AND behavior (RTI Tiers II-III); behaviorally has good days and bad, with academic skills very fragile, slow progress; OR frequent behavioral outbursts requiring out of classroom time or in-class discipline (several times a week) AND dropping grades. _____

Appendix C

Case Vignettes

Tommy is a 10 year-old fifth grade student who moved to the City Elementary school district last year. Tommy was identified as a student in need of special education services for reading and written language disorders in second grade at his old school, Smith Elementary. His academic growth over the years has been minimal while his behavioral disruptions have significantly increased. He is failing all academic classes except math where he has a 60 % average. He does not appear to have any positive peer relationships at school and states that hates school and everyone in it. Reports from his old school indicate that Tommy would often "shut down" when he was asked to do a task that he felt was too hard. Since the beginning of this school year, Tommy has been referred to the office a total of 24 times and suspended eight days in a six-month period. He has been suspended for smoking at school, fighting, and threatening teachers and other staff members. His juvenile probation officer has encouraged the school to file paperwork with the courts identifying Tommy as a person in need of supervision.

Score: _____

Tiffany is a 9 year-old third grade student at Greendale Elementary. Her teacher referred her to the child study team based on behavioral and academic concerns. Tiffany is diagnosed with selective mutism, and she rarely speaks to other students or staff. Over the past three months, Tiffany has had accidents wetting her pants in class and does not tell anyone when this happens. Academically, she is a brilliant reader who reads at a high school level. While she appears to have the ability to write or type responses to questions, she rarely does. Math is challenging for her. While she can add three digit problems with regrouping, she struggles with subtraction, and her multiplication facts are limited to "2s", "5s", and "10s". Tiffany is rarely seen initiating or engaging in play or study with her peers in the classroom or at recess. At times she seems reluctant to pass through the doorway into the classroom appearing to not understand what is

expected of her. At recess and dismissal, she requires teacher assistance to put on her coat and boots, but she is able to remove her things upon arrival to school and return from recess.

Score: _____

Steven is a 4th grader who is identified as a student with a learning disability in reading. Initially, Steven was identified as a child with an Early Childhood Developmental Delay. Steven is described as a pleasant and kind student who has a good sense of humor who wants to achieve. His reading intervention primarily takes place out of the classroom where he uses the Read 180 curriculum. He has good attendance and is presently getting C and C + grades in his core academic content areas. He knows 64% of the high frequency sight words for his grade level, he is able to spell 68% of his grade level spelling words accurately. Comparatively, his comprehension is a strength in reading while making inferences is considered a weakness. Steven tested in the partially proficient range on the MEAP math and reading assessments last year. His teacher adds that Steven has shown great improvement in his writing, but he is still inconsistent about turning in homework.

Score: _____

Nico is a 9-year-old fourth grade student at Apple Elementary School. Nico is an "A" student in the talented and gifted program. He is a hard working student who is well liked by peers and adults. His previous report cards indicate that he has always been a very good student who participates well in class, is extremely well behaved both in and out of the classroom, and who is caring and considerate toward his peers. Each year, he has been nominated and won a special student award both within the class and this year, he has won the Outstanding Student award for

the school. He participates in Lego League, his local scouting organization; plays baseball, hockey, and soccer; and helps with the school's recycling program. In the summer, Nico participates in the local college "Little Einsteins" program that incorporates education for the arts and environmental education programming into a day camp format.

Score: _____

Patty is a second grade student who is struggling academically. Patty is described as polite, helpful, and friendly. She enjoys using the computer and has demonstrated the ability to use it independently. She works better in smaller learning communities than she does in whole class settings. She has a good attitude toward learning and appears to try her best. She can however become frustrated when she cannot readily perform a task. Her basic calculation skills appear adequate for completing work, but math concepts appear difficult for her. Patty began kindergarten downstate where she remained for part of her first grade year. Patty has recently moved to the area; anecdotal reports from her previous teachers indicate that Patty received intervention in reading (Title 1) and math calculation. Language and reading scores as measured by Gates-McGinitie were all in the lower 3rd stanine. Patty's most recent DRA (4) and SORT (1.3) indicate that she is below grade level in reading. Patty fully participates in the second grade general education curriculum and receives 20 minutes of daily reading intervention outside the classroom. Current interventions are geared toward improving Patty's phonemic awareness skills. She receives support in reading through the SRA *Early Interventions in Reading* curriculum which is designed to increase letter-sound recognition and fluency. Patty is progress monitored on a weekly basis using Aimsweb Reading Curriculum Based Measures (R-CBM).

Progress monitoring data indicate that she is showing adequate improvement to eventually meet grade level standards.

Score: _____

Kenny is a 7 year-old first grader at Jones Elementary. Kenny is receiving additional reading support services to improve his decoding and fluency. Teachers report he is responding well to the interventions and expect that he will be released from the intervention by the start of the fourth quarter. Kenny has several positive peer relationships, and he is generally respectful to teachers and other adults in the school.

Score: _____

James is a 9 year-old 4th grade student who was referred to the Educational Support Team for concerns over reading and writing. James is diagnosed with ADHD and takes stimulant medication to help with symptoms. James has had intervention support for reading, but he has not progressed past a 1st grade instructional level. While he can identify letter sounds in isolation, he struggles to blend sounds and read words beyond simple c-v-c words, and he has not yet mastered second grade level sight words. His teacher notes that James frequently attempts to engage others in conversation during instructional time. He blurts out answers and struggles to wait his turn. In addition to his reading and behavioral struggles, James has been absent 19 of the first 100 days of school. His teacher adds that James is a sweet boy who is eager to please. He excels at sports and has many positive peer relationships within the school community.

Score: _____

Jake is an 11 year-old fifth grader at Sunnyside Elementary. He is diagnosed with a Traumatic Brain Injury that occurred in 2003; he has suffered from previous seizures, and it is unclear what his school functioning was or could have been before his injury. Jake no longer takes any of his medications that 1. Help him focus and function within the classroom and 2. Help with any seizures that could occur because his mother thinks he can make the right decision to take or in this case, not take his medications. Jake frequently acts impulsively and can be seen sitting in the principal's office because of something that he has done (stabbing bus seats with a pencil, fighting in the hallway with his friend, staying after for intramurals after counselor and principal had enforced that he go home on the 1st bus). He gets angry about certain issues and lacks appropriate social skills. Within the counseling session, he knows the appropriate responses but does not generalize them with peers. He does sit with a circle of girls and sometimes boys at lunch but carryover into classes or outside of school is slim. Sports are a huge motivator for him even though he doesn't play too often. Jake bends the truth quite often (says he scored two touchdowns in last night's game, but did not play; told the police that an older student threw pills at him and told him to sell them, but he stole them from his mother's drawer and brought to school, etc.) Academically, Jake rides the line for failing classes; he receives Special Education services and needs adult assistance to keep him on task. He has below average skills on verbal and nonverbal reasoning skills and low processing speed and memory skills. Academics are tough but math is his favorite subject (he is in self-contained math and blends for all other classes). He is described as a naughty (not bad) kid who requires adult supervision for a considerable portion of the day to keep him from making bad choices. He has a sarcastic sense of humor and is usually compliant upon making requests or demands of him.

