AN ALTERNATE ASSESSMENT FOR STUDENTS WITH MULTIPLE DISABILITIES: A PSYCHOMETRIC EVALUATION AND MEASUREMENT VALIDATION STUDY

A DISSERTATION

SUBMITTED TO THE FACULTY

OF

THE GRADUATE SCHOOL OF APPLIED AND PROFESSIONAL PSYCHOLOGY

OF

RUTGERS,

THE STATE UNIVERSITY OF NEW JERSEY

BY

LINDSEY ZAHRA

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF PSYCHOLOGY

NEW BRUNSWICK, NEW JERSEY OCTOBER 2015

APPROVED:

_____
Ryan Kettler, PhD.

_____
Nancy Fagley, PhD.

_____
Melissa Kasmin, PhD.

DEAN:

_____
Stanley Messer, PhD.

Abstract

Federal legislation requiring the inclusion of students with disabilities (SWDs) in standardized testing and annual accountability reports has increased the need for developmentally appropriate and psychometrically sound alternate assessments that are reflective of the state's curricular standards. Few validated measures exist that adequately address the unique needs of this population and integrate evaluation of academic achievement and functional behavior, the latter having been deemed paramount for SWDs. The P.G. Chambers Outcomes Measurement Tool (PGS-OMT) assesses various areas of functioning of students with severe, multiple disabilities to glean a more representative depiction of their capabilities. The main goal of this study was to gather evidence to establish the reliability and validity of conclusions drawn from the scores of this rating scale. The PGS-OMT measures student outcomes and progress and includes assessment of skills in the following areas: Communication, Social, Personal Care, Physical Navigation, Academics, and Functional. The PGS-OMT was used to assess 117 students enrolled in an out-of-district placement in New Jersey. The PGS-OMT was found to be reliable with excellent internal consistency. Internal structure validity evidence was obtained via factor analyses and correlations between PGS-OMT domains. Results indicated a three-factor structure, further supported by the pattern of intercorrelations of PGS-OMT domains. The PGS-OMT shared substantial variance with an established measure of adaptive behavior. Although no significant differences were found between mean scores of advantaged and disadvantaged ethnic groups for two of the three factor domains, further investigation is required to assess instrument bias due to ethnic heterogeneity in the disadvantaged group. Teacher perceptions of the PGS-OMT were favorable and Evaluation Survey results supported the tool's utility.

Acknowledgements

This project would not have been possible without the support and encouragement of many researchers, educators, and family members. I want to extend my sincerest gratitude to the students and faculty who participated in the study, including everyone at the P.G. Chambers School. I also want to thank the transdisciplinary team members who created this measure and facilitated data collection, and in particular, Amber Hummer and Kelly Faber who were always available to answer the numerous questions I had along the way.

I would also like to thank a wonderful dissertation committee, including Drs. Ryan Kettler, Nancy Fagley, and Melissa Kasmin. I especially want to thank Ryan for allowing me this opportunity and for his tireless support during the iterative revision process. I truly appreciate your help and guidance in completing this study.

Above all, none of this would have been possible without my spectacular family. I am incapable of expressing the depth of my gratitude to my grandparents who never bored of hearing my progress and encouraged me to continually strive for the best. I will forever be grateful to my parents, Denise, Kenny, and Joe, who have loved, supported, and encouraged me throughout my academic career and all other life endeavors. I can only hope for the opportunity in the future to help them as much as they have helped me.

L.P.Z.

Table of Contents

List of Tables

List of Figures

Chapter I

Introduction

Alternate assessment for students with disabilities is a commonly used practice across the United States and is perpetually being revised to improve upon existing measures to ensure instruments are yielding reliable scores from which valid inferences can be drawn. In the past, these students were excluded or given exemption from standardized testing practices and so their academic progress was not reflected in summary reports of student progress. Federal law provides regulations for alternate assessment in order to protect and ensure appropriate education for children with disabilities regardless of their needs or severity. The Individuals With Disabilities Education Act (IDEA) of 1997 and Individuals With Disabilities Education Improvement Act (IDEIA) of 2004 state that instruction should not only be tailored for students in a manner that aligns with their specific academic needs, but also that these students appropriately access the tenets and skills addressed by the general curriculum.

Prior to the amendments to IDEA in 1997, most states adopted policies of exemption for students with disabilities (SWDs) participating in statewide assessments (Thurlow, 2004). The exclusion of those with the most severe cognitive disabilities subsequently resulted in a loss of accountability of schools to adequately track their academic progress (Thurlow, 2004). Similarly, this loss of accountability extended beyond informing interested parties about student progress and performance. Relevant stakeholders were also unaware of the degree of instructional adherence to curricular standards to which other students were being held (Erickson, Thurlow, & Thor, 1995).

Amendments to the 1997 IDEA called for alternate assessment of SWDs for whom general state assessment is inappropriate, even with proposed accommodations (20 U.S.C.

1412(a)(16). This legislation highlighted the need to align alternate assessments with the general education curricular content to maintain continuity within the system, and serve as a protective component to ensure that SWDs were instructed in the appropriate content areas. However, much was left for interpretation of how to successfully devise these standards. This resulted in ambiguity and discontinuity across the nation in terms of how states designed measures that adhered to the standards and integrated components of adaptive behavior (Thompson & Thurlow, 2000).

IDEA (1997) additionally mandated that students, including SWDs, participate in statewide testing and addressed concerns regarding SWDs with significant impairments for whom instruction and achievement standards consistent with that provided to general education population would be inappropriate. This revision allowed the creation of alternate standards for no greater than 1% of the students being assessed, and adds that while these standards are adjusted, they must still be in accordance with the standards dictated by the state. Similarly, in 2003, the U.S. Department of Education allowed states to develop and implement alternate assessment measures to better assess those students with the most severe cognitive disabilities who are otherwise unable to sit for state exams. States were also instructed to include alternate assessment results in their aggregated reports of student performance with the caveat that the proportion of inclusion of SWDs could not exceed 1% of the total population of students taking the general and alternate assessment statewide (U.S. Department of Education, 2003). The assessment became known colloquially as the 1% assessment.

What is Alternate Assessment?

Alternate assessment has been defined as "data collection procedures used in place of the

typical assessment when students cannot take standard forms of assessment" (Ysseldyke & Olsen, 1999). Further, according the U.S. Department of Education, "an alternate assessment must be aligned with the state's content standards, must yield results separately in both reading/language arts and mathematics, and must be designed and implemented in a manner that supports the use of the results as an indicator of AYP (adequate yearly progress)" (USDOE 2005, p.15). Alternate assessments are designed for those students with the most severe intellectual or cognitive disabilities for which participation in standardized statewide testing, regardless of accommodations provided, would be otherwise inappropriate (Towles-Reeves, Kleinert, & Muhomba, 2009).

Because of the ambiguity in the law as to how these assessments should be designed and how the content be represented, there are various approaches to alternate assessments of SWDs including portfolio assessment, performance assessment, and comprehensive rating scales of achievement (Browder et al., 2003; Elliott & Roach, 2007). Despite the differing nature of the three approaches and their nomenclature, each tends to incorporate some collection of classwork or other sampling of students' tangible academic productions (Elliott & Roach, 2007). A pervasive trend in the literature highlights differences across states in how these assessments are created, as well as what states are actually using. In terms of prevalence, the majority of states (59%) utilize a portfolio or examination of some concordant body of evidence (Cameto et al., 2009). Forty-one percent of states evaluate students on performance tasks or events that are observed at some point in time during the academic year (Cameto et al., 2009). Lastly, one quarter of states employ rating scales or checklists as their method of alternate assessment; however, it is important to note that these percentages exceed 100% because multiple responses

were accepted as some states use a combination of approaches in their alternate assessment procedure (Cameto et al., 2009).

There are certain assumptions embedded within inclusionary practices through alternate assessment. Past researchers anticipated that alternate assessment would increase visibility and consideration of these SWDs and impart a broader accountability system for the quality of instruction. By including these students in some form of assessment, states would have an indication of how they were performing and if they were making adequate progress in their education. These alternate assessments would also ensure SWDs had access to the curriculum in a manner parallel to their general education counterparts and would subsequently improve the quality of their classroom instruction. This led to increasing achievement expectations, which had previously been low for SWDs. Because little was expected and required of them, SWDs, in turn, had poor performance, which was suggested to be a self-fulfilling prophecy (Browder et al., 2003). The objective of increased visibility through alternate assessment extended beyond simply tracking progress to utilizing that information to better educate SWDs and enhance their overall learning experience within the academic setting (Kleinert & Thurlow, 2001).

Functional vs. Academic Focus in Alternate Assessments

Some have purported that a student with a severe cognitive disability should learn and master functional skills, such as cleaning his or her room or implementing an appropriate hygiene routine, prior to reading, mathematics, language arts, or other core academic content areas (Ysseldyke & Olsen, 1999). Premier versions of alternate assessments typically focused on attainment of adaptive skills as indicators of progress or growth (Browder, et al., 2005). In the early stages of alternate assessment, less than half of the states in the country had actually developed measures that aligned with curricular standards. Some solely focused on functional

and adaptive skills in their assessments and followed curricula based on maximizing an individual's ability to become a contributing member of society (Ysseldyke & Olsen, 1999).

Students with cognitive impairments were shown to spend less time in academically focused tasks in the classroom than those with learning disabilities (Tindal et al., 2003). Some teachers perceived access to the general curriculum to be of secondary importance for SWDs with higher consideration given to functional behavior and the acquisition of appropriate social skills (Agran, Alper, & Wehmeyer, 2002). Considering the varying needs and levels of functioning within this population, adequate measures of student academic achievement also integrated some assessment of adaptive skills (Miller & Linn, 2000).

At the turn of the century, there was a shift in focus toward bolstering SWDs' access to the general curriculum by means of assessing their performance and linking functional skills to the general education standards (Browder et al., 2005). Consultation with various stakeholders within the system allowed professionals to incorporate tenets of adaptive behavior to linear skills outlined in the state's curricular standards (Browder et al., 2003). Gleaning meaningful assessment results meant more than determining if or how well SWDs could perform three-digit multiplication or describe the outcomes of the major world wars. Rather, professionals were urged to apply the underlying foundational skills in a germane manner to the creation of alternate content standards (Thompson, Quenemoen, Thurlow, & Ysseldyke, 2001).

Some states attempted to determine ways in which SWDs could demonstrate mastery of those standards or revised the standard so that it represented some functional skill (Ford, Davern, & Schnorr, 2001). Browder et al., (2005) provided useful examples of how to demonstrate this in practice. For example, to "develop historical perspective," SWDs could demonstrate proficiency through the "use of a personal calendar." In contrast to those who suggested that

these SWDs were learning material different than what was prescribed by standard curriculum, clinicians designing alternate assessments sought to evaluate students on the same material, just in a different, accommodating manner (Thompson et al., 2001).

While adaptive behavior and functional skills are paramount domains for SWDs, there has been a recent decline in states' integration of these areas into alternate assessments, mostly as a result of the No Child Left Behind Act (2001) mandates that students be assessed in the areas of language arts, mathematics, and science (Browder et al., 2005; Cameto et al., 2009; Towles-Reeves, Kleinert, & Muhomba, 2009; Wakeman, Browder, Meier, & McColl, 2007). Concerns have been raised about the lack of adaptive skill assessment and the potential for this area to be subsumed by a sole focus on academics. Some voiced criticism of the educational systems' emphasis on passing statewide testing and suggested that educators are solely "teaching to the test." In other words, because of stringent testing and AYP guidelines, teachers based their lessons on specific material evaluated by standardized tests rather than instruction tailored to the needs of the students. The design or selection of an assessment should center on the relevance for the target population and the associated needs (Hager & Slocum, 2008). However, academic constructs have still been shown to share some degree of variance with functional skills (Kettler et al., 2010; Miller & Linn, 2000).

The P.G. Chambers School incorporated academic and functional domains within the P.G. Chambers School Outcomes Measurement Tool (PGS-OMT) as a way to better represent the skill sets and areas of concentration for the student population in question. Considering the unique needs of the school's population, the transdisciplinary team responsible for the creation of the tool were concerned that assessment results would yield much less meaningful information if the sole focus was on academic functioning without a measure of adaptive behavior. A

combination of the two would potentially reveal larger gains and be sensitive to a greater deal of progress than academics alone. A much more representative picture of student progress could be gleaned from this incorporation. While a connection between adaptive behavior and academic skills exists, educators and academic professionals designing alternate assessments based on alternate achievement standards must make concerted efforts to ensure that measurement of these constructs does not become so intertwined as to measure the same construct (Kettler et al., 2010; Miller & Linn, 2000). It is unknown whether states will continue to include these functional domains, and researchers query how stakeholders will perceive assessments that are solely based in academics (Browder et al., 2005).

Utility of Alternate Assessments and Practical Application of Techniques

Title I of the Elementary and Secondary Education Act mandated that "State assessment shall be aligned with the State's challenging content and student performance standards and provide coherent information about student attainment of such standards" (§1111[b][3][B]). Examination of states' alternate assessments has shown that alternate methods to standardized testing are present but the various methodologies vary greatly in terms of design and the content assessed (Browder et al., 2005). Alternate assessments have not only been unclear in the implementation and scoring of measures, but also lacking in continuity and technical quality in how the alternate standards aligned with the core curriculum (Browder et al., 2005) and state standards (Elliott & Roach, 2007).

The Common Core State Standards address the greater level of need of SWDs and are consistent with goals highlighted in the research to increase their visibility in annual reports (National Governors Association Center for Best Practices [NGA] & Council of Chief State School Officers [CCSSO], 2010; Kleinert & Thurlow, 2001). These standards were created in

part to ensure all students are provided with an appropriate education reflective of what is deemed necessary to excel in later education and careers (NGA & CCSSO, 2010; Browder et al., 2003). But like the research that discusses the creation and use of assessments that are psychometrically sound and congruent with the rigorous content standards, the Common Core State Standards do not explicitly delineate how to do so in practice.