Score: _____

Isaac is an 8 year-old first grade student who enrolled at South Elementary three months ago. He has an IEP and receives services as a student with Emotional Impairment. Isaac was previously in a hospital based residential facility before moving from out-of-state to live with his biological father. Since his move, Isaac has been suspended from school nine times for acts of physical aggression that included biting a teacher, choking a classmate, repeatedly kicking his one-to-one aid, and for attempting to gouge the eyes of a child on the playground. Psycho-educational assessment was halted due to Isaac's unwillingness to cooperate, but social emotional checklists filled out by his teachers and father indicate clinical impairment on internalizing and externalizing scales. Isaac has medical diagnoses from a child psychiatrist that include Post-traumatic Stress Disorder, Major Depressive Disorder, and Conduct Disorder (childhood onset, severe).

Score: _____

Danny is a 6 year-old first grader at Smith Elementary. Danny's teachers describe him as a nice boy who has lots of energy. He says that his favorite part of school is running and racing. He frequently needs reminders to stay on task, speak more quietly, and to stay in control of his body. Twice, near the beginning of the school year, Danny was referred to the principal's office for running and sliding in the halls. Danny's math and reading skills are said to be in the average range, and he scores well on weekly spelling tests, but his writing is often messy and incomplete. He has several positive peer relationships in and out of the classroom and he is respectful and polite to teachers and staff.

Score: _____

Hailey is a 9 year-old fourth grader at Jones Elementary. Hailey's behavior has improved significantly since second grade where she used to spend half of her school day in a 12:1:1 classroom. Since 3rd grade, she has spent all her time in a blended classroom. She receives Special Education based upon her classification of "Autism and Hyperkinesia of Childhood Developmental Delays" (basically ADHD). She is solidly average cognitively and academically but has difficulties with cooperative peer relationships. She receives Speech and Counseling services; both of whose main focus are social interactions and pragmatics. She tends to dominate conversations, has her own agenda for ideas, conversations, and completing group work. She perseverates on topics, objects, colors, etc., and will talk incessantly. She currently has an FBA/BIP to address her talking and (secondary) inability to follow directions. She has difficulty with transitions, particularly when she has to transition from reading (desired activity) to anything to do with math (undesired activity), especially because math is her hardest subject. She is a very bright, sweet, and honest child. Parent involvement is minimal and requests for follow through have been unfruitful.

Score: _____

Allison is a 10 year-old fifth grade student at North Elementary School. Teachers describe her as a model student in the classroom. Last month, Allison was nominated as the student of the quarter based on her classroom performance, behavior, and volunteer efforts within the school. As part of a community project, Allison mentors Kindergarten and 1st grade students and helps them with their schoolwork after school. She has received the perfect attendance award three of the past five years and is part of the new peer mediation program that has been implemented at the school.

Score: _____

Braden is a 12 year-old fifth grader who was referred for special education evaluation based on poor academic performance and trouble focusing in class. Historically, Braden has struggled with reading, writing, and math. He finished his first grade well below grade level in reading. He improved in reading by the end of his second grade year, but he was still a year behind in reading skills (word recognition, decoding, blending). Braden was retained in second grade due to low academic performance and repeated the grade with the same teacher. The second year of 2nd grade helped Braden catch up to his peers and several interventions were put into place. Braden's slow academic progress through third and fourth grade was accompanied by behavioral problems. When frustrated, he would shut down and become argumentative. Braden's inability to focus became more apparent in fourth grade and despite environmental accommodations (preferential seating, focus stations, fidget toys, etc.), he showed poor attention. In this, his fifth grade year, Braden continues to struggle with academics, attention, and his self-esteem appears to be affected as a result. He receives reading intervention using the Read 180 program. He has been diagnosed this year with ADHD, but he does not as yet take medication for symptoms.

Score: _____

Alyssa is a 10 year-old 5th grade student at Bryant Elementary. Alyssa's teachers describe her as a very bright girl who is capable of excellent work and her grades reflect this; however, she needs frequent reminders during a class period to stop talking to peers and to refrain from interrupting the teacher. When Alyssa receives consequences for continued talking (lunch detention), she pouts and stops working altogether. Alyssa will ask to go to the nurse on average 3 times during the week, and she frequently requires bathroom breaks (2 a day in addition to classroom scheduled breaks). Alyssa is well liked by peers and adults alike. She is active in school sports, community theatre, and various school based service clubs.

Score: _____

Annie is a 6 year-old first grade student who receives speech and language services for an articulation disorder. Her teacher indicates that Annie is very distractible and inattentive in the classroom. Her work quality and production is inconsistent. She was evaluated for learning problems. Her scores fell within the average range on standardized cognitive and achievement tests; however, Annie does appear to struggle with phonemic awareness skills and basic numeracy. Both Annie's mother and teacher completed ADHD rating scales. Scores differed considerably between the two respondents indicating that Annie manifests far fewer ADHD symptoms at home. Annie's teacher is concerned that Annie does not have the academic skills to progress to second grade. Annie is described as a happy-go-lucky little girl, but she rarely observed playing with other children. Annie started Kindergarten as a 4 year-old and has a late October birthday, so her teacher is reluctant to identify her as immature.

Score: _____

Appendix D

Email to Participants

Dear Sir or Madam:

My name is Joe Palamara and I am a graduate student in the Counseling and School Psychology Department at Alfred University. I am currently completing my doctoral dissertation in school psychology, and I would very much like your participation in a research study I am conducting. This investigation is intended to extend the literature base and practice regarding universal screening of school-based behaviors. The purpose of this study is to gather information about the reliability of a new screening tool that can be utilized by teachers and school psychologists to assess the overall functioning of students.

The study requires you to visit the TBAISD Moodle site where I have created a class that will help to train you in using this new assessment tool. Additionally, the site contains an informed consent document that you are required to sign electronically. All the tools you will need to complete the study can be found on this page, including a link to the actual study which is presented using surveymonkey.com.

It is estimated that the entire process from the time you log into the site will take about 20 minutes to complete.

At the conclusion of the study, you will be invited to enter your name into a random drawing to win a new Apple Ipad. In order to be eligible, you are asked to complete the survey in its entirety.