The Common Core State Standards provide practitioners with a linear trajectory of information and requisite skills for multiple content areas, as well as resources for understanding the development of the standards. It does not, however, provide explicit instructions for tailoring these benchmarks for SWDs, resulting in the abovementioned common misalignment of test content with state standards (Browder et al., 2005; Elliott & Roach, 2007). Accommodations such as assistive technology and multimodal instruction are provided as options to facilitate student access to the required content but are explicitly stated as not changing the standard (NGA & CCSSO, 2010). So how do practitioners design psychometrically sound measures for students with severe multiple disabilities, like those enrolled at the P.G. Chambers School, that are both developmentally appropriate and consistent with the rigor of the standards? The need for structured information and strategies to align content to state standards is not unique to the PGS-OMT.

Some assessments, such as the Wisconsin alternate assessment, were found to be misaligned with the state's Language Arts and Science content standards, posing a threat to the validity of the scores yielded by the measure (Roach, Elliott, & Webb, 2005). The researchers of this study sought to examine the depth and range of alternate performance indicators as compared to the areas of academic content outlined in the state's Model Academic Standards. Much research in this area discusses academic content standards versus academic achievement

standards, and has highlighted the difficulty most states have had in creating these alternate standards (Elliott & Roach, 2007; Kearns, Towles-Reeves, Kleinert, Kleinert, & Thomas, 2011). Students are expected to learn and demonstrate mastery of specific skills and material that are detailed in content standards, while achievement standards reflect the level at which students can demonstrate these skills, or indicate how well they have grasped the material. With alternate achievement standards there is "an expectation of performance that differs in complexity from a grade-level achievement standard" (NCLB, 2001). One reason states have had difficulty developing alternate assessments that are in accordance with federal guidelines is that the content of the state academic standards was perceived as irrelevant and the level proficiency required was deemed inapplicable for SWDs (Elliott & Roach, 2007). There is a stance among professionals that academic based instruction for SWDs is inappropriate, and so they tend to forgo traditional teaching for a more functional approach. However, an alternate assessment's strong alignment to content standards is paramount to the validity of the inferences to be drawn from the results of testing (Elliott, Compton, & Roach, 2007; Tindal et al., 2003).

A more systemic approach for creating standards for those students with the most severe disabilities would be quite useful in providing a more meaningful assessment of student outcomes and better guide decision-making at the school and district levels (Browder et al., 2005; Hager & Slocum, 2005). Alternate assessments previously deemed unfit for standardized use were revised and later considered adequately aligned to state content standards utilizing the Webb approach to alignment (Roach, Elliott, & Webb, 2005). An additional strategy for creating or revising performance indicators is to have experts in the field and special education teachers, who are often responsible for completing the evaluation process, determine the appropriateness and relevance of the standards for the target population (Browder et al., 2005). IEP teams can

select target tasks that are explicit and well defined. This would increase the likelihood that the assessment aligns with IEP goals and objectives and further support ongoing progress monitoring (Hager & Slocum, 2005). By integrating IEPs and alternate assessment procedures, input from parents and special education teachers could help guide the selection of appropriate target skills rather than the use of a catalogued curriculum that may sacrifice individualized instruction for time efficiency and ease of use. Differentiating between standardization and individualization is a challenging endeavor, but in doing so, professionals can maximize student access to the curriculum and provide accurate and clear instructional recommendations to teachers (Hager & Slocum, 2005).

Reliability, Validity, and Alternate Assessments

NCLB further required that assessments "be used for purposes for which such assessments are valid and reliable, and be consistent with relevant, nationally recognized professional and technical standards" (20 U.S.C. 6311 § 1111 (b)(3)(C)(iii). Prior to establishing the validity evidence in support of inferences from of an assessment, an instrument must be shown to yield reliable scores, measuring the same thing consistently. Used to indicate how well items within an instrument fit together to measure a single construct, Chronbach's alpha is often employed by professionals to demonstrate adequate internal consistency (Sattler, 2008).

The *Standards for Educational and Psychological Measurement* define validity as the "degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (AERA, APA, NCME, 1999). The most recent version of the *Standards* conceptualizes validity as classifications of various forms of evidence including evidence based on test content, evidence based on response processes, evidence based on internal structure, evidence based on relations to other variables, and evidence based on the consequences of

testing. More specifically, internal structure validity is established by demonstrating that the various parts or domains of an instrument fit together, and is typically evaluated through factor analytic techniques. Further, concurrent validity evidence based on relations to other variables is often obtained by evaluating the degree to which a measure converges with related measures of similar constructs and diverges with differing constructs. It is additionally recommended that multiple sources of evidence be used in establishing adequate validity for an instrument.

Johnson and Arnold (2004) published a study examining the alternate assessment practices in Washington State. The authors evaluated the Washington Alternate Assessment System (WAAS), a portfolio assessment that relied on teacher ratings of work samples, and subsequently determined there were "serious shortcomings" to the procedures. Various sources of content and structural invalidity were found within the WAAS, including unclear scoring criteria and misalignment to state standards in a portion of the portfolios sampled.

In the course of any new policy or procedure, there are challenges in aligning the new with the preexisting conditions. More specifically, states have experienced some difficulty aligning the alternate standards with the grade level standards, for both achievement and content (Rabinowitz, Sato, Case, Benitez, & Jordan, 2008). This is especially true when the process adopts a subjective nature in allowing teachers to choose, modify, or design the alternate version of standards. Relatedly, teacher subjectivity and ability to adhere to assembly guidelines when completing portfolio assessments was another noted threat to validity in Johnson and Arnold (2004).

The ability of teachers to compile a portfolio in accordance with the state guidelines was shown to influence student scores on the assessment and so the generalizability of results was questioned (Johnson & Arnold, 2004). This subjectivity and reliance on teacher judgment has

been frequently examined throughout the research. Teachers are often asked for insight regarding student functioning, as they are generally the individuals who are present in the classroom and have the most contact with these students. While teachers tend to be good sources from which to glean information surrounding academic competence, research has shown there is variability in teacher judgments.

Hoge and Caladarci (1989) conducted an empirical review examining how well teacher ratings aligned with the outcomes of a standardized tool measuring academic success. The authors aggregated research from 16 studies that found teacher judgments often accurately reflected student achievement as opposed to another academically related construct. These findings lend credence to the comprehensive rating scale approach of alternate assessment that so heavily relies on teacher observations and opinions. However, it is important to note a high degree of variability between teachers' ratings within the same classroom, which led the authors to conclude that individual differences in teachers' abilities was a moderating variable between their predictions and actual student achievement (Hoge & Caladarci, 1989). Similar to the WAAS and the concerns highlighted in Johnson and Arnold (2004), teacher responses on the PGS-OMT solely indicate how the student scores and how their performance is represented on the measure. They are responsible for correctly completing the protocol and any misunderstanding or inability to do so reliably will greatly impact the assessment results and their interpretation. So, the decisions and implications of the information that will later be applied to IEP activities could be misrepresentative of the students' needs and actual level of functioning.

Evaluating Technical Quality in Practice

Sufficient assessment reaches beyond the narrow scope of the explicit skill definition being measured. It has larger implications for how that knowledge generalizes to the overarching abilities of the students. Two major goals of large-scale alternate assessment are to improve outcomes for students and better inform stakeholders, which means merely meeting federal reporting requirements is inadequate. These assessments should be able to measure performance with a wide enough breadth of coverage to best illustrate the student's developmental trend as they are exposed to more instruction and develop mastery of both lower level and higher ordered skills (Tindal et al., 2003). This is a challenging task in practice. For example, to demonstrate AYP in Utah's alternate assessment, two performance goals were selected to represent the students' ability to independently demonstrate skills in language arts and mathematics (Hager & Slocum, 2008). Faculty members and teachers opined that this was not a comprehensive assessment of the skills, which is consistent with others' concerns pertaining to the limited number of items on performance tasks (Miller & Linn, 2000). That is, construct underrepresentation had occurred where an appropriate judgment of student achievement could not be gleaned from that relatively small sampling of abilities (Hager & Slocum, 2008).

Past research in this area typically utilized the various sources of validity delineated in the *Standards*; however, while each validity argument should be able to stand alone, aggregating multiple sources of evidence provides a more comprehensive evaluation of the validity of scores yielded from the assessment (Hager & Slocum, 2008). The multitrait-multimethod (MTMM) matrix is commonly used by researchers to garner information from multiple sources and reflective of multiple traits to provide validity evidence for a new measure (Campbell & Fiske,

1959; Elliott, Compton, & Roach, 2007; Kettler et al., 2010). This has proved useful in practice for evaluating a variety of constructs, such as academic skills and performance, as well as adaptive behavior (Kettler et al., 2010).

Employing a multimethod approach, Elliott, Compton, & Roach (2007) provided a framework for further validation efforts by comparing the Idaho Alternate Assessment (IAA) with related, standardized measures (i.e., Academic Competence Evaluation Scales [ACES] and Vineland Adaptive Behavior Scale [VABS]) (Idaho Department of Education, 1999; DiPerna, & Elliott, 2000; Sparrow, Balla, & Cicchetti, 1985). The authors sought convergent and discriminant validity evidence, the combination of which provided professionals with a strong validity argument for conclusions made from the alternate assessment scores. The authors found that the content assessed by the IAA adequately reflected the state's academic content standards, but other validity evidence was mixed yet promising. Results also demonstrated significantly higher correlations between IAA subscale scores in Reading, Language Arts, and Mathematics with measures of adaptive behavior on the VABS ($r = 0.60$) and academic enablers on the ACES ($r = 0.52$) than with measures of academic skills ($r = 0.26$). Interestingly, the component skills of the academic enablers scale included behaviors pertaining to interpersonal skills, motivation, engagement, and study habits, which are more functional in nature than the ACES academic skills scale based solely on reading/language arts, mathematics, and critical thinking. This suggests that the IAA has more shared variance with adaptive behavior, though the assessment does not contain a scale to measure that construct.

Kettler et al. (2010) extended this work and examined the relationship between alternate assessments based on alternate achievement standards (AA-AAS), a general achievement test, and two norm-referenced teacher rating scales in six states. The authors found a high degree of

shared variance between reading and mathematics scores, indicating they could potentially represent the same construct. Further, analyses revealed that scores from the states' AA-AAS were strongly related to adaptive behavior as measured by the Vineland Adaptive Behavior Scale, with over half of the correlations being in the very large range or higher (15 of 25 pairings). Interestingly, in two states (i.e., Idaho and Indiana) these correlations were higher for those SWDs who were eligible to participate in the state's alternate assessment as compared to those SWDs who were deemed ineligible and participated in the regular state testing. Similar trends were found throughout the various analyses, including between the Vineland Adaptive Behavior Composite and the ACES Academic Skills scores, with the correlation being stronger for the students with more significant impairments ($r = 0.50$) than the other group ($r = 0.44$). A concluding point highlighted that constructs measured by AA-AAS were related to academic constructs, including readiness and academic skills, and while there was a great deal of variation observed, the correlations tended to be within the medium or large range. However, these areas were more related to functional behavior, with correlations tending to be within the very large range.

The results of alternate assessments must be reliable and yield valid inferences, so states must make concerted efforts to demonstrate sound psychometric properties in their measures (Elliott, Compton, & Roach, 2007). Researchers advise professionals to incorporate a viable plan for continual validity monitoring in their procedures to yield consistently meaningful results and promote ongoing confidence in those who use and review the measures (Elliott & Roach, 2007; Goldstein & Behuniak, 2011). In addition to being in accordance with the *Standards* and contiguous with state standards, alternate assessments must also be reflective of the unique needs and learning styles of SWDs. Appropriately describing this population and successfully

measuring their progression of learning is a vital task in the development of alternate assessment (Marion & Pellegrino, 2006). Compiling validity evidence from multiple sources through the use of multiple methods can aid professionals in synthesizing and analyzing information from various domains in a meaningful, coherent manner.

Summary

The current study was inspired by a growing need for a psychometrically sound alternate assessment that evaluated students with the most severe disabilities in the requisite academic areas, as well as adaptive behavior. Research has shown that standardized testing or typical assessments do not adequately assess students with profound or severe disabilities. These measures, which serve to evaluate students in areas that align with common core standards, do not yield information and results that are meaningful or interpretable when considering a student's current progress or level of functioning. Such discrepancy is particularly significant for out-of-district, academic placements that serve SWDs. Educators and professionals employed at P.G. Chambers School, an organization for students with severe, multiple disabilities, recognized the need for an evaluation tool that could adequately capture the impact of academic programming on the progress and outcomes of the students in various areas (i.e., Social, Communication, Functional, Physical Navigation, Personal Care, and Academic).

Created by a transdisciplinary team that included teachers and teaching assistants, an occupational therapist, a physical therapist, a speech-language therapist, and a nurse, the P.G. Chambers School Outcomes Measurement Tool (PGS-OMT) was designed to better evaluate students with multiple disabilities enrolled in an academically oriented setting. This study sought to provide information regarding the reliability of scores and validity of inferences drawn from the PGS-OMT, as well as insight into the perceived utility of the measure. These findings

have the potential to influence the type of assessment administered to the students at the P.G. Chambers School. Educators at the facility could further use the information to better guide decisions surrounding goals and objectives suited for the needs of SWDs.

Research Questions and Predictions

(1) Does the PGS-OMT yield reliable scores that demonstrate adequate internal consistency? Coefficient alpha was predicted to be within the acceptable to excellent range as other research has demonstrated these results in the evaluation of alternate assessment instruments (Wisconsin Department of Public Instruction, 2010).

(2) Does the PGS-OMT demonstrate adequate internal structure validity evidence of match to its target constructs? Confirmatory factor analysis was predicted to yield overall good fit indices for the proposed six-factor model and confirm that items within the domains appropriately loaded onto their respective factors.

(3) How well do scores from the PGS-OMT converge with scores from other measures that are established as psychometrically sound? The PGS-OMT was predicted to demonstrate larger correlations with related Vineland-II domains and small correlations between unrelated domains (e.g., Academic vs. Motor Skills). Similar results have been found when evaluating other alternate assessment instruments (Elliott, Compton, & Roach, 2007; Kettler et al., 2010).

(4) Is there a significant difference between mean scores of advantaged and disadvantaged ethnic groups for any domain measured by the PGS-OMT? Results were hypothesized to indicate no statistically significant difference between groups across the three factor domains and overall composite score. Previous researchers investigated the presence of demographic bias in the Idaho Alternate Assessment by correlating demographic

characteristics with measured constructs, and found small correlations (Elliott, Compton, & Roach, 2007).

(5) What are teacher impressions regarding the process of completing the PGS-OMT and its perceived utility? It was expected that teachers find the PGS-OMT time efficient and easy to use.

Chapter II

Method

Participants

Students enrolled in the P.G. Chambers School Program were the participants in the study. As a New Jersey Approved Private School for Children with Disabilities, the P.G. Chambers School serves individuals between the ages of three and fourteen, preschool through grade eight, and offers various services including speech and occupational therapy, as well as comprehensive educational instruction as per the New Jersey Common Core State Standards. Most students enrolled in this placement have severe, multiple disabilities including cognitive and physical impairments. Passive informed consent letters were distributed to the sending school districts for each student regarding participation in the assessment process. Students attending P.G. Chambers School were placed by their Local Education Authority in accordance with IDEA (2004), as it was determined to be the least restrictive environment by which their educational needs could be appropriately met. The P.G. Chambers School educates students under the auspices of the school district's permission, and therefore informed consent was directed to the sending districts. Students for whom participation was not declined participated passively in the study by being assessed by their teachers and other members of the transdisciplinary team.

Of the program's total population of 119 students, 117 individuals were included in the sample. One student was excluded from participation due to lack of consent, and the other because the team did not have time to complete the data collection procedure on the day of assessment. Table 1 depicts the demographic characteristics of the student sample.

Table 1.

Student Demographic Characteristics (N = 117)

	N	Percentage
Gender		
Male	66	56
Female	51	44
Grade*		
Early Childhood	50	43
Elementary	37	32
Middle School	29	25
Ethnicity		
White/Non-Hispanic	89	76
Black/Non-Hispanic	6	5
Hispanic	11	9
American Indian/Alaskan Native/Pacific Islander	0	0
Asian	11	9
Vision*		
Within Normal Limits	65	56
Impaired	47	40
Hearing*		
Within Normal Limits	94	80
Impaired	16	14
Communication*		
Verbal	50	43
Non-Verbal	65	56
Mobility*		
Ambulatory	61	52
Non-Ambulatory	55	47

Note. *Frequencies do not total 117 because of missing information or rounded values.

Measures

P.G. Chambers School Outcomes Measurement Tool (PGS-OMT). The PGS-OMT is a measure of student abilities and outcomes. Professionals at the P.G. Chambers School asserted that a psychometrically sound tool was still needed to appropriately assess students with severe, multiple disabilities who attend academically based programs. The team sought a unique measure that not only assessed functional progress, but also academic gains in their students. A transdisciplinary team designed an instrument intended to effectively measure and communicate the impact of the school program on student outcomes to stakeholders, with regard to several areas of functioning.

Multiple versions of the PGS-OMT were used in a series of pilot phases to further develop the tool. Following these pilot phases, a content validity review was conducted to modify the instrument, and additional measures were surveyed and considered. The team determined the set of items that was parsimonious and comprehensive in order to identify indicators and composites deemed reflective of the needs of the student population. Items were subsequently reworded for content and skills were operationally defined to streamline the scoring process.

The present version of the PGS-OMT includes items referring to specific skills in the following areas: Communication, Social, Personal Care, Physical Navigation, Academics, and Functional domains. Each skill is evaluated on a 7-point Likert scale. The Social, Communication, Physical Navigation, and Personal Care domains are rated as follows according to the ability or frequency with which the student demonstrates a particular skill: 0 = Never, 1 = Maximum Support, 2 = Moderate Support, 3 = Minimal Support, 4 = Emerging (25%), 5 = Sometimes (50%), 6 = Often (75%), 7 = Always (100%). The Academic and Functional

domains are rated as follows: 0 = Never, 1 = Introduced (Participation), 2 = Attempted (Exploring), 3 = Minimal (Infrequent and Inconsistent Performance), 4 = Emerging (Beginning Performance), 5 = Sometimes (Regular Performance/Not Always Correct), 6 = Often (More Often Correct but Not Mastered), 7 = Mastered. The Likert scales are accompanied by skill level definitions that guide the use of the rating scale.

A brief examination of academic content was also conducted to compare the progression of material within the Academic domain to the New Jersey Common Core State Standards. The seven mathematic items within this domain were best aligned with the first grade state standards targeting basic addition and subtraction, understanding whole number relationships, and basic geometric concepts (e.g., congruence and symmetry) (NGA & CCSSO, 2010). Consistent misalignment was found for ELA/literacy standards, which stipulate a progression of skills with increasing complexity. The seven items targeting phonics, reading fluency, and comprehension are best represented by the first grade reading standards for foundational skills (NGA & CCSSO, 2010). Items relating to literature, informational texts, and writing are grossly lacking.

Vineland Adaptive Behavior Scales, Teacher Report Form-Second Edition (Vineland-II). The Vineland-II was utilized to collect information about the students' adaptive behavior, which is all of the combined behavior and skills needed to take care of oneself in regular daily activities. These abilities continue to develop as the child matures but there are certain capabilities that are expected at each age. This measure concentrates on what the child actually does on an everyday basis rather than what she or he is able to do. The Vineland-II includes four domains (i.e., Communication, Daily Living Skills, Socialization, and Motor Skills) that comprise an overall measure of skills, the Adaptive Behavior Composite. Items are rated on a 3-point Likert Scale (0 = Never, 1 = Sometimes, 2 = Always). During the data

collection process, evaluators were instructed to begin each domain with the first item, as administration procedures suggest start points may be lowered if there is a suspected developmental delay (Sparrow, Cicchetti, & Balla, 2006).

Extensive statistical analysis has been conducted on the Vineland-II to establish it as a psychometrically sound measure whose scores are reliable and lead to valid inferences in practice. To demonstrate reliability, internal consistency of the Vineland-II subdomains, domains, and overall Adaptive Behavior Composite has been evaluated using coefficient alpha. For the four domains, reliability coefficients were found to be high, ranging from .86 to .98 across the various ages. The reliability coefficient for the Adaptive Behavior Composite across ages three through 18-years was on average .98, with scores ranging from .97 to .99 (Sparrow, Cicchetti, & Balla, 2006).

Validity evidence for the scores of the Vineland-II has been presented for the following areas: test content, response processes, internal structure, and relations to other variables. The authors provided evidence based on test content strongly rooted in theory with empirical linkage to the behaviors deemed critical to adaptive behavior. Prior research has assessed the degree to which the internal structure of the Vineland-II adheres to the theoretical model of adaptive behavior through hierarchical confirmatory factor analysis and an examination of intercorrelations among subdomains. Results supported the model proposed by the authors, and all domain factors have high loadings (.86 or higher) on the overall factor of adaptive behavior. Similarly, a majority of the loadings on the subdomain level were high with most falling between .70 and .80. However, loadings for Written, Personal, and Gross Motor subdomains were lower than .70 and were attributed to floor and ceiling effects for the former two subdomains. Additionally, the Gross Motor subdomain loading of .53 was suggested to be a result of weaker

relations to the Motor Skills factor, and the broader Motor Domain in general tended to not load as strongly onto the other domains because of the nature of the skills embedded within it (Sparrow, Cicchetti, & Balla, 2006).

During the standardization procedures, correlational analyses were conducted between scores from the Vineland-II and several other measures to examine the relationship between this measure and other instruments commonly used to assess adaptive behavior and cognitive functioning. More specifically, the Vineland-II was compared to the Behavior Assessment System for Children, Second Edition (BASC-2). Although the BASC-2 mostly measures problem behaviors and internalizing concerns, it also includes several measures of adaptive functioning. Large correlations were demonstrated between the Adaptive Behavior Composite of the Vineland-II and the Adaptive Skills Composite of the BASC-2. Correlations between the adaptive components of the two measures were reported for three age groups and were as follows: $r = .65$ for ages 3-5, $r = .73$ for ages 6-11, and $r = .73$ for ages 12-18. The authors concluded that the correlational evidence demonstrated by their analyses provided further support for the construct validity of Vineland-II scores.

Student Demographic Questionnaire. The Student Demographic Questionnaire (Appendix A) was completed by teachers to gather information regarding individual student characteristics. The questionnaire included items pertaining to gender and ethnicity, as well as diagnosis and mobility.

Evaluation Survey. The Evaluation Survey (Appendix B) was designed to obtain teacher feedback on their experience completing the PGS-OMT and the Vineland-II, as well as the perceived utility of the PGS-OMT. Adapted from Huai (2004), the Evaluation Survey was created to assess teacher perceptions of completing a screening system used to evaluate academic

functioning. This study's Evaluation Survey included seven questions, four of which were answered on a four-point Likert scale (1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Agree*, and 4 = *Strongly Agree*), including "I consider the time spent completing the P.G. Chambers School Outcomes Measurement Tool for each of my students reasonable" and "Overall, the P.G. Chambers School Outcomes Measurement Tool is easy to use." The Evaluation Survey also included a list of nine features from which teachers could select what they "liked best" and "liked least" about the tool. Finally, an open response box allowed teachers to provide general feedback and comments that they wanted the transdisciplinary team to consider for future uses.

Procedure

Each transdisciplinary team, consisting of one teacher and any two other team members, including but not limited to occupational therapists, physical therapists, speech specialists, teachers, teaching assistants, nursing staff, and a learning consultant, completed data collection in the following order: (a) student demographic questionnaire, (b) PGS-OMT, and (c) Vineland Adaptive Behavior Scale, Second Edition- Teacher Rating Form (Vineland-II) for each student in the classroom on the same day. All information was completed in its entirety before proceeding to the next item or domain. Items were not to be skipped. Teachers were instructed to refer to the corresponding skill level definitions when recording data for the PGS-OMT and to select the number in the appropriate box that *best* describes how often the student demonstrates that respective skill. Finally, teachers were administered the Evaluation Survey, a brief, anonymous questionnaire to evaluate their perceptions of the PGS-OMT.

Evaluation materials and protocols belong to the P.G. Chambers School, and permission to release the data to the principal investigator was granted by the transdisciplinary team in the interest of establishing the psychometric properties of the measure. All research information was

deidentified and the data received was considered extant. Student participants were given identification numbers by the team that were used on all evaluation material. The principal investigator did not have access to any information to identify original participants.

Data Analysis

This study was designed to evaluate the reliability of scores from the PGS-OMT, as well as the validity of inferences drawn from those scores, as an alternate assessment for students with severe, multiple disabilities enrolled in the P.G. Chambers School. Quantitative techniques were used to analyze the data. Pre-analysis data screening was conducted to assess for missing data and the presence of outliers. Correlational analyses were used to determine relationships among variables on the PGS-OMT and compare the relationship between the PGS-OMT and Vineland-II domains. The internal structure of the PGS-OMT was assessed via factor analyses and a multitrait-multimethod matrix. To assess for instrument bias, the mean scores of factor domains between advantaged and disadvantaged ethnic groups were compared using independent samples *t*-tests. Table 2 presents a summary of the proposed analyses.

Table 2.

Data Analytic Plan for Evaluating the Psychometric Properties of the PGS-OMT

	Data Analytic Techniques
Reliability	Cronbach's Alpha
Internal Structure	Confirmatory Factor Analysis/ Exploratory Factor Analysis Intercorrelations Multitrait-Multimethod Matrix
Relations to Other Variables	Correlations between PGS-OMT and Vineland-II domains
Demographic Bias	Independent Samples *t-tests*
Utility	Descriptive Statistics

Reliability. Cronbach's alpha was used to examine the internal consistency of the PGS-OMT. Coefficient alpha is often used to indicate how well a set of items within a scale fit together to measure a single construct (Sattler, 2008). A range between .70 and .90 has been recommended for acceptable to excellent results (Meyers, Gamst, & Guarino, 2006). For the present study, reliability was demonstrated by examining the PGS-OMT's internal consistency on the domain, factor domain, and total scale levels.

Internal Structure Validity. The internal structure of the PGS-OMT was evaluated in two ways. Correlations between the domains within the instrument were examined (See Table 3 for initial predictions), and medium to large, positive correlations demonstrate that the domains are related while remaining as distinct areas that could function as independent constructs (Cohen, 1992). The internal structure of the PGS-OMT was further evaluated via confirmatory factor analysis to determine if the variables within the scale appropriately loaded onto the intended constructs.

Table 3.

Predicted Intercorrelational Ranges of PGS-OMT Domains

	Social	Communication	Physical Navigation	Personal Care	Academic	Functional
Social	-	-	-	-	-	-
Communication	Very Large	-	-	-	-	-
Physical Navigation	Small	Small	-	-	-	-
Personal Care	Medium	Small	Large	-	-	-
Academic	Medium	Medium	Medium	Medium	-	-
Functional	Medium	Medium	Small	Medium	Medium	-

Correlational Ranges
0.10-0.29 = Small 0.30-0.49 = Medium 0.50-69= Large 0.70-0.89 = Very Large 0.90-0.99 = Nearly Perfect

Several precision indicators were evaluated utilizing the AMOS software including the normed fit index (NFI), goodness of fit index (GFI), and the normed comparative fit index (CFI), which are used to provide information pertaining to how well the items of a scale are organized with regard to proposed models. More specifically, through comparison of the null and target models, the NFI indicates how well the current model improves fit relative to the independent model. The NFI has a range of 0-1.0, with higher values indicating a better fit to the model, but has been shown to be sensitive to small sample sizes in that even perfect models may not reach 1.0. Further, the CFI compares the current model to a null or baseline model, the latter of which assumes that there is no observed correlation among all the variables (Bentler, 1990).