The directions for creating a moodle account and enrolling in the course are found in the following attachment available for download (moodle document attachment here).

Again, thank you for your consideration.

Sincerely,

Joseph D. Palamara, M. A.
Doctoral Candidate in School Psychology
Alfred University
Department of Counseling and School Psychology

Appendix E

Informed Consent Document

I agree to participate in this survey willingly and am aware that I can discontinue my participation in this study without penalty at any time. I hereby acknowledge that all of the information provided will remain strictly confidential. The data will only be viewed by the principle investigators and will be maintained on a password protected computer. I understand that no information regarding the school district will be released and all identifying information will be removed from participant surveys. All information will be analyzed by groups. No individual data will be obtained and/or used for individual identification or analysis. Informed consent will be retained for three years, and subsequently destroyed according to APA guidelines. If you have any questions regarding the survey or results, please feel free to contact Joseph Palamara at jdp6@alfred.edu or Dr. Mark Fugate at ffugate@alfred.edu. If you have any questions regarding your rights as a participant in this study please contact Alfred University's Human Subjects Committee at hsrc@alfred.edu. Thank you for your participation in this research.

Sincerely,

Joseph D. Palamara, M.A.
Doctoral Candidate in School Psychology
Alfred University
Email: jdp6@alfred.edu
Phone: 267-918-9542

Dr. Mark Fugate, Ph.D.
Associate Professor
Alfred University
Email:ffugate@alfred.edu
Phone: 607-871-2732

Dr. Danielle D. Gagne, Ph.D.
Human Subjects Research Committee
Alfred University
Email: gagne@alfred.edu
Phone607-871-2213

APPENDIX F

Global Assessment of School Functioning

Instructions: Rate student over the past month; identify numeric range that captures his/her functioning, and estimate within the range to assign a single numeric rating; read descriptions above and below to verify placement.

91-100
Meets all academic and social expectations, a model student. Superior functioning day in and day out. No problems with attendance or truancy.

81-90
Completes work with no reminders, quality of work is good, does not get upset when making mistakes, takes correction easily, and meets most social expectations; OR meets most academic expectations and all social expectations (is polite, raises hand, considerate of others); participates in wide range of activities. No problems with attendance or truancy.

71-80
Some occasional difficulties in schoolwork or behavioral regulation (may be due to psychosocial stressors); occasionally falls behind in schoolwork; demonstrates ability to make and maintain positive peer relationships typical for age; Participates in some activities. If identified as a special education student, is nearing exit based on remediation of skills deficits. Minor attendance problems.

61-70
Mild academic difficulties (occasional truancy, gets in some trouble, poor grades in one or two classes), but produces adequate academic work; if identified as a special education student, is making good progress toward goals; OR behavior generally appropriate with occasional difficulty (may have to leave room or be disciplined once a quarter at most). Absences or tardies may be affecting performance.

51-60
Moderate academic difficulty and at risk for educational failure – could be failing several classes but never identified for special education classes; if identified as a special education student, passing most classes only with support OR few friends; conflicts with peers; behavior may require some form of intervention due to weekly behavioral disturbances. Rare school-activity participation (may play on a sports team). Attendance problems may be affecting ability to learn.

41-50
Academic performance is more than one grade level behind current grade level placement in more than one subject area; if identified with a special education disability, is making modest gains toward goals. OR Social, behavioral, academic difficulties may be attributed to poor attendance. Demonstrates difficulty making and maintaining positive peer relationships. AND/OR Attendance severely impacting school performance. Is at-risk for retention based on truancy or absences.

31-40
Requires significant intervention for academics (1:1) AND behavior; behaviorally has good days and bad, with academic skills very fragile, slow progress; OR frequent behavioral outbursts requiring out of classroom time or in-class discipline (several times a week) AND dropping grades. OR Demonstrates weekly absences or more than 12 absences in a semester (7 to 8 in a trimester).

21-30
Severe academic difficulty. Identified with a disability (receiving special education services) but services and interventions having no positive impact; failing in several academic subjects despite interventions AND behavioral problems - at serious risk of being placed out of district due to behavior; multiple behavior problems per week.

11-20

Inability to function in school; educational needs cannot be met due to significant handicaps, severe impairments, or behavior that is out of control; impairment renders child unresponsive to interventions in present setting.

1-10

Assessed to be unable to benefit from structured academics or academic instruction beyond purely functional skills. Danger to self and/or others OR unable to maintain appropriate hygiene OR gross impairment in communication; requires institutional placement, residential setting.

Appendix G

Survey

Global Assessment Measure for Schools

Demographic Information

1. Please include your age

☐ Under 24 ☐ 25 - 30 ☐ 31 - 35 ☐ 36 - 40 ☐ 41 - 45 ☐ 46 - 50 ☐ 51 - 55 ☐ 56 - 60 ☐ Over 60

2. Gender

○ Male ○ Female

3. Please tell us about where you work.

City/Town:

County

4. Please specify you job title.

☐ Teacher

☐ School Psychologist

5. Please indicate your years of experience in your current position.

○ 1st year

○ 2nd year

○ 3 - 5 years

○ 6 - 10 years

○ 11 - 15 years

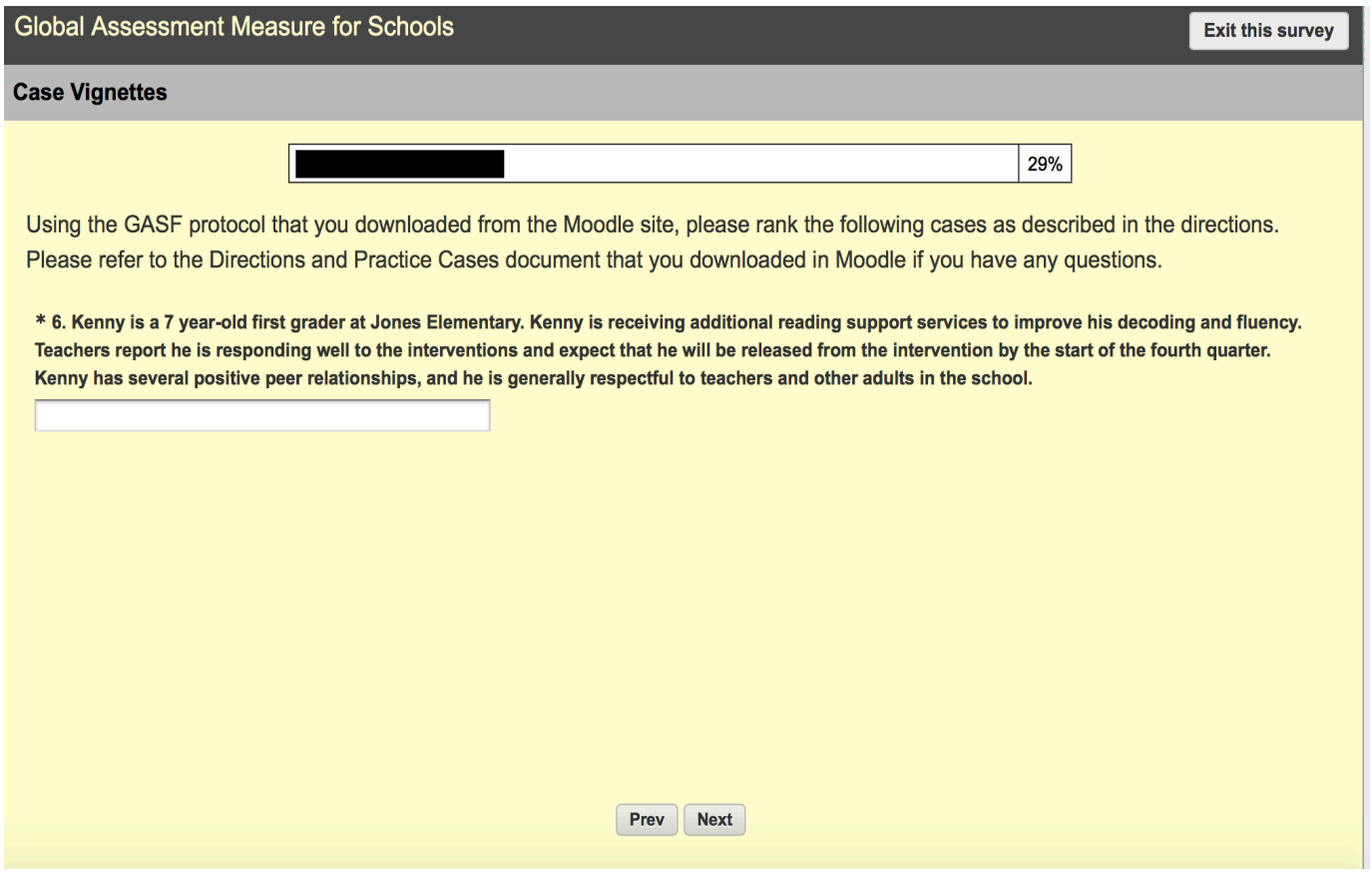
○ 16 - 20 years

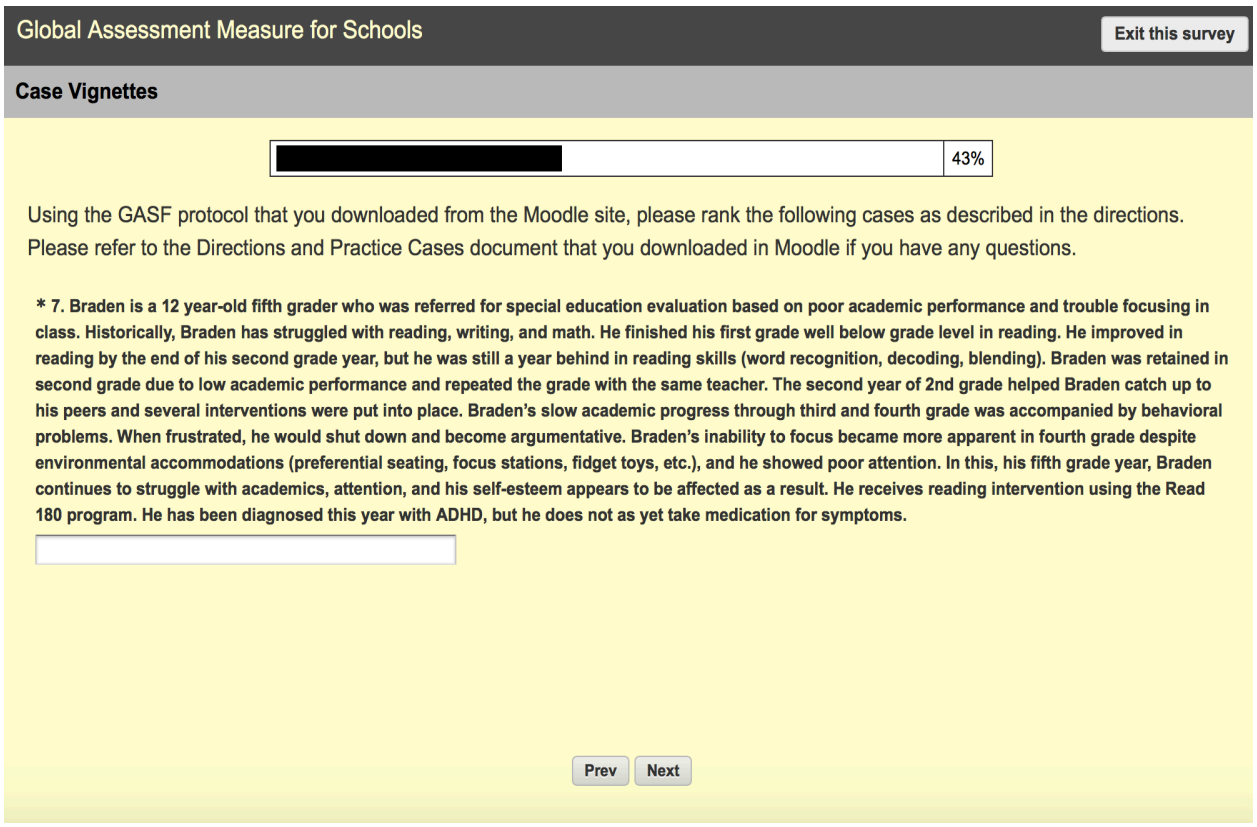
○ over 20 years

Case Vignettes

Using the GASF provided in your participant letter, please rank the following cases as described in the directions. Please refer to the Directions and Practice Cases document provided in your participant letter if you have any questions.