Generally speaking, to determine the model that best represents the organization of the data, it is recommended to compare the fit of several models with fit indices closer to one being more preferred or indicative of better fit (Hair Jr., Black, Babin, & Anderson, 2010). When evaluating fit indices and organizational models, there are different suggestions for acceptable values of each precision indicator though a statistic of .92 is typically considered adequate or acceptable (Hair Jr. et al., 2010). The current study compared the relative fit of multiple models, including a six-factor model that consisted of all six domains as independent factors and a two-factor model that assumed relationships among domains as they purported to measure comparable skills across related areas of functioning.

Validity Evidence Based on Relations to Other Variables. Concurrent validity evidence was initially sought through comparison of PGS-OMT domains with domains of the Vineland-II (See Table 4 for predictions). However, evidence was instead established through the examination of correlations between the three PGS-OMT factor domains with Vineland-II domains. Due to the differing nature of domains across the measures, as well as the operational

definitions or exact constructs measured by each tool, those domains that purported to represent specific areas of functioning were compared (e.g. PGS-OMT Interpersonal vs. Vineland-II Socialization). Cohen's (1992) work in classifying effect sizes provides guidelines for the strength of correlations. Cohen (1992) suggests values for small (r = .10 or r = -.10), medium (r = .30 or r = -.30), and large (r = .50 or r = -50) effect sizes. Social science researchers typically regard these values as the inner boundaries of each range. Hopkins (2002) suggests further ranges that extend Cohen's (1992) work to include values for nonexistent correlations (r = .00), very large correlations (r = .70 or -.70), and nearly perfect correlations (r = 90 or r = -.90).

Table 4.

Predicted Correlational Ranges of Vineland-II Domains with PGS-OMT Domains

	Vineland-II, TRF			
	Communication	Daily Living Skills	Socialization	Motor Skills
PGS-OMT				
Social	Medium	Medium	Very Large	Small
Communication	Very Large	Small	Medium	Small
Personal Care	Small	Very Large	Small	Very Large
Academic	Large	Medium	Medium	Small
Functional	Small	Very Large	Small	Medium
Physical Navigation	Small	Medium	Small	Very Large

Correlational Ranges
0.10-0.29 = Small 0.30-0.49 = Medium 0.50-69= Large 0.70-0.89 = Very Large 0.90-0.99 = Nearly Perfect

Utility. The argument for the perceived utility of the PGS-OMT was based on descriptive analyses of the Evaluation Survey completed by the teachers at the end of the data collection procedure. Descriptive analyses of items one through six were conducted. More

specifically, for items one through four, a higher percentage of endorsements would indicate that teachers considered the PGS-OMT easy to use, and that their time spent completing the PGS-OMT, Vineland-II, and other materials, was reasonable. The final question existed in open-ended form and would have been qualitatively evaluated by identifying trends in comments and teacher suggestions on how to potentially improve the ease of use of the PGS-OMT during the next revision process. However, none of the seven participating teachers responded to that item.

Chapter III

Results

Internal Consistency

To answer the first research question, the internal consistency coefficient (Cronbach's alpha) was calculated for each of the six domains on the PGS-OMT and total scale composite. The PGS-OMT possesses excellent internal consistency at the total scale level (.99). Table 5 details the Cronbach's alpha values, including the domain level coefficients, which were all within the excellent range. As predicted, these scores indicate that items on the scale fit well together.

Table 5.

Reliability Coefficients (Cronbach's alpha) of PGS-OMT Domains and Total Scale

PGS-OMT Factor	Number of Items	Cronbach's Alpha	Lowest Item-to-Total Correlation	Highest Item-to-Total Correlation
Social	10	.94	.61	.85
Communication	9	.95	.70	.86
Personal Care	10	.97	.69	.89
Academic	17	.95	.60	.93
Functional	8	.98	.68	.92
Physical Navigation	10	.95	.81	.94
Total Scale	**64**	**.99**	.44	.88

Item-to-total correlations were also evaluated at the domain and total scale level to determine which items, if any, did not function in a manner consistent with the progression of other items within the measure. Significant improvements in reliability statistics were not observed for the deletion of any item at the domain and total scale level. Appendix C presents the item-to-total correlations. Specifically at the domain level, no correlation was below the

commonly noted .30 criterion of acceptability (Sattler, 2008). This was also found true for the total scale level, and all item-to-total correlations were larger than .44 so no items were removed due to small item-to-total correlations.

Evidence for Internal Structure

Confirmatory Factor Analysis. To answer the second research question and further examine the internal structure of the PGS-OMT, multiple models were evaluated via confirmatory factor analysis. Fit indices were compared for each model respectively. The first two-factor (2F) model included the factors Adaptive Behavior and Academics.

While individual factor loadings were high (i.e., greater than .50) for each of the items within the factors (see Table 6 for a summary of the factor loadings for the two tested models), the Chi-Square Goodness of Fit statistic was significant ($\chi^2 = 6534.81$, $p < .001$). Indicating a significant difference between predicted and observed relationships, the model did not have adequate fit. Analyses of the 2F model revealed poor fit statistics for the NFI (.52), CFI (.61), and GFI (.26), none of which approached the requisite .92 criterion for acceptability. Additionally, the Standardized Root Mean Square Residual (SRMR) was much larger (.85) than tolerable, as values should be as small as possible and generally less than .05 (Meyers, Gamst, & Guarino, 2006). Values greater than .10 for the Root Mean Square Error of Approximation (RMSEA) are generally deemed unacceptable, and the RMSEA was found to be higher than that criterion in the current analyses (.14).

While analyses demonstrated better fit indices, the proposed six-factor (6F) model that regarded each domain of the PGS-OMT as an independent factor was not a good fit, as the model surpassed the threshold for significance and was subsequently rejected ($\chi^2 = 5034.44$, $p < .001$). Poor fit was also demonstrated by the NFI (.61), CFI (.71), and GFI (.39), all below the .92

suggestion of appropriate fit. Although the SRMR (.55) and RMSEA (.12) were smaller compared to the previous model, these values were still unacceptable according to the criteria of less than .05 and .08, respectively. Table 7 depicts the Goodness of Fit Indices for both the 2F and 6F models.

Table 6.

Factor Loadings for Domain Items for 2F and 6F Models

Domain/ Item	2F	6F
Comm.		
C01	**.73**	.83
C02	**.77**	.86
C03	**.57**	.69
C04	**.84**	.89
C05	**.58**	.71
C06	**.84**	.86
C07	**.80**	.86
C08	**.86**	.86
C09	**.82**	.85
Academic		
A01	**.81**	.79
A02	**.92**	.93
A03	**.86**	.86
A04	**.93**	.93
A05	**.90**	.91
A06	**.89**	.89
A07	**.91**	.93
A08	**.89**	.88
A09	**.89**	.91
A10	**.83**	.82
A11	**.89**	.91
A12	**.81**	.82
A13	**.61**	.62
A14	**.83**	.84
A15	**.85**	.86
A16	**.83**	.83
A17	**.90**	.91

Domain/ Item	2F	6F
Social		
S01	*.57*	.77
S02	*.83*	.91
S03	*.82*	.89
S04	*.55*	.69
S05	*.75*	.84
S06	*.67*	.79
S07	*.74*	.86
S08	*.64*	.77
S09	*.67*	.69
S10	*.53*	.62
S01	*.57*	.77
Phys. Nav.		
PN01	*.62*	.82
PN02	*.72*	.95
PN03	*.69*	.93
PN04	*.76*	.95
PN05	*.53*	.84
PN06	*.67*	.86
PN07	*.73*	.85
PN08	*.56*	.86
PN09	*.70*	.89
PN10	*.78*	.89

Domain/ Item	2F	6F
Pers. Care		
PC01	*.81*	.93
PC02	*.76*	.91
PC03	*.80*	.73
PC04	*.76*	.76
PC05	*.70*	.70
PC06	*.67*	.86
PC07	*.73*	.85
PC08	*.56*	.86
PC09	*.80*	.90
PC10	*.75*	.87
Functional		
F01	*.74*	.83
F02	*.73*	.86
F03	*.74*	.96
F04	*.71*	.65
F05	*.72*	.96
F06	*.76*	.84
F07	*.71*	.92
F08	*.66*	.93

Note. Italicized 2F item loadings comprise the Academic Factor in the hypothesized two-factor model, while bolded 2F item loadings comprise the Adaptive Behavior factor.

Table 7.

Fit Indices for Tested Models

	Fit Indices	2F Model	6F Model
Relative Fit Indices	NFI	.522	.632
	CFI	.606	.734
Absolute Fit Indices	GFI	.256	.413
	SRMR	.846	.546
	RMSEA	.143	.118
Model Comparison	AIC	6806.81	5326.441

Exploratory Factor Analysis. In response to the inadequate fit for both proposed models, an exploratory factor analysis (EFA) was conducted in SPSS to determine if there was an underlying structure present in the data, as well as to identify the organizational loadings of the items to emergent factors. Prior to conducting the EFA, the data set was assessed for suitability for the analyses using the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO). A KMO greater than .70 indicates that factor analysis is appropriate for the data set (Meyers, Gamst, & Guarino, 2006), and the KMO in the current analyses surpassed that criterion (KMO = .92). Principal axis factoring was utilized with a promax rotation and yielded a six-factor model that accounted for 81% of the variance. This six factor model, as well as a four and five-factor model, did not best represent the pattern of items. The six-factor model had minimal loadings with negative values for four of the eight item loadings. While the five-factor model had an eigenvalue greater than 1.00 and accounted for 77% of the variance, it was comprised of items that divided the Personal Care domain and had no other loadings. A four-factor model was also considered that accounted for 75% of the variance. Item loadings were minimal and divided a domain whose items were better explained by another factor. A scree plot of the eigenvalues

(Figure 1) demonstrated that the loadings were best explained by three larger factors with eigenvalues ranging from 27.34 to 3.34.

The EFA was again conducted; this researcher constrained the analyses to three factors and suppressed small coefficients with an absolute value less than .40. This model accounted for 72% of the variance. Items were then eliminated if they did not load onto any individual factor, or if the items cross loaded onto multiple factors with less than a .20 difference between the loadings. Items are often also removed on the basis of unacceptable item-to-total correlations; however, in the present analyses, no single item was found to have exceptionally low item-to-total values. Twelve of the 64 items were ultimately removed, and the EFA was once again conducted. The resulting three-factor model cumulatively accounted for 73% of the variance. Table 8 details item loadings for this model.

This resulting structure was comprised of the following factors: Adaptive behavior, containing 17 items assessing an individual's ability to physically navigate his or her surroundings and demonstrate adequate custodial self-care; Interpersonal, 14 items assessing an individual's ability to engage in appropriate interpersonal interactions, including the use of reciprocity and manners; and Functional Academics, a total of 21 items assessing an individual's ability to read, follow directions, and perform elementary mathematics computations. These factor domains consisting of 52 items were substituted for the original six PGS-OMT domains in the MTMM and instrument bias analyses. The three factor domains and original six domains were both used in the analyses discerning relations to other variables when compared to Vineland-II domains.

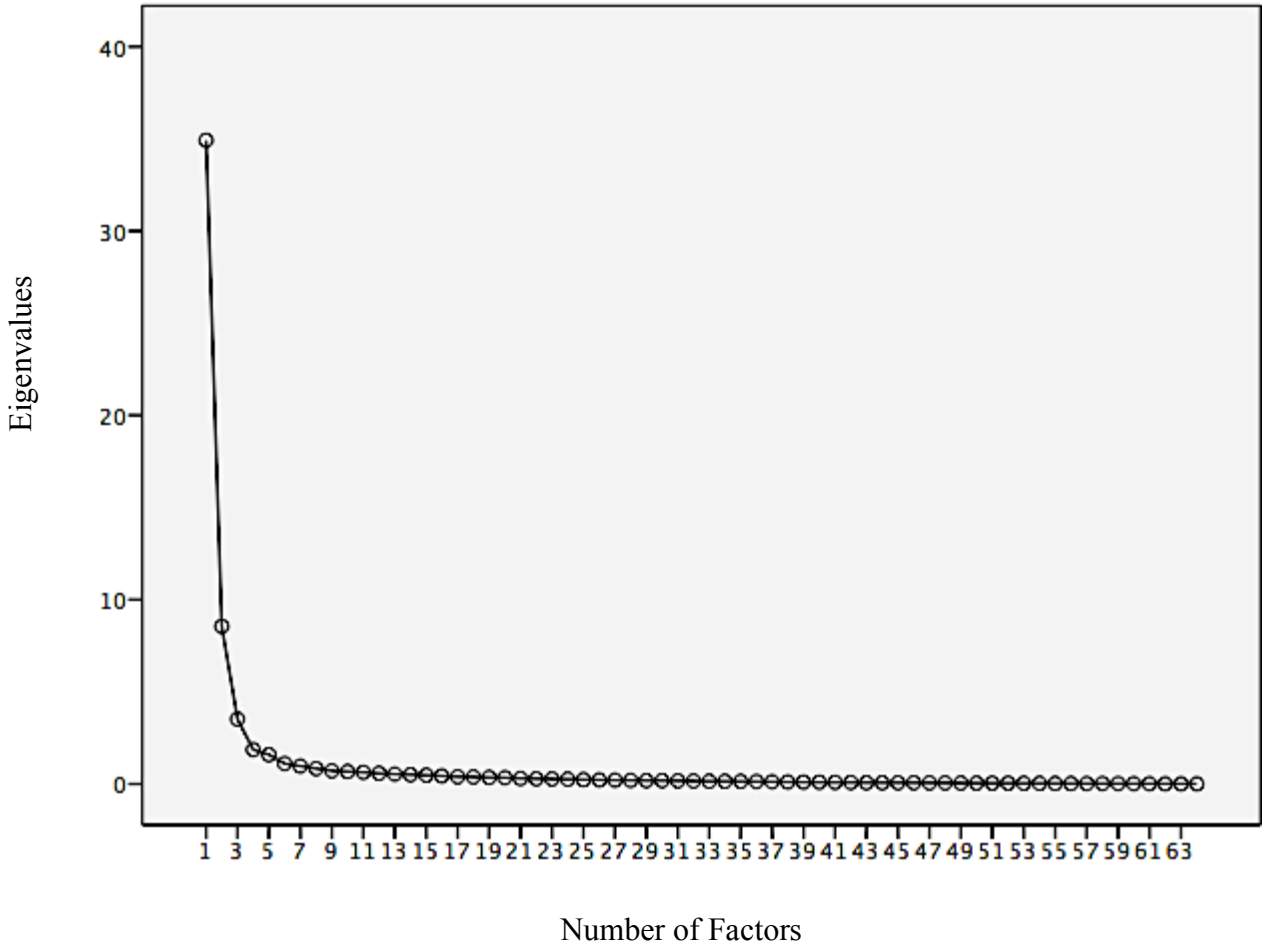

Figure 1. Scree Plot for the EFA of the PGS-OMT. This figure illustrates the number of extracted factors and corresponding eigenvalues from the current data sample.

Table 8.

EFA Item Loadings for Constrained 3F Model

Item	**Factor 1**
A02	.76
A03	.56
A04	.78
A05	.72
A06	.77
A07	.81
A09	.84
A11	.88
A12	.94
A13	.92
A14	.74
A15	.83
A16	.90
A17	.93
F01	.74
F02	.89
F03	1.01
F05	1.00
F06	.74
F07	.95
F08	1.03

Item	**Factor 2**
PN01	.85
PN02	.98
PN03	.95
PN04	.91
PN05	1.03
PN06	.94
PN07	.83
PN08	1.00
PN09	.89
PN10	.82
PC01	.73
PC02	.75
PC04	.70
PC05	.76
PC06	.44
PC09	.68
PC10	.67

Item	**Factor 3**
S01	1.09
S02	.74
S03	.57
S04	.68
S05	.70
S06	.66
S07	.83
S08	.73
S10	.57
C01	.83
C02	.73
C03	.89
C05	.81
C07	.51

To contrast the above EFA, an additional EFA was conducted with a varimax rotation to discern if a clearer distinction between item loadings could be achieved. The varimax rotation strives to achieve a "simple structure" in the data whereby factors are considered unrelated and are kept "independent" of each other during the analysis (Meyers, Gamst, & Guarino, 2006). Based on the scree plot of eigenvalues, the analysis yielded a four-factor model that accounted for 75% of the variance. However, the fourth factor had minimal loadings that divided the Academic domain. Of those items, most cross-loaded onto other factors, and the difference between the loadings was less than .20.

The EFA with a varimax rotation was again conducted, but the analyses were constrained to extract three factors and to suppress small coefficients with an absolute value less than .40. Similar to the previous EFA with a promax rotation, items were then removed if they cross-loaded onto other factors with less than a .20 difference between the loadings. Fifteen of the 64 items were ultimately removed, and twelve of these fifteen items were identical to the items removed in the previous EFA with the promax rotation. The EFA was again conducted, resulting three-factor model that accounted for 74% of the variance, and results were virtually identical to the three-factor EFA with a promax rotation reported earlier. Factor 1, accounting for 32% of the variance, combined the Functional and Academic domains; Factor 2, accounting for 25% of the variance, combined the Physical Navigation and Personal Care domains; and Factor 3, accounting for 17% of the variance, combined the Social and Communication domains.

Intercorrelations Between PGS-OMT Domains

To provide further evidence addressing the third research question, intercorrelations among domains are summarized in Table 9. Results were generally consistent in following the

pattern of predictions, but differed in the magnitude of the relationships, with correlations tending to be larger than expected.

Table 9.

Intercorrelations of PGS-OMT Domains

	Social	Communication	Physical Navigation	Personal Care	Academic	Functional
Social	-	-	-	-	-	-
Communication	.87*	-	-	-	-	-
Physical Navigation	.45*	.44*	-	-	-	-
Personal Care	.68*	.71*	.82*	-	-	-
Academic	.73*	.84*	.35*	.64*	-	-
Functional	.73*	.81*	.40*	.67*	.91*	-

Correlational Ranges
0.10-0.29 = Small 0.30-0.49 = Medium 0.50-69= Large 0.70-0.89 = Very Large 0.90-0.99 = Nearly Perfect
*Correlation is significant at the 0.05 level (1-tailed)

The correlation between the Social and Communication domains was in the very large range (r = .87), and was the only pairing to have a correlation in the predicted range. Patterns were present with other correlations; however, results indicated stronger relationships than what was originally anticipated. For example, correlations between the Physical Navigation domain and others were expected to be small, but were instead within the medium range in three of five pairings. The correlation between the Academic and Functional domains was within the nearly perfect range (r = .91). This finding indicates that these constructs are highly related and could overlap to the point of multicollinearity, whereby both are measuring the same construct.

Correlations were also computed at the factor domain level (Table 10). Patterns were consistent with the correlations at the domain level, and related areas demonstrated larger

relationships than unrelated areas. The correlation between the three factor domains and PGS-OMT Total composite score were all in the very large range (r = .81 to r = .87), which suggests scores of these factor domains comparably contribute to an individual's overall level of performance on the PGS-OMT.

Table 10.

Intercorrelations of PGS-OMT Factor Domains

	Functional Academics	Adaptive Behavior	Inter-personal	PGS-OMT Total
Functional Academics	**.98**	-	-	-
Adaptive Behavior	.45*	**.98**	-	-
Interpersonal	.75*	.55*	**.96**	-
PGS-OMT Total	.86*	.81*	.87*	**.98**

Correlational Ranges
0.10-0.29 = Small 0.30-0.49 = Medium 0.50-69= Large 0.70-0.89 = Very Large 0.90-0.99 = Nearly Perfect
*Correlation is significant at the 0.05 level (1-tailed)
Note: Cronbach's Alpha for each factor domain is indicated in bold.

Validity Evidence Based on Relations to Other Variables

In response to the third research question, the constructs delineated by the three identified factor domains of the PGS-OMT were compared with the domains of the Vineland-II to gather concurrent validity evidence (Table 11). One-tailed significance tests of the correlations were conducted to assess relationships within and across the measures. Results were qualitatively categorized using Cohen's (1992) classifications with Hopkins' (2002) extension for very small and very large relationships.

Consistent with researcher hypotheses, stronger relationships were observed between conceptually related domains and weaker relationships were observed between conceptually unrelated domains. For example, very large correlations were demonstrated between the

Interpersonal factor domain of the PGS-OMT and two of four pairings with Vineland-II domains (i.e., Communication and Socialization). The Adaptive Behavior factor domain was predicted to have small relationships with two of the four Vineland-II domains. Correlations were in the medium or large range for three of four pairings, with the exception of Motor Skills (.82), which was expected to result in a very large relationship. Although the correlations were larger than originally predicted, the pattern of these relations were in accordance with researcher expectations.

Correlations also tended to be higher between the individual factors and domains of the measures as compared to the correlation between the Adaptive Behavior Composite of the Vineland-II and the PGS-OMT Total composite score (.53). Greater variability was present in the strength of correlations when evaluating domain and factor level relationships than total score comparisons.

Table 11.

Correlations of Vineland-II Domains and PGS-OMT Factor Domains

	PGS-OMT				Vineland-II				
	Functional Academics	Adaptive Behavior	Inter-personal	PGS-OMT Total	Communication	Daily Living Skills	Socialization	Motor Skills	Adaptive Behavior Total
PGS-OMT									
Functional Academics	1.00								
Adaptive Behavior	.45*	1.00							
Interpersonal	.75*	.55*	1.00						
PGS-OMT Total	.86*	.82*	.87*	1.00					
Vineland-II									
Communication	.69*	.36*	**.71***	.67*	1.00				
Daily Living Skills	**.63***	.52*	.68*	.71*	.88*	1.00			
Socialization	.70*	.47*	**.83***	.76*	.83*	.83*	1.00		
Motor Skills	.44*	**.82***	.48*	.71*	.46*	.62*	.50*	1.00	
Adaptive Behavior Total	.45*	.38*	.54*	.53*	.88*	.90*	.78*	.53*	1.00

Correlational Ranges
0.10-0.29 = Small 0.30-0.49 = Medium 0.50-69= Large 0.70-0.89 = Very Large 0.90-0.99 = Nearly Perfect
*Correlation is significant at the 0.05 level (1-tailed)
Note. Bold values indicate pairings of similar constructs across measures.

Multitrait-Multimethod Matrix (MTMM). In order to discern whether intercorrelations among the constructs measured by both the PGS-OMT and Vineland-II demonstrated additional validity evidence, Campbell & Fiske's (1959) multitrait-multimethod organization of correlations was utilized. The first group was categorized by similar constructs across measures. For example, the Interpersonal factor domain of the PGS-OMT had an assumed relationship with the Socialization domain of the Vineland-II, and so larger correlations were expected. Second, domains measuring dissimilar constructs within the PGS-OMT were compared (e.g., Functional Academics and Adaptive Behavior). Correlations among dissimilar constructs within the Vineland-II were also considered. These correlations were expected to be smaller in magnitude than those involving related domains across measures. A third grouping was organized for dissimilar constructs across the two measures, such as those domains expected to have little to no relation to each other (e.g., Interpersonal factor domain of the PGS-OMT and Motor Skills domain of the Vineland-II). These were hypothesized to have lower correlational values than the former two groupings. See Table 12 for a full depiction of the different combinations for each of the three groups.

The first group, comprised of variables of similar constructs from different measures, resulted in correlations within the large and very large ranges (r = .63 to r = .83) with three of the four pairings in the very large range. The second group, comprised of variables of different constructs within the same measure, had notable variability in the strength of the correlations with values ranging from the medium to very large ranges on the PGS-OMT. Values ranged from r = .46 to r = .75. However, only one pairing, between the PGS-OMT Functional Academics and Interpersonal factor domains, was within the very large range. The removal of the problematic pairing resulted in values within the medium and large ranges. Dissimilar

constructs within the Vineland-II demonstrated smaller correlations than the first grouping; however, values for the two Vineland-II domain pairings within this group were not larger than what was found for the majority of comparisons in the third group. These two correlations were medium and large (r = .46 and r = .50, respectively). The last grouping, which should have resulted in the smallest correlations of all three groups, compared different constructs from different measures. Correlations were within the medium to very large ranges and more than half of the values (four of six pairings) were in the medium range.

Table 12.

MTMM: Organization of Correlations by Group

Group 1) Similar Construct-Different Measure	***r***
PGS-OMT Interpersonal & Vineland-II Communication	.71
PGS-OMT Interpersonal & Vineland-II Socialization	.83
PGS-OMT Functional Academics & Vineland-II Daily Living Skills	.63
PGS-OMT Adaptive Behavior & Vineland-II Motor Skills	.82
Group 2) Different Construct-Same Measure	***r***
PGS-OMT Functional Academics & PGS-OMT Adaptive Behavior	.45
PGS-OMT Functional Academics & PGS-OMT Interpersonal	.75
PGS-OMT Interpersonal & PGS-OMT Adaptive Behavior	.55
Vineland-II Communication & Vineland-II Motor Skills	.46
Vineland-II Socialization & Vineland-II Motor Skills	.50
Group 3) Different Construct- Different Measure	***r***
PGS-OMT Functional Academics & Vineland-II Socialization	.70
PGS-OMT Functional Academics & Vineland-II Motor Skills	.44
PGS-OMT Interpersonal & Vineland-II Motor Skills	.48
PGS-OMT Interpersonal & Vineland-II Daily Living Skills	.68
PGS-OMT Adaptive Behavior & Vineland-II Communication	.36
PGS-OMT Adaptive Behavior & Vineland-II Socialization	.47

Correlational Ranges
0.10-0.29 = Small 0.30-0.49 = Medium 0.50-69= Large 0.70-0.89 = Very Large 0.90-0.99 = Nearly Perfect

Instrument Bias

To address the fourth research question assessing bias based on demographic characteristics, independent samples *t*-tests were conducted to assess whether individuals who are identified as a minority group had significantly different scores than those who identified as a majority group. Ethnicity categorizations of participants were dichotomized into advantaged (i.e., White/Non-Hispanic and Asian) and disadvantaged (i.e., Black/Non-Hispanic, Hispanic, and American Indian/Alaskan Native/Pacific Islander) groups. Independent samples *t*-tests were conducted to determine any statistically significant difference between the means of the two groups for each of the three PGS-OMT factor domains, as well as the total scale score (Table 13). Results indicated no statistically significant difference between groups for two out of three factor domains, with the only exception being the Adaptive Behavior factor domain where mean scores were significantly higher for those in the disadvantaged group, t (115) = -2.11, p = .04. The effect size, as measured by Hedge's g, was medium (d = .55).

Table 13.

Comparison of Domain Score Means for Advantaged and Disadvantaged Groups

	Advantaged	**Disadvantaged**		
	(n = 100) *M (SD)*	(n = 17) *M (SD)*	p	Hedge's g
PGS-OMT Factor				
Functional Academics	28.57 (37.88)	17.12 (28.51)	.24	
Interpersonal	47.02 (26.76)	44.24 (26.20)	.69	
Adaptive Behavior	44.58 (38.66)	66.18 (41.66)	.04*	.55
PGS-OMT Total	118.82 (87.83)	125.35 (83.86)	.78	

Note. Hedge's g was used to calculate effect size for significant differences found for Adaptive Behavior. *p < .05

Utility

Teacher responses to items two and four of the Evaluation Survey provided evidence for the utility of the PGS-OMT. Mean response scores for each statement were between three and four on the four-point Likert scale (1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Agree*, and 4 = *Strongly Agree*) for six of the seven teachers who responded. These findings suggest that the majority of the teachers consider the amount of time spent completing the PGS-OMT to be reasonable (86%), and all participating teachers endorsed that the measure itself is easy to use. However, most of the teachers (n = 5) were dissatisfied with the data collection process in its entirety and answered 2 (*Disagree*) to statement three. This could be attributable to teacher dissatisfaction with completing the Vineland-II. Table 14 presents the percentages of teacher responses to statements one through five of the Teacher Evaluation Survey.

Table 14.