Page 1


Global Assessment Measure for Schools
Exit this survey
Case Vignettes
29%
Using the GASF protocol that you downloaded from the Moodle site, please rank the following cases as described in the directions. Please refer to the Directions and Practice Cases document that you downloaded in Moodle if you have any questions.
* 6. Kenny is a 7 year-old first grader at Jones Elementary. Kenny is receiving additional reading support services to improve his decoding and fluency. Teachers report he is responding well to the interventions and expect that he will be released from the intervention by the start of the fourth quarter. Kenny has several positive peer relationships, and he is generally respectful to teachers and other adults in the school.
Prev
Next


Global Assessment Measure for Schools
Exit this survey
Case Vignettes
43%
Using the GASF protocol that you downloaded from the Moodle site, please rank the following cases as described in the directions. Please refer to the Directions and Practice Cases document that you downloaded in Moodle if you have any questions.
* 7. Braden is a 12 year-old fifth grader who was referred for special education evaluation based on poor academic performance and trouble focusing in class. Historically, Braden has struggled with reading, writing, and math. He finished his first grade well below grade level in reading. He improved in reading by the end of his second grade year, but he was still a year behind in reading skills (word recognition, decoding, blending). Braden was retained in second grade due to low academic performance and repeated the grade with the same teacher. The second year of 2nd grade helped Braden catch up to his peers and several interventions were put into place. Braden's slow academic progress through third and fourth grade was accompanied by behavioral problems. When frustrated, he would shut down and become argumentative. Braden's inability to focus became more apparent in fourth grade despite environmental accommodations (preferential seating, focus stations, fidget toys, etc.), and he showed poor attention. In this, his fifth grade year, Braden continues to struggle with academics, attention, and his self-esteem appears to be affected as a result. He receives reading intervention using the Read 180 program. He has been diagnosed this year with ADHD, but he does not as yet take medication for symptoms.
Prev
Next

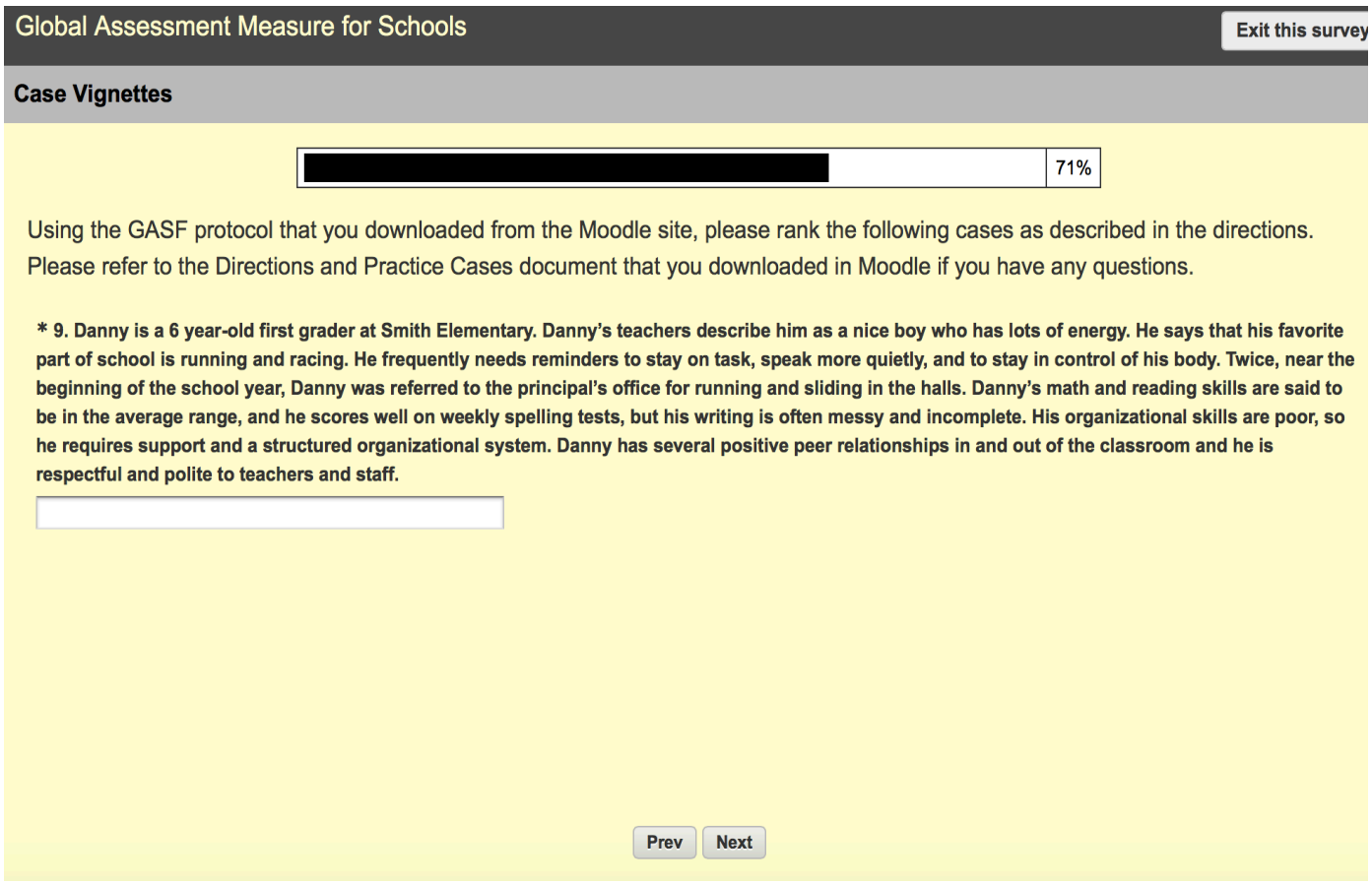
Case Vignettes

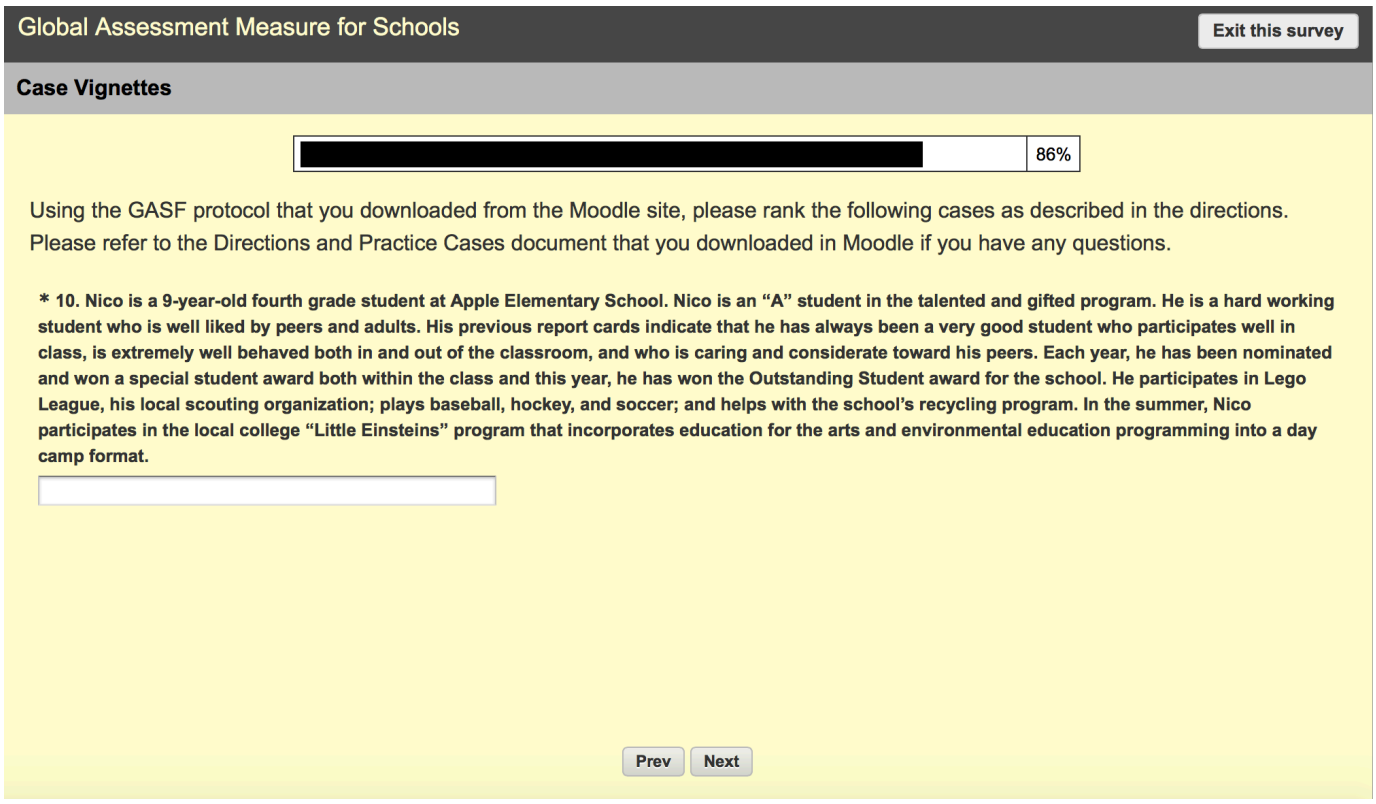
57%

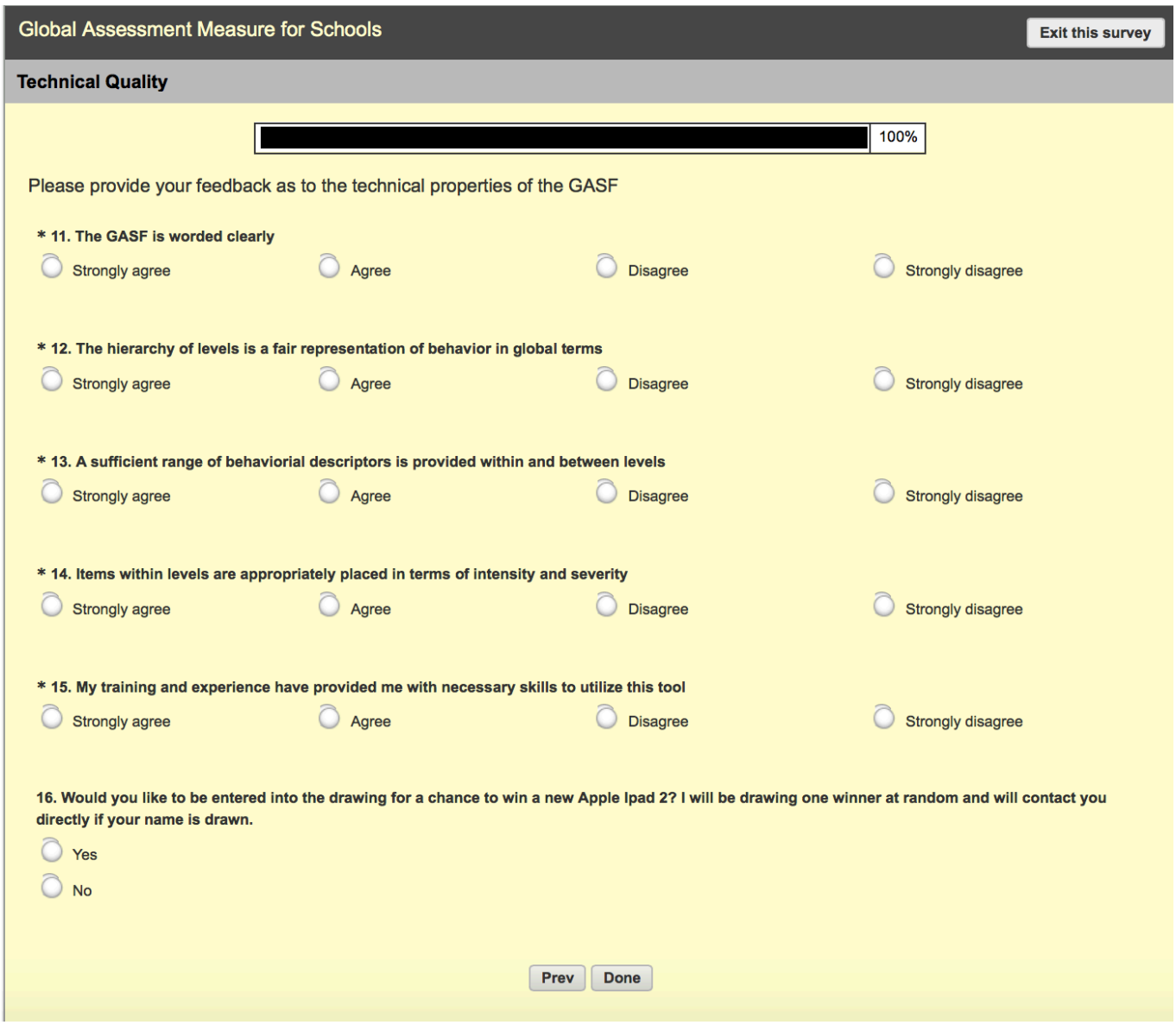
Using the GASF protocol that you downloaded from the Moodle site, please rank the following cases as described in the directions. Please refer to the Directions and Practice Cases document that you downloaded in Moodle if you have any questions.