Percentages of Teacher Endorsements of Evaluation Survey Items

	Strongly Disagree	Disagree	Agree	Strongly Agree
Statement				
1. I consider the time spent on completing the Vineland for each of my students reasonable.	0%	43%	57%	0%
2. I consider the time spent on completing the P.G. Chambers School Outcomes Measurement Tool for each of my students reasonable.	0%	14%	86%	0%
3. I consider the overall time spent for data collection reasonable.	0%	71%	29%	0%
4. Overall, the P.G. Chambers School Outcomes Measurement Tool is easy to use.	0%	0%	86%	14%

Questions six and seven on the Evaluation Survey allowed teachers to select which aspects of the tool contributed to or hindered the ease of use of the instrument. A majority of the teachers (71%) opined that the PGS-OMT's ability to accurately capture students' functional capabilities was the strongest characteristic of the measure followed by the skill-level definitions that guide ratings (43%). There were minimal endorsements of perceived negative aspects of the PGS-OMT and two or fewer teacher ratings for any of the ten given characteristics. Table 15 details the percentages of teacher endorsements for each subjective characteristic of the scale being considered. Percentages are regarded individually rather than totaled, as teachers were able to select multiple indicators, or none at all, of what they liked best and least. The final question was open-ended and allowed teachers to provide explicit suggestions regarding subsequent revisions and improvements of the measure. None of the participating teachers opted to respond or provide feedback.

Table 15.

Percentages of Teacher Endorsements of Characteristics of the PGS-OMT

Characteristic	Percent Liked Best	Percent Liked Least
User friendliness	14%	0%
Clarity of instructions	13%	14%
Skill-level definitions	43%	14%
Rating scale	29%	29%
Comprehensiveness of tool	14%	0%
Adaptability for students with disabilities	14%	14%
Transdisciplinary team administration	29%	14%
Alignment of tool with core standards	0%	14%
Ability of tool to assess students' functional abilities	71%	0%
Ability of tool to assess students' academic progress	14%	14%

Chapter IV

Discussion

The availability of adequate alternate assessments that integrate evaluation of both academic functioning and adaptive behavior is generally lacking, as is the research assessing the psychometric properties of these assessments. The primary purpose of this study was to gather evidence for the technical quality and utility of an alternate assessment designed for students with severe, multiple disabilities in a New Jersey non-public school. While the PGS-OMT possesses adequate reliability, the validity of inferences to be drawn from scores requires further investigation. The initial six-domain structure of the measure was a poor fit, as indicated by factor analysis and correlations, and a three-factor domain organization better represented the data. Relationships between areas of functioning across the PGS-OMT and an established measure of adaptive behavior were found to be congruent with initial hypotheses, demonstrating stronger correlations between conceptually related areas. Limitations and suggestions for future research are discussed later in this chapter.

Reliability

The first research question addressed the reliability of the PGS-OMT, and strong internal consistency was established. Cronbach's alpha was calculated at the domain, factor domain, and total scale levels, and alpha values were consistent with initial predictions. Alpha values for each of the three factor domains were excellent and, at the total scale level, alpha was similarly robust (.99). Item-total correlations were high for all six domains and additionally suggest that the items fit well together. These results indicate that the PGS-OMT yields reliable scores that demonstrate adequate internal consistency.

Evidence for Internal Structure Validity

To answer the second research question, construct validity of the PGS-OMT was assessed by examining the tool's internal structure through multiple analyses. Results of confirmatory factor analysis demonstrated a poor fit for both models tested and an exploratory factor analysis indicated a three-factor model organization of the data. Intercorrelations between domains and factor domains of the PGS-OMT were additionally computed, and relationships were consistent with hypothesized patterns.

Confirmatory Factor Analysis. Additional evidence for the internal structure validity of the PGS-OMT was obtained through factor analysis. Confirmatory factor analysis (CFA) was used to examine a two-factor and six-factor model, both of which were difficult to interpret due to their high factor loadings but poor fit indices. Both CFAs revealed overall poor fit to the respective models, which were subsequently rejected.

While the choice to use CFA is generally driven by the measure's theoretical framework, it is important to note that the transdisciplinary team members responsible for creating this tool included items and skills they considered illustrative of the six content areas delineated by the PGS-OMT domains. As there was no empirical foundation upon which to base hypotheses of potential factor organization, these six domains were interpreted as individual factors and assessed accordingly. While individual factor loadings were high for the six-factor (6F) model, the overall model was a poor fit, as evidenced by a significant Chi-Square Goodness of Fit statistic. Meyers, Gamst, & Guarino (2006) provided guidelines for appropriate values of fit indices for factor models, and according to the authors' criteria, the 6F model fit indices were unacceptable. Similar results were found when examining a separate, hypothesized two-factor (2F) model.

Exploratory Factor Analysis. Because the evidence of internal structure validity derived from confirmatory factor analyses was poor, a subsequent exploratory factor analysis (EFA) was conducted to discern if any underlying factor structure was present in the data. Based on these results, a three-factor model accounted for 73% of the variance. The first factor is a combination of the Functional and Academic domains, which is more representative of the skills targeted in the school program rather than separate constructs purported to measure educational progress and functional skills independently. The second factor combines the Personal Care and Physical Navigation domains and is reflective of adaptive behavior in that items are reflective of an individual's mobility and ability to care for her or himself. Lastly, the third factor incorporates the Social and Communication domains and includes items about characteristics of interpersonal relationships. Similar results were demonstrated with an additional EFA using a varimax rotation.

Although the PGS-OMT was originally designed with six independent domains measuring distinct areas, the subjective rationale for inclusion of test items by transdisciplinary team members may have resulted in an a priori assumption of the relationship between those items and their respective domains. EFA results suggest that the PGS-OMT is actually measuring latent constructs represented by the three factor domains. These three areas better capture the relationships between the various areas of functioning and their shared variance.

Intercorrelations between PGS-OMT Domains. Evidence for the internal structure validity of the PGS-OMT was also obtained through examination of the intercorrelations between domains of the PGS-OMT and between factor domains, as well. There was variability in the strength of correlations between factor domains (r = .45 to r = .75), but patterns were again consistent with initial hypotheses of larger correlations between related areas. Patterns of the

correlations across the six domains of the PGS-OMT are supportive of the three factor domains illustrated by the EFA. The largest correlations, those between the Academics and Functional domains (.91), Social and Communication domains (.87), and Physical Navigation and Personal Care domains (.82), are consistent with EFA item loadings and further support the three-factor structure of the measure.

The link between academics and functional behavior is of great importance in the present study. Previous evaluations of alternate assessments found varying degrees of shared variance between measures of academic skills and adaptive behavior. Researchers have posited that it is ideal to keep functional skills and academic achievement distinct enough to be separate constructs (Kettler et al., 2010; Miller & Linn, 2000). That does not seem to be the case for the PGS-OMT. Intercorrelations were nearly perfect and were further supported by the factor analysis, which linked the Academic and Functional domains as loading strongly onto one factor. This may suggest that the PGS-OMT is measuring a separate Functional Academics factor that is more reflective of the needs of this specific student population rather than the curricular content designated by the Common Core State Standards.

This is consistent with a study that examined six states' alternate assessments of alternate academic achievement standards. Very large correlations were found between reading/language arts and mathematics subscales with the Adaptive Behavior Composite of the Vineland-II in three states with values ranging from .78 to .90 for elementary school students (Kettler et al., 2010). The authors similarly suggested that, in order to comply with federal legislation mandating clear assessment of academic areas, modification of alternate assessments should focus on making a clear distinction between the skills required for academic achievement and

adaptive behavior with a stronger emphasis on relating to the core academic domains required by each state.

Adaptive behavior research was originally derived from studies examining individuals with mental retardation and primarily highlighted self-care and social competence as indicators of an individual's adaptive functioning (Ysseldyke & Olsen, 1999). Extended to include conceptual skills and community navigation (e.g., buying items and making change), social skills (e.g., forming and maintaining interpersonal relationships), and practical skills (e.g., caring for one's personal hygiene and obtaining employment), adaptive behavior by this definition is represented in the content domains of the PGS-OMT. It seems that in contrast to six distinct areas, the PGS-OMT domains are interrelated. Some degree of overlap across content areas is expected and often acceptable, but when these areas become too strongly related, the construct validity of the measure is negatively impacted (Miller & Linn, 2000).

Relatedly, alignment to the state standards has been shown to be notably difficult in the research and is similarly true in this study (Thompson & Thurlow, 2000). Items on the Academic domain are indeed more reflective of what is considered adaptive behavior, such as measuring number skills as a function of phone numbers and home addresses rather than addressing the content delineated in the Common Core Standards (e.g., the relationship between multiplication and division). By examining the alignment of the Academic domain to the content standards outlined in the state's Common Core Standards, the PGS-OMT appears to poorly adhere to the standardized guidelines. It is important to note that distinct sets of standards are available for each grade level beginning at kindergarten and extending through twelfth grade with the following year's criteria building upon the foundational skills supposedly acquired in the previous year. The PGS-OMT is maintaining the same set of rating factors for students in

early education as for those in middle school. Thus, this measure is assuming that students will be continuing to work on the same skills for the duration of the enrollment in their program, given they remain at the school through the eighth grade.

Validity Evidence Based on Relations to Other Variables

Addressing the third research question, validity evidence based on relations to other variables was established through correlational analyses between the factor domains of the PGS-OMT and domains of the Vineland-II. Results were consistent with initial hypotheses demonstrating stronger relationships between related domains than those for unrelated domains. Individually, these results suggest that the compared domains measure similar constructs. For example, the Interpersonal factor domain of the PGS-OMT and Socialization domain of the Vineland-II demonstrated the strongest relationship with a correlation within the very large range (r = .83). However, similar very large associations (r = .71) were also found between the Interpersonal factor domain of the PGS-OMT and Communication domain of the Vineland-II, which questions what underlying construct the Interpersonal factor domain is measuring. Further, large and very large correlations between the Functional Academics factor domain of the PGS-OMT and the Communication, Daily Living Skills, and Socialization domains of the Vineland-II suggest that this component shares significant variance with areas of adaptive behavior.

Consistent with the evidence of internal structure validity demonstrated by the EFA and strong relations between PGS-OMT domains, the pattern of correlations in the multitrait, multimethod matrix (Campbell & Fiske, 1959) was additionally supportive of the three-factor model and was congruent with what was expected in the groupings. Promising patterns were demonstrated in the first grouping that compared similar constructs across different measures

(e.g., Interpersonal and Socialization). These correlations were the largest amongst the various comparisons, with the second grouping demonstrating slightly smaller correlations, and the smallest relations in the third grouping of divergent constructs across different measures.

The very large correlation demonstrated by the pairing of the Functional Academics and Interpersonal PGS-OMT factor domains (r = .75) within the second grouping was more consistent with the strength of pairings in the first group. The strength of this relationship could signify that the constructs are not being measured appropriately. It is likely that an individual's performance on the Functional Academics factor domain is impacted by the communication component of the Interpersonal factor domain. Functional Academics items require some degree of adequate communication to convey responses. The reliance on skills from an unrelated domain to demonstrate competency in another area of functioning can confound assessment results. For example, a mathematics assessment designed to measure a student's understanding of basic operations should not use lengthy word problems. Doing so transforms the assessment from one of mathematic computation to include reading comprehension. Some of an individual's ability to perform well is dependent on his or her ability to read and comprehend written material, rather than the ability to solely conduct mathematical calculations. The intertwined nature of requisite skills across areas results in an impure assessment of the targeted construct.

Demographic Bias

The fourth research question involved the presence of bias toward individual demographic variables, specifically including ethnicity. Participants were dichotomized into groups: advantaged (i.e., White/Non-Hispanic and Asian) and disadvantaged (i.e., Black/Non-Hispanic, Hispanic, and American Indian/Alaskan Native/Pacific Islander) and mean scores

across each factor domain were compared. Statistically significant differences between groups indicate difference mean ratings in that specific PGS-OMT domain. Independent *t*-test results revealed no significant differences in scores between advantaged and disadvantaged ethnic groups for a majority of PGS-OMT factor domains. Adaptive Behavior was the only exception, $t(115) = -2.11$, $p = .04$, and results indicate that ratings were higher in the disadvantaged group ($M = 66.18$). Individuals in the disadvantaged group were rated as having stronger gross motor skills (e.g., avoiding obstacles, navigating uneven surfaces) than those in the advantaged group ($M = 44.58$).

Although initial hypotheses predicted no significant differences between groups across all areas, the finding for this factor domain elicits less concern than a significant finding for the other two factor domains. Gross motor skills are less sensitive to cultural differences than other areas such as social skills and academic skills. As demonstrated by the correlations between Vineland-II domains during its evaluation process, the Motor Skills domain had the smallest correlations with the other domains (Sparrow, Cicchetti, & Balla, 2006). This result is also less concerning than a significant difference in Interpersonal scores, for example, because of the unique nature of student disabilities in this setting. Differences are more characteristic of specific impairments rather than cultural influences. However, it is important to note that the disadvantaged group was ethnically heterogeneous, which may limit the generalizability of results to any particular group. Further investigation is required to make substantiated conclusions.

Utility of the PGS-OMT

The fifth research question addressed teacher perceptions of the PGS-OMT and its

strengths and weaknesses. Evaluation Survey results indicate that teachers completing the PGS-OMT found the measure to be useful and time efficient for evaluating students with severe, multiple disabilities. Strongest support was found in the percentage of teachers (86%) endorsing that they agree with statements that the PGS-OMT is easy to use and the time to complete the measure is reasonable. Negative evaluations of the overall data collection process were attributed to the requirement of completing the Vineland-II in addition to the PGS-OMT, with 43% of teachers disagreeing with the statement that the time spent completing the Vineland-II for each of their students was reasonable. This finding does not detract from the encouraging utility support for the PGS-OMT because the Vineland-II was only required to obtain concurrent validity evidence for this study and is not typically utilized by P.G. Chambers School faculty.