*** 8. Isaac is an 8 year-old first grade student who enrolled at South Elementary three months ago. He has an IEP and receives services as a student with Emotional Impairment. Isaac was previously in a hospital based residential facility before moving from out-of-state to live with his biological father. Since his move, Isaac has been suspended from school nine times for acts of physical aggression that included biting a teacher, choking a classmate, repeatedly kicking his one-to-one aid, and for attempting to gouge the eyes of a child on the playground. Psycho-educational assessment was halted due to Isaac's unwillingness to cooperate, but social emotional checklists filled out by his teachers and father indicate clinical impairment on internalizing and externalizing scales. Isaac has medical diagnoses from a child psychiatrist that include Post-traumatic Stress Disorder, Major Depressive Disorder, and Conduct Disorder (childhood onset, severe).**

Prev Next


Global Assessment Measure for Schools
Exit this survey
Case Vignettes
71%
Using the GASF protocol that you downloaded from the Moodle site, please rank the following cases as described in the directions. Please refer to the Directions and Practice Cases document that you downloaded in Moodle if you have any questions.
* 9. Danny is a 6 year-old first grader at Smith Elementary. Danny's teachers describe him as a nice boy who has lots of energy. He says that his favorite part of school is running and racing. He frequently needs reminders to stay on task, speak more quietly, and to stay in control of his body. Twice, near the beginning of the school year, Danny was referred to the principal's office for running and sliding in the halls. Danny's math and reading skills are said to be in the average range, and he scores well on weekly spelling tests, but his writing is often messy and incomplete. His organizational skills are poor, so he requires support and a structured organizational system. Danny has several positive peer relationships in and out of the classroom and he is respectful and polite to teachers and staff.
Prev
Next


Global Assessment Measure for Schools
Exit this survey
Case Vignettes
86%
Using the GASF protocol that you downloaded from the Moodle site, please rank the following cases as described in the directions. Please refer to the Directions and Practice Cases document that you downloaded in Moodle if you have any questions.
* 10. Nico is a 9-year-old fourth grade student at Apple Elementary School. Nico is an "A" student in the talented and gifted program. He is a hard working student who is well liked by peers and adults. His previous report cards indicate that he has always been a very good student who participates well in class, is extremely well behaved both in and out of the classroom, and who is caring and considerate toward his peers. Each year, he has been nominated and won a special student award both within the class and this year, he has won the Outstanding Student award for the school. He participates in Lego League, his local scouting organization; plays baseball, hockey, and soccer; and helps with the school's recycling program. In the summer, Nico participates in the local college "Little Einsteins" program that incorporates education for the arts and environmental education programming into a day camp format.
Prev
Next


Global Assessment Measure for Schools
Exit this survey
Technical Quality
100%
Please provide your feedback as to the technical properties of the GASF
* 11. The GASF is worded clearly
Strongly agree
Agree
Disagree
Strongly disagree
* 12. The hierarchy of levels is a fair representation of behavior in global terms
Strongly agree
Agree
Disagree
Strongly disagree
* 13. A sufficient range of behaviorial descriptors is provided within and between levels
Strongly agree
Agree
Disagree
Strongly disagree
* 14. Items within levels are appropriately placed in terms of intensity and severity
Strongly agree
Agree
Disagree
Strongly disagree
* 15. My training and experience have provided me with necessary skills to utilize this tool
Strongly agree
Agree
Disagree
Strongly disagree
16. Would you like to be entered into the drawing for a chance to win a new Apple Ipad 2? I will be drawing one winner at random and will contact you directly if your name is drawn.
Yes
No
Prev
Done

Appendix H

Directions and Practice Cases

The GASF is used to report overall student functioning by teachers, school psychologists, or other professional staff who know the child well enough to make an informed estimate of his overall functioning. The GASF is divided into ten ranges consisting of descriptors that cover academic behavioral severity and functioning. When considering a student's academic and behavioral functioning, DO NOT include impairment in functioning due to physical (or environmental) limitations. The following method is recommended when assigning a GASF rating:

Step 1: Begin with the first level and evaluate each range and ask, " is either the individual's behavioral severity OR level of functioning worse than what is stated within the indicated range description?"

Step 2: Continue moving down the scale until the best descriptive range is found indicating the student's behavioral severity OR the level of functioning that is determined – **which ever is worse.**

Step 3: Consider the range beneath the previously determined range to ensure against prematurely stopping. This range should be deemed too severe both in terms of severity **and** functioning. If this range is indeed too severe, the previously determined range is accurate. If not, continue moving down the scale repeating steps 2 and 3.

Step 4: When determining the specific GASF score within the selected range, consider whether the student's functioning is at the higher or lower end of the range. For example, for a child who is functioning in the 80 -71 range who is experiencing only minimal difficulty in one or two academic areas, the rater will likely provide a score of 77 or 78. If the same child is also occasionally falling behind in school work, and is also struggling occasionally with behavioral regulation, the rater will likely score the child at a 72 or 73.

REMEMBER TO PROVIDE **ONLY ONE** NUMERIC RATING PER CASE.

Appendix I.

Multiple Post Hoc Comparisons

Tukey HSD

Dep. Variable	(I) Occupation	(J) Occupation	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Kenny	Gen. Ed.	Sp. Ed.	-4.206	1.879	.125	-9.17	.76
		Psych	.994	1.616	.927	-3.28	5.26
		Other	-3.806	3.159	.626	-12.15	4.54
	Sp. Ed.	Gen. Ed.	4.206	1.879	.125	-.76	9.17
		Psych	5.200	2.146	.084	-.47	10.87
		Other	.400	3.461	.999	-8.75	9.55
	Psych	Gen. Ed.	-.994	1.616	.927	-5.26	3.28
		Sp. Ed.	-5.200	2.146	.084	-10.87	.47
		Other	-4.800	3.325	.478	-13.59	3.99
	Other	Gen. Ed.	3.806	3.159	.626	-4.54	12.15
		Sp. Ed.	-.400	3.461	.999	-9.55	8.75
		Psych	4.800	3.325	.478	-3.99	13.59
Braden	Gen. Ed.	Sp. Ed.	-8.522	3.540	.087	-17.88	.83
		Psych	-2.156	3.043	.893	-10.20	5.89
		Other	4.111	5.951	.900	-11.61	19.84
	Sp. Ed.	Gen. Ed.	8.522	3.540	.087	-.83	17.88
		Psych	6.367	4.043	.401	-4.32	17.05
		Other	12.633	6.519	.223	-4.59	29.86
	Psych	Gen. Ed.	2.156	3.043	.893	-5.89	10.20
		Sp. Ed.	-6.367	4.043	.401	-17.05	4.32
		Other	6.267	6.263	.750	-10.28	22.82
	Other	Gen. Ed.	-4.111	5.951	.900	-19.84	11.61
		Sp. Ed.	-12.633	6.519	.223	-29.86	4.59
		Psych	-6.267	6.263	.750	-22.82	10.28
Isaac	Gen. Ed.	Sp. Ed.	-3.028	2.551	.637	-9.77	3.71
		Psych	-5.761	2.193	.052	-11.56	.03
		Other	7.639	4.288	.292	-3.69	18.97
	Sp. Ed.	Gen. Ed.	3.028	2.551	.637	-3.71	9.77
		Psych	-2.733	2.913	.784	-10.43	4.96
		Other	10.667	4.697	.116	-1.74	23.08
	Psych	Gen. Ed.	5.761	2.193	.052	-.03	11.56
		Sp. Ed.	2.733	2.913	.784	-4.96	10.43

		Other	13.400*	4.513	.022	1.48	25.32
	Other	Gen. Ed.	-7.639	4.288	.292	-18.97	3.69
		Sp. Ed.	-10.667	4.697	.116	-23.08	1.74
		Psych	-13.400*	4.513	.022	-25.32	-1.48
Danny	Gen. Ed.	Sp. Ed.	2.967	2.524	.645	-3.70	9.64
		Psych	3.667	2.170	.338	-2.07	9.40
		Other	-3.667	4.243	.823	-14.88	7.54
	Sp. Ed.	Gen. Ed.	-2.967	2.524	.645	-9.64	3.70
		Psych	.700	2.882	.995	-6.92	8.32
		Other	-6.633	4.648	.488	-18.91	5.65
	Psych	Gen. Ed.	-3.667	2.170	.338	-9.40	2.07
		Sp. Ed.	-.700	2.882	.995	-8.32	6.92
		Other	-7.333	4.465	.363	-19.13	4.47
	Other	Gen. Ed.	3.667	4.243	.823	-7.54	14.88
		Sp. Ed.	6.633	4.648	.488	-5.65	18.91
		Psych	7.333	4.465	.363	-4.47	19.13
Nico	Gen. Ed.	Sp. Ed.	-1.222	.834	.464	-3.43	.98
		Psych	-.422	.717	.935	-2.32	1.47
		Other	.778	1.402	.945	-2.93	4.48
	Sp. Ed.	Gen. Ed.	1.222	.834	.464	-.98	3.43
		Psych	.800	.953	.835	-1.72	3.32
		Other	2.000	1.536	.565	-2.06	6.06
	Psych	Gen. Ed.	.422	.717	.935	-1.47	2.32
		Sp. Ed.	-.800	.953	.835	-3.32	1.72
		Other	1.200	1.476	.848	-2.70	5.10
	Other	Gen. Ed.	-.778	1.402	.945	-4.48	2.93
		Sp. Ed.	-2.000	1.536	.565	-6.06	2.06
		Psych	-1.200	1.476	.848	-5.10	2.70

Note. *The mean difference is significant at the 0.05 level.

Appendix J

Comparing Elements of the Children's Global Assessment Scale (CGAS) to the Global Assessment of School Functioning (GASF)

CGAS	GASF
70-61 Some difficulty in a single area but generally functioning pretty well (eg., sporadic or isolated antisocial acts, such as occasionally playing hooky or petty theft; consistent minor difficulties with school work; mood changes of brief duration; fears and anxieties which do not lead to gross avoidance behaviour; self-doubts); has some meaningful interpersonal relationships; most people who do not know the child well would not consider him/her deviant but those who do know him/her well might express concern.	**61-70** Mild academic difficulties (occasional truancy, gets in some trouble, poor grades in one or two classes), but produces adequate academic work; if identified as a special education student, is making good progress toward goals; OR behavior generally appropriate with occasional difficulty (may have to leave room or be disciplined once a quarter at most). Absences or tardies may be affecting performance.
60-51 Variable functioning with sporadic difficulties or symptoms in several but not all social areas; disturbance would be apparent to those who encounter the child in a dysfunctional setting or time but not to those who see the child in other settings.	**51-60** Moderate academic difficulty and at risk for educational failure – could be failing several classes but never identified for special education classes; if identified as a special education student, passing most classes only with support OR few friends; conflicts with peers; behavior may require some form of intervention due to weekly behavioral disturbances. Rare school-activity participation (may play on a sports team). Attendance problems may be affecting ability to learn.
40-31 Major impairment of functioning in several areas and unable to function in one of these areas (ie., disturbed at home, at school, with peers, or in society at large, eg., persistent aggression without clear instigation; markedly withdrawn and isolated behaviour due to either mood or thought disturbance, suicidal attempts with clear lethal intent; such children are likely to require special schooling and/or hospitalisation or withdrawal from school (but this is not a sufficient criterion for inclusion in this category).	**31-40** Requires significant intervention for academics (1:1) AND behavior; behaviorally has good days and bad, with academic skills very fragile, slow progress; OR frequent behavioral outbursts requiring out of classroom time or in-class discipline (several times a week) AND dropping grades. OR Demonstrates weekly absences or more than 12 absences in a semester (7 to 8 in a trimester).

Joseph Dennis Palamara

2372 Montmorency Lane• Traverse City, Michigan 49686
Phone: (231) 409-5528 • E-Mail: jpalamara@mac.com

Education

Psy.D. School Psychology Alfred University. Department of Counseling and School Psychology. Alfred, NY

M.A., C.A.S. School Psychology Alfred University. Department of Counseling and School Psychology. Alfred NY

M.S. Special Education. CX Endorsement. Emotionally Impaired Concentration. Eastern Michigan University. Ypsilanti, MI

B.S. English Language and Literature, Teaching Endorsement 8-12. Eastern Michigan University. Ypsilanti, MI

Experience

- Traverse Bay Area Intermediate School District, School Psychologist, Intern. Traverse City, MI — Sept., 2010 – June, 2011
- Sunapee Middle High School, Behavior Specialist. Sunapee, NH — Sept., 2002 – June, - 2007
- Lansing School District, Special Education Teacher. Lansing, MI — Sept., 1997 – June, 2002
- Downriver Community Conference, Education Coordinator. Southgate, MI — Sept., 1993 – Aug., 1997

Publications/Presentations

- Accommodating Students with ADHD. Presented at the Annual Conference of the Michigan Association of Teachers of Emotionally Disturbed Children (MATEDC) — Spring, 1999
- Considering Global Assessment Scales in Schools. Presented to the Traverse Bay Area Intermediate School District School Psychologists. — Spring, 2011

Research Experience

- Investigation of the Psychometric Properties of a Global Assessment Measure of School Functioning. Dissertation. — April, 2015

Affiliations/Memberships

- National Association of School Psychologists — 2007 to present
- American Psychological Association — 2007 to present
- Michigan Association of School Psychologists — 2011 to present

Professional Development

- School-Wide Improvement System (SWIS). Mancelona, MI
- Extensive training in curriculum-based measures including Aimsweb suite, and DIBELS. Alfred, NY.
- Implementing PowerSchool. Sunapee, NH

Interests

- Global assessment in schools, progress monitoring, and behavior.
- Coaching and teambuilding utilizing a structured system of identifying both personal and group strengths and limitations. Utilizing data to implement individualized programs to help individuals and teams to improve communication, identify challenges, set goals, implement plans, assess progress, and modify plans when needed to achieve goals.