Analysis of teacher endorsements of characteristics of the PGS-OMT liked best demonstrated a notable discrepancy between the number of teachers touting the tool's ability to assess students' functional skills (71%) versus their academic progress (14%), which further speaks to the underlying functional concentration of the items. The dearth of responses to open-ended questions on the survey prevented qualitative analysis of suggestions to improve this rating scale. Overall, these findings from the Evaluation Survey indicate that teachers perceived the PGS-OMT to be a useful and timely measure to best evaluate their students' levels of functional behavior.

Summary of Findings

Aggregating the findings in the current study, evidence is provided that PGS-OMT scores are a reliable assessment of SWDs. The PGS-OMT has good reliability, as evidenced by excellent Cronbach's alpha coefficients for the overall measure, three factor domains, and six

individual domains. This was an especially encouraging result considering the relatively small number of items comprising each domain.

While the PGS-OMT had large and very large correlations with the Vineland-II, providing some concurrent validity evidence, the constructs being assessed are ambiguous and convergent. The PGS-OMT is not a pure measure of academic content, based on the questionable underlying factor structure and intercorrelations between multiple PGS-OMT domains. The Social and Communication domains shared a relationship in the very large range (.87), and the correlation between Academic and Functional domains was nearly perfect (.91). Factor analyses demonstrated that these domains are better viewed as three broader factor domains, which is additionally supported by intercorrelations.

Adaptive functioning and academic achievement have been highly correlated in previous research, and it was suggested that clinicians take necessary precautions to ensure that they remained distinct constructs (Ysseldyke & Olsen, 1999; Miller & Linn, 2000; Browder et al., 2005; Kettler et al., 2010). The nearly perfect association between the Academic and Functional domains of the PGS-OMT suggests that they are measuring a unitary construct, coined Functional Academics in the present study. The skills embedded within this Functional Academics factor are not wholly reflective of the skills outlined in the state's curricular standards and require revision to be an adequate academic assessment. Conversely, should the transdisciplinary team wish to maintain this area on the PGS-OMT, additional items must be added that better align to the standards and encompass the skills and content detailed for each grade level being assessed.

Practical Implications

Results support the internal consistency of the PGS-OMT, suggesting that the items function consistently within the measure. Results also indicate that the original six content areas within the PGS-OMT (i.e., Social, Communication, Academic, Functional, Physical Navigation, and Personal Care) are better explained by three areas of functioning that encompass definitions of adaptive behavior and functional abilities. Revisions should be tailored to maintaining the three factor domains derived from the EFA, while refining the targeted constructs. Organizing the measure that remains parsimonious with this structure will result in a better representation of the relationships between items.

Indicated by correlations of PGS-OMT domains and factor domains with Vineland-II domains, related areas across the measures shared some portion of the variance in student scores; however, similarly large relationships were also demonstrated between unrelated constructs (i.e., PGS-OMT Interpersonal and Vineland-II Daily Living skills). Further supported by very large intercorrelations between PGS-OMT factor domains (i.e., Interpersonal and Functional Academics), the targeted constructs across areas were unclear and require refinement. Revision of factor domain items should integrate MTMM findings to ensure that the content measured by one domain does not require competency of skills from another. These results limit the support that can be given to the validity of inferences to be drawn from PGS-OMT scores.

Further, the inclusion of a combined score or overall aggregation of scores to indicate a predetermined level of proficiency is a requisite for measures to be utilized as an alternate assessment (Elliott & Roach, 2007). Supported by the excellent alpha coefficient at the total scale level, future revisions of the PGS-OMT could focus on combining information from the various domains to assess a broader concept of adaptive functioning. Like the Vineland-II,

which synthesizes data from three or four separate domains dependent on age to yield a comprehensive measure of an individual's level of adaptive behavior, scores from the PGS-OMT could be totaled, transformed to standard scores, and analyzed as a measure of an individual's functional capabilities.

The present version of the PGS-OMT possesses characteristics that should be maintained and utilized in future revisions. The multi-rater, team based approach to evaluating students ensures that at least one member has had direct interaction or observation of the student being rated. Teacher involvement on both the transdisciplinary team and in the data collection process encourages commitment and support for the measure, which is considered an important aspect of the alternate assessment process (Roach, Elliott, & Webb, 2005). However, much work is needed to improve the technical quality and alignment of the PGS-OMT.

Limitations

The generalizability of the results of this study is limited by multiple factors. The convenience sample of participants hinders the applicability of the findings to the general population and contributed to the disparity between the advantaged and disadvantaged groups. The heterogeneity of ethnicity of the disadvantaged group's participants makes it difficult to draw conclusions about demographic bias despite significant differences found between the means of the two groups in one area of functioning (i.e., Adaptive Behavior). The number of participants in the present study is also of particular concern with regard to the performed data analysis, specifically the factor analyses. CFA is typically utilized in studies with larger sample sizes (e.g., 200-400 subjects per approximately 15 variables) (Meyers, Gamst, & Guarino, 2006). Heuristics have been provided such as having a minimum of 100 subjects but preferring more than 200 subjects. Meyers, Gamst, & Guarino (2006) suggest that, for analyses with greater than

10 variables, anything less than 200 subjects could result in unstable parameter estimates. Because of the large number of variables (i.e., test items) being examined, the small sample size of 117 participants limited the strength of the analysis and, had there been a larger pool of subjects, the precision of the indices may have been improved.

An additional limitation was the sole comparison of the PGS-OMT to a measure of adaptive behavior. The transdisciplinary team sought to design a measure that not only incorporated assessment of functional skills but also a more reflective component of the academic capabilities of their unique population at the P.G. Chambers School. Realistically, the current analyses are only capturing one component of that goal, as they were restricted to examining the relationship of PGS-OMT domains with those of the Vineland-II. The inclusion of a measure of academic achievement may have improved the study and yielded further information regarding the validity of conclusions to be drawn from PGS-OMT scores.

Future Research

More research is needed to examine the psychometric properties and utility of alternate assessments, and more specifically, the PGS-OMT. Replication studies with larger sample sizes should be considered as the PGS-OMT is revised and improved. Past evaluations of alternate assessments have found significant threats to the validity of scores, most focusing on the misalignment of test content with the core standards (Browder et al., 2005; Roach, Elliott, & Webb, 2005). This is a flaw with regard to the content validity of these measures and should be addressed in future research of the PGS-OMT. The PGS-OMT may be lacking in the breadth and depth of skills delineated in the state standards and future investigation of the PGS-OMT's alignment to the standards by a panel of experts may provide explicit information about its adherence to those guidelines. Relatedly, an additional adjustment mentioned previously was to

improve upon the static nature of items for consecutive grades by increasing the complexity or difficulty in test items for older students. For example, assuming that an individual student masters the use of a calculator in the second grade (Item A17), scores from subsequent administrations of the PGS-OMT would not indicate acquisition of higher ordered proficiencies expected when administering the measure in sixth grade, such as understanding statistical variability (NGA & CCSSO, 2010).

In addition to content misalignment, deficient scoring is another prominent area of concern for alternate assessments (Elliott & Roach, 2007). An acceptable scoring rubric should incorporate three key criteria measuring the correctness of response, how students are able to apply their knowledge to other settings, and the level of support needed to successfully demonstrate the response or skill (Elliott & Roach, 2007). In its current form, the PGS-OMT encompasses two of those three characteristics with more emphasis needed on the generalizability of skills. Additional information is required to determine the impact of this deficit. Future research could be longitudinal in nature and involve a predictive validity component to assess how well the measure predicts future performance. A criterion-related validation study could compare measures of current adaptive functioning with future measures that are indicative of desired student outcomes (e.g., getting a job).

Future research should also focus on gathering additional evidence for the construct validity of PGS-OMT scores and include a comparison to a validated, standardized measure of academics, such as the New Jersey Alternate Proficiency Assessment (APA) that is designed to evaluate SWDs in the requisite content areas of language arts, mathematics, and science (New Jersey Department of Education, 2001). Integration of the APA with the simultaneous

administration of the Vineland-II allows for a pertinent triangulation analysis and could improve the overall strength of later replications.

Lastly, more information on the utility of the PGS-OMT should also be gathered and could be obtained by revising open-ended questions on the Evaluation Survey. Teachers were more likely to complete items with finite responses and so restructuring the last test item on the survey could increase the likelihood of teacher responding. A checklist or Likert-scale ratings similar to the initial five items may improve the ease of use. Qualitative analysis was not possible in the present study due to the lack of responses from all participating teachers regarding improvements and revisions. Detailed information about the perceived strengths and weaknesses of the PGS-OMT would be helpful in the revision process to maintain those strengths and specifically target any weaknesses or barriers to completion.

Conclusions

Federal regulations mandate the use of appropriate and psychometrically sound assessments that adequately reflect the requisite content areas in order to best inform educationally related decisions made at the school, district, and state levels (NCLB, 2001; IDEA, 2004). This is especially true for SWDs who were historically disregarded in accountability systems and did not often receive meaningful instruction in the abovementioned core content areas or functional areas deemed necessary to build their activities of daily living (Erickson, Thurlow, & Thor, 1995; Thurlow, 2004). Few alternate assessments exist that successfully integrate these two areas, and this integration has been shown to be an arduous, equivocal task in practice (Thompson & Thurlow, 2000). The PGS-OMT is one such tool that attempts to measure achievement and adaptive behavior for students with severe, multiple disabilities for whom standardized testing, with or without accommodations, is otherwise inappropriate.

Evidence from the current study indicates that the PGS-OMT yields reliable scores and is positively viewed by teachers in terms of its utility, specifically including the "reasonable" amount of time it takes to complete the measure in its entirety. While promising relationships were revealed between the PGS-OMT and an established rating scale of adaptive behavior, the overlapping nature of domain content and ambiguous structure of the six PGS-OMT domains obscure concrete or definitive conclusions about the validity of inferences drawn from PGS-OMT scores. Although the PGS-OMT is still new and the revision process is ongoing, the current study provides promising evidence that the PGS-OMT can eventually be used as a form of alternate assessment, and more proximally as a way to measure academic and adaptive growth of students enrolled at the P.G. Chambers School.

References

Agran, M., Alper, S., & Wehmeyer, M. (2002). Access to the General Curriculum for Students with Significant Disabilities: What It Means to Teachers. *Education And Training In Mental Retardation And Developmental Disabilities*, *37*(2), 123-33.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). Standards for educational and psychological testing. Washington, DC: Author.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238-246.

Browder, D. M., Ahlgrim-Delzell, L., Flowers, C., Karvonen, M., Spooner, F., & Algozzine, R. (2005). How states implement alternate assessments for students with disabilities. *Journal Of Disability Policy Studies*, *15*(4), 209-220.

Browder, D. M., Spooner, F., Algozzine, R., Ahlgrim-Delzell, L., Flowers, C., & Karvonen, M. (2003). What we knew and need to know about alternate assessment. *Exceptional Children*, *70*(1), 45.

Cameto, R., Knokey, A., Nagle, K., Sanford, C., Blackorby, J., Sinclair, B., & Riley, D. (2009). National Center for Special Education Research, National Profile on Alternate Assessments Based on Alternate Achievement Standards: A Report From the National Study on Alternate Assessments. NCSER 2009-3014. *National Center For Special Education Research*

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Rabinowitz, S., Sato, E., Case, B. J., Benitez, D., & Jordan, K. (2008). *Alternate Assessments for Special Education Students in the Southwest Region States*, (44), 7-19.

Elementary and Secondary Education Act, 20 U.S.C. § 6311 (b)(3)(C)(ii) (2002).

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

Elliott, S. N., Compton, E., & Roach, A. T. (2007). Building Validity Evidence for Scores on a State-Wide Alternate Assessment: A Contrasting Groups, Multimethod Approach. *Educational Measurement: Issues And Practice*, *26*(2), 30-43.

Elliott, S. N., & Roach, A. T. (2007). Alternate Assessments of Students with Significant Disabilities: Alternative Approaches, Common Technical Challenges. *Applied Measurement In Education*, *20*(3), 301-333. Doi:10.1080/08957340701431385

Erickson, R. N., Thurlow, M. L., & Thor, K. (1995). State special education outcomes, 1994.Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes (ERIC Document Reproduction Service No. ED 404 799).

Goldstein, J., & Behuniak, P. (2011). Assumptions in Alternate Assessment: An Argument-Based Approach to Validation. *Assessment For Effective Intervention*, *36*(3), 179-191.

Hager, K. D., & Slocum, T. A. (2005). Using Alternate Assessment to Improve Educational Outcomes. *Rural Special Education Quarterly*, *24*(1), 54-59.

Hager, K. D., & Slocum, T. A. (2008). Utah's Alternate Assessment: Evidence regarding Six Aspects of Validity. *Education And Training In Developmental Disabilities*, *43*(2), 144-161.

Hair, J., Black., W., Babin, B., Anderson, R. (2010). *Multivariate Data Analysis.* Upper Saddle River, NJ: Prentice Hall.

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

Hopkins, W. G. (2002). A scale of magnitudes for effect statistics. *A New View of Statistics*. Retrieved March 20, 2014, from http://sportsci.org/resource/stats/effectmag.html

Huai, N. (2004). *Development and validation of a universal screening instrument: The Brief Academic Competence Evaluation Screening System for identifying children at-risk of school failures*. Unpublished dissertation: University of Wisconsin-Madison.

Idaho Department of Education (1999). *Idaho Alternate Assessment*. Boise, ID: Author.

Individuals with Disabilities Education Act (IDEA) of 1997, 20 U.S.C. §1400 *et seq.*

Individuals with Disabilities Education Improvement Act (IDEIA) of 2004, 20 U.S.C. §1400 *et seq*.

Johnson, E., & Arnold, N. (2004). Validating an alternate assessment. *Remedial and Special Education*, *25*(5), 266– 275.

Kearns, J., Towles-Reeves, E., Kleinert, H., Kleinert, J., & Thomas, M. (2011). Characteristics of and implications for students participating in alternate assessments based on alternate academic achievement standards. *Journal of Special Education*, *45*(1), 3-14.

Kettler, R. J., Elliott, S. N., Beddow, P. A., Compton, E., McGrath, D., Kaase, K. J., Bruen, C., Ford, L., & Hinton, K. (2010). What do alternate assessments of alternate academic achievement standards measure? A multi trait-multi method analysis. *Exceptional Children*, *76*(4), 457-474.

Kleinert, H., & Thurlow, M. L. (2001). An introduction to alternate assessment. In H. L. Kleinert & J. F. Kearns (Eds.), *Alternate assessment: Measuring outcomes and support for students with disabilities* (pp. 1-12). Baltimore: Paul H. Brookes.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). Applied multivariate research: Design and interpretation. Thousand Oaks, CA: Sage.

Marion, S. F., & Pellegrino, J. W. (2006). A Validity Framework for Evaluating the Technical Quality of Alternate Assessments. *Educational Measurement: Issues And Practice*, *25*(4), 47-57.

Miller, M. D., & Linn, R. L. (2000). Validation of performance-based assessments. *Applied Psychological Measurement*, 2, 367–378.

National Governors Association Center for Best Practices (NGA Center), & Council of Chief State School Officers (CCSSO). (2010). *Common Core State Standards*. Retrieved from: www.corestandards.org

No Child Left Behind Act of 2001, PL 107-110, 115 Stat. 1425, 20 U.S.C. §§6301 *et seq.*

Rabinowitz, S., Sato, E., Case, B. J., Benitez, D., & Jordan K. (2008). Alternate assessments for special education students in the southwest region states (Issues & Answers Report, REL 2008–No. 044). Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Roach, A. T., Elliott, S. N., & Webb, N. L. (2005). Alignment of an alternate assessment with state academic standards: Evidence for the content validity of the Wisconsin alternate assessment. *Journal Of Special Education*, *38*(4), 218-231.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations (5th ed.)*. La Mesa: Jerome M. Sattler, Publisher, Inc.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1985). *Vineland adaptive behavior scales: Classroom edition*. Circle Pines, MN: American Guidance Service.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (2006). *Vineland adaptive behavior scales: Teacher rating form manual, second edition*. Minneapolis, MN: Pearson.

Thompson, S. J., Quenemoen, R. F., Thurlow, M. L., & Ysseldyke, J. E. (2001). *Alternate assessments for students with disabilities.* Thousand Oaks, CA: Corwin Press.

Thompson, S. J., & Thurlow, M. L. (2000). State alternate assessments: Status as IDEA alternate assessment requirements take effect (Tech. Rep. No. 35). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M. L. (2004). How state policies and practices for alternate assessment impact who is included in NAEP state assessments. *National Assessment Governing Board*: Conference on Increasing the Participation of SD and LEP Students in NAEP.

Tindal, G., McDonald, M., Tedesco, M., Glasgow, A., Almond, P., & Hollenbeck, K. (2003). Alternate Assessments in Reading and Math: Development and Validation for Students With Significant Disabilities. *Exceptional Children*, *69*(4), 481.

Towles-Reeves, E., Kleinert, H., & Muhomba, M. (2009). Alternate assessment: Have we learned anything new?. *Exceptional Children*, *75*(2), 233-252.

Wakeman, S. Y., Browder, D. M., Meier, I., & McColl, A. (2007). The implications of No Child Left Behind for students with developmental disabilities. *Mental Retardation And Developmental Disabilities Research Reviews*, *13*(2), 143-150.

Wisconsin Department of Public Instruction. (2010). 2009-10 Wisconsin alternate assessment for students with disabilities: Technical Manual. Monterey, CA: CTB/McGraw-Hill.

U.S. Department of Education. (2003). *Title I Regulation of Alternate Achievement Standards: Questions and Answers.* Washington, DC: Author.

Ysseldyke, J. E., & Olsen, K. (1999). Putting alternate assessments into practice: what to measure and possible sources of data. *Exceptional Children*, *65*(2), 175-185.

Appendix A

Student Demographic Questionnaire

Thank you for your participation in the Outcomes Measurement Study. Please complete the following survey for each student to the best of your ability based on your experience with the student.

1. **D.O.B**.: _____ **Age**: _____ years _____ months
2. **Diagnosis**: _____
3. **Date of enrollment**: _____
4. **Gender**: Male ___ Female ___
5. **Grade**: _____

   Grade Level: Early Childhood ____ Elementary _____ Middle School _____
6. **Ethnicity**:

   White/Non-Hispanic ___ Black/Non-Hispanic ___ Hispanic ___

   American Indian/Alaskan Native ___ Asian ___ Pacific Islander ___
7. **Has this student had extended absences? Length?** _____
8. **Vision** (check one):

   Within Normal Limits ___ Wears glasses ____

   Impaired-Unaided ___ Cortical Visual Impairment ___

   Eligible for services through Commission for the Blind ___
9. **Hearing** (check one):

   Within Normal Limits_____ Impaired-aided _____ Impaired-Unaided_____
10. **Communication**:

    Verbal ____ Non-verbal _____ Augmentative Device _____

    Student's Primary Mode of Communication _____
11. **Mobility**: Ambulatory _____ Non-Ambulatory _____

    Primary Mode of Mobility _____

Appendix B

Evaluation Survey

Thank you for participating in the Outcomes Measurement Data Collection and Study on October 14, 2013. We invite you to complete this brief questionnaire to help us understand how the Vineland and Outcomes Measurement Tool worked in your classroom. Your responses, constructive comments, and evaluation efforts are appreciated and will help us make improvements to our process in the future.
Please select the response that best characterizes the strength of your agreement or disagreement with each item. There are no right or wrong answers. Please do not skip any items.

***1. I consider the time spent on completing the Vineland for each of my students reasonable.**

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree

Please estimate the average amount of time in minutes needed to complete the Vineland for one of your students.

***2. I consider the time spent completing the P.G. Chambers School Outcomes Measurement Tool for each of my students reasonable.**

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree

Please estimate the average amount of time in minutes needed to complete the P.G. Chambers School Outcomes Measurement Tools for one of your students.

***3. I consider the overall time spent for data collection reasonable.**

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree

***4. Overall, the P.G. Chambers School Outcomes Measurement Tool is easy to use.**

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree

***5. Please indicate the feature(s) you like best about the P.G. Chambers School Outcomes Measurement Tool.**

- ☐ User friendliness
- ☐ Clarity of instructions
- ☐ Skill-level definitions
- ☐ Rating scale
- ☐ Comprehensiveness of tool
- ☐ Adaptibility for students with disabilities
- ☐ Transdisciplinary team administration
- ☐ Alignment of tool with core standards
- ☐ Ability of tool to assess students' functional abilities
- ☐ Ability of tool to assess students' academic progress

Other (please specify)

***6. Please indicate the feature(s) you like least about the P.G. Chambers School Outcomes Measurement Tool.**

- ☐ User friendliness
- ☐ Clarity of instructions
- ☐ Skill-level definitions
- ☐ Rating scale
- ☐ Comprehensiveness of tool
- ☐ Adaptibility for students with disabilities
- ☐ Transdisciplinary team administration
- ☐ Alignment of tool with core standards
- ☐ Ability of tool to assess students' functional abilities
- ☐ Ability of tool to assess students' academic progress

***7. Please add any comments, suggestions, or opinions with regard to the P.G. Chambers School Outcomes Measurement Tool that you think should be considered when revising it.**

Appendix C

Item-to-Total Correlations for PGS-OMT Domains

Social Domain

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
S01	26.95	276.20	.76	.68	.94
S02	28.93	265.51	.83	.79	.93
S03	29.15	274.91	.83	.80	.93
S04	28.35	284.02	.71	.62	.94
S05	28.74	274.37	.79	.68	.94
S06	29.03	275.26	.78	.65	.94
S07	28.29	272.57	.85	.76	.93
S08	28.59	273.18	.79	.67	.94
S09	29.43	287.37	.68	.55	.94
S10	27.45	292.70	.61	.39	.94

Communication Domain

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
C01	20.68	214.67	.83	.78	.94
C02	20.92	220.23	.86	.75	.94
C03	19.28	230.91	.70	.67	.95
C04	22.59	230.16	.82	.75	.94
C05	21.18	222.72	.74	.68	.95
C06	23.01	237.89	.81	.77	.94
C07	22.16	211.86	.84	.81	.94
C08	22.46	231.30	.81	.74	.94
C09	21.97	229.68	.81	.72	.94

Physical Navigation Domain

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
PN01	26.73	558.41	.81	.75	.97
PN02	27.62	543.05	.94	.91	.97
PN03	28.42	558.95	.90	.89	.97
PN04	28.09	551.60	.92	.91	.97
PN05	27.33	549.22	.86	.87	.97
PN06	28.33	555.93	.86	.82	.97
PN07	28.60	558.67	.82	.76	.97
PN08	27.15	545.88	.88	.90	.97
PN09	27.94	563.75	.88	.82	.97
PN10	28.33	559.59	.86	.84	.97

Personal Care Domain

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
PC01	19.07	332.32	.89	.90	.94
PC02	19.21	335.44	.86	.87	.94
PC03	19.90	356.23	.73	.62	.95
PC04	17.61	324.48	.76	.74	.95
PC05	17.01	326.58	.69	.68	.95
PC06	18.93	338.00	.82	.70	.94
PC07	18.95	327.86	.76	.67	.95
PC08	19.96	342.40	.79	.76	.94
PC09	19.05	335.20	.87	.80	.94
PC10	19.48	336.89	.83	.77	.94

Academic Domain

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
A01	23.22	915.92	.78	.78	.98
A02	24.80	912.92	.93	.91	.98
A03	23.98	911.16	.85	.85	.98
A04	25.14	924.15	.92	.89	.98
A05	24.55	902.97	.91	.95	.98
A06	25.15	938.56	.89	.90	.98
A07	24.80	908.97	.93	.96	.98
A08	24.45	926.96	.87	.87	.98
A09	24.79	930.95	.90	.84	.98
A10	24.36	939.46	.81	.74	.98
A11	25.21	941.22	.90	.93	.98
A12	25.50	960.55	.81	.92	.98
A13	26.14	1004.21	.60	.72	.98
A14	25.21	953.92	.84	.82	.98
A15	25.28	948.03	.84	.84	.98
A16	25.42	952.04	.80	.85	.98
A17	25.26	923.14	.90	.90	.98

Functional Domain

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
F01	7.71	124.05	.84	.79	.94
F02	7.86	129.60	.85	.74	.94
F03	8.30	132.42	.92	.91	.94
F04	6.31	129.27	.68	.560	.96
F05	8.56	139.15	.91	.92	.94
F06	8.00	131.90	.81	.73	.95
F07	8.36	132.43	.88	.83	.94
F08	8.60	141.86	.87	.87	.94

Total Scale

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
S01	138.74	11220.81	.59	.87	.99
S02	140.73	11073.58	.82	.92	.98
S03	140.95	11130.02	.82	.93	.98
S04	140.15	11249.52	.57	.81	.99
S05	140.54	11136.25	.77	.89	.98
S06	140.83	11183.18	.67	.88	.99
S07	140.09	11157.04	.75	.90	.98
S08	140.39	11181.89	.66	.88	.99
S09	141.22	11211.88	.67	.82	.99
S10	139.25	11275.48	.53	.73	.99
C01	139.86	11111.61	.74	.90	.98
C02	140.09	11123.67	.81	.92	.98
C03	138.45	11208.18	.64	.84	.99
C04	141.76	11141.74	.88	.97	.98
C05	140.35	11183.95	.62	.86	.99
C06	142.18	11229.48	.79	.95	.98
C07	141.33	11060.40	.81	.94	.98
C08	141.63	11200.80	.75	.96	.98
C09	141.15	11179.51	.78	.90	.98
PN01	139.23	11152.33	.56	.88	.99
PN02	140.12	11091.31	.66	.96	.99
PN03	140.92	11143.50	.63	.96	.99
PN04	140.60	11084.26	.71	.97	.99
PN05	139.84	11218.12	.44	.93	.99
PN06	140.84	11140.50	.60	.94	.99

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
PN07	141.10	11083.39	.68	.89	.99
PN08	139.65	11191.87	.48	.94	.99
PN09	140.44	11148.03	.64	.94	.99
PN10	140.84	11072.60	.73	.91	.98
PC01	141.50	11147.74	.75	.96	.98
PC02	141.64	11177.70	.69	.95	.99
PC03	142.33	11197.98	.78	.87	.98
PC04	140.04	11073.11	.71	.91	.99
PC05	139.44	11088.82	.65	.87	.99
PC06	141.37	11131.80	.79	.89	.98
PC07	141.39	11029.89	.83	.92	.98
PC08	142.39	11160.88	.75	.90	.98
PC09	141.49	11155.06	.75	.90	.98
PC10	141.92	11177.10	.69	.90	.99
A01	140.22	11077.62	.74	.90	.98
A02	141.80	11102.02	.80	.96	.98
A03	140.98	11085.29	.76	.94	.98
A04	142.14	11117.67	.84	.96	.98
A05	141.55	11072.80	.79	.98	.98
A06	142.15	11180.13	.77	.95	.98
A07	141.80	11084.66	.81	.98	.98
A08	141.45	11126.65	.79	.95	.98
A09	141.79	11168.69	.75	.93	.98
A10	141.36	11147.73	.78	.90	.98
A11	142.21	11202.82	.74	.97	.98
A12	142.50	11261.11	.66	.97	.99

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
A13	143.14	11373.17	.53	.88	.99
A14	142.21	11236.98	.70	.93	.99
A15	142.28	11190.50	.78	.95	.98
A16	142.42	11201.33	.74	.97	.98
A17	142.26	11124.18	.80	.96	.98
F01	142.07	11142.91	.76	.94	.98
F02	142.22	11179.17	.79	.95	.98
F03	142.66	11235.35	.76	.97	.98
F04	140.67	11129.69	.75	.90	.98
F05	142.92	11287.12	.75	.97	.99
F06	142.36	11189.68	.78	.96	.98
F07	142.68	11231.17	.74	.95	.98
F08	142.96	11313.70	.69	.96	.99