OBSERVATIONAL PLAY ASSESSMENT IN RECREATIONAL THERAPY (OPART): APPLICATIONS FOR THE VALIDATION OF ASSESSMENTS IN RECREATIONAL THERAPY

by

Nicole Christine Fox

A Senior Honors Project Presented to the

Honors College

East Carolina University

In Partial Fulfillment of the

Requirements for

Graduation with Honors

by

Nicole Christine Fox

Greenville, NC

May, 2018

Approved by: Dr. David P. Loy

Department of Recreation and Leisure Studies: College of Health and Human Performance

ABSTRACT

The need for standardized assessments in healthcare is rapidly growing as new discoveries are made. In the field of Recreational Therapy (RT), there are primarily agency specific developed assessments that have not yet been analyzed and proven reliable or valid[1]. As a growing occupation, it is critical that RT assessments are evaluated to provide evidence of the effectiveness of treatment. This study analyzed the reliability and validity of a new RT Assessment, the *Observational Play Assessment in Recreational Therapy* (OPART). The assessment was designed to provide recreational therapists the opportunity to assess levels of physical, cognitive, emotional, and social skills of children with disabilities while observed in play settings. Reliability and validity evidence was generated through the comparison of the OPART to another psychometrically sound assessment used in RT, the *Comprehensive Evaluation in Recreation Therapy* (CERT). Results suggested that the OPART had some evidence of interrater reliability (overall % agreement = .73; section agreement ranging from .60-.87%) with higher agreements when measuring functional skills with physical attributes (e.g., gross motor skills, fine motor skills, endurance/weakness). The OPART had some evidence of convergent validity with overall (r_s= .611; p=.016) and adjusted (r_s = .738; p=.002) scores when compared with the CERT. Further analysis of individual sub-sections of the OPART with the CERT suggested limited evidence of convergent validity as only one of the four sections approached statistical significance. The OPART also demonstrated *some* evidence of predictive validity in its' ability to accurately classify the functional level of the participants when compared to those assigned by staff familiar with individuals observed. The OPART was accurate in 8/15 (53.3%) cases while the standardized CERT was slightly more accurate at classifying participants in 9/15 (60.0%) cases. Overall, the OPART had some evidence of

convergent validity with the CERT in terms of adjusted scores (r_s=.738; p= .002) and total scores (r_s=.611; p= .016) with total scores. While this initial step to validate a new instrument was promising, it is clear that significant development and validation is indicated. Recommendations and future considerations are provided for the validation of agency specific RT assessments.

INTRODUCTION

Recreational Therapy (RT) is a growing field with practical implications amongst multiple settings. In order to create individualized treatment for clients, the APIE (assessment, plan, intervention, evaluation) process is used[1]. Of the APIE process, assessment is the most important component in order to identify information about the client for baseline records, safety precautions, goals and objectives[1]. Each year, the government spends millions of dollars to assess the outcomes of various student programs, however, recently it has become more important to incorporate assessments for children with disabilities[2]. Specifically for children with autism, it is important to determine the outcomes being assessed, which are commonly communication, social skills, and stereotyped behaviors[3]. Successful studies have also incorporated demographic and clinical assessment components to determine the severity of autism diagnosis and potential dual diagnosis[4].

While it is extremely important to develop an assessment for children with autism that measures the correct outcomes, it is also necessary to ensure that the instrument is reliable and valid. In a study conducted in 2009, the interrater reliability of the *Checklist for Autism in Toddlers* (CHAT)[5] was evaluated by determining percent agreement of the initial diagnosis[5]. This study determined the importance of an early diagnosis of autism with a reliable instrument in order to improve later treatment[5]. An additional study measured the convergent validity and interrater reliability of the *Pediatric Anxiety Rating Scale*[6] by using the Cronbach's alpha score

and comparing the assessment to the CGI-Severity assessment[6].

The *Failure Mode and Effects Analysis*[7], an assessment tool used in many healthcare facilities was analyzed to determine face, content, criterion, and construct validity[7]. This study emphasized the importance of validity when using assessments in healthcare, because without psychometric evidence there is often skepticism and little improvement to treatment[7].

It is extremely necessary to provide assessments in the healthcare field that are reliable and valid. Psychometrically sound assessments provide more accurate baseline data, discharge data, contraindications and precautions, and better goals and objectives. Recreational therapy is a growing field, however, currently many of the assessments within the field are developed specifically for an agency without testing psychometrics. While agency specific assessments can fulfill some benefits, standardized assessments are far superior and trusted in providing evidence of the effectiveness of treatment and completing the APIE (Assessment, Plan, Implement, Evaluate) process. For children with disabilities, standardized assessments can validate that treatment is effective and provide a measure to track the therapeutic progress of the individual[6].

Currently, professionals working with children with autism critically need assessments and programs with properly trained staff[6]. The following study provides an example of a newly conceptualized RT assessment instrument, the Observational Play Assessment in Recreational Therapy (OPART), and procedures utilized to obtain evidence of reliability and validity for the instrument.

METHODS

Development of the OPART

The OPART is an observational assessment of a child's functional skills based on a 7-point Likert scale ranging from 1= "total assistance" to 7= "complete independence". Recreational therapy graduate students at East Carolina University originally developed the OPART. The 24 original functional sections of the OPART were chosen based on an extensive literature review to determine traits and functional needs of children with developmental disabilities.

After visiting with agency staff and observing participants during play sessions, the primary researchers modified the assessment to include additional behavioral descriptions commonly exhibited by children with autism. The recreational therapists on staff were asked their input on the assessment, and additional modifications were made. The researchers also determined that the length of the original OPART was too long for many typical play sessions, therefore, the number of outcome sections was therefore combined and reduced to seven functional constructs. The following outcomes were measured: gross motor skills & mobility, fine motor skills/grasp and release, endurance & weakness, visual acuity & hearing deficits & perceptual problems, communication & social skills, motivation & attitude & self-confidence, and cognitive skills & decision making. Higher scores on the OPART indicated higher functional levels.

Comprehensive Evaluation in Recreational Therapy-(CERT)- Psychological/Behavioral[8]

When testing the validity of a new instrument, it is important to compare results to an already established instrument or a validation instrument[3]. The CERT was used as a validation instrument to help provide evidence of validity for the OPART. The CERT is a reliable and valid

assessment[8] with a physical and cognitive version. The CERT cognitive can be scored in 20 minutes when assessing a group of 15-20 clients[8]. The outcomes are scored on a scale of 0-4. Lower scores on the CERT indicate higher function. However, for the purpose of this research, the CERT scores were reversed to better compare to the OPART (0=4, 1=3, 2=2, 3=1, 4=0). Table 1.1 outlines the similar outcomes between the OPART and CERT tested to determine convergent validity.

Table 1.1: Assessment Outcome Comparisons

OPART	CERT
Gross Motor Skills & Mobility posture	General coordination and
Endurance & Weakness	Strength/Endurance
Communication & Social Skills	Ability to form relationships, style of group interactions, ability to handle conflict when directly/indirectly involved, Response to group structure, leadership ability in groups, group conversation, sexual role in group
Motivation, Attitude, & Self Confidence	Attendance, performance in organized activities, performance in free activities, competition in groups, Attitude toward RT, response to therapist structure, expression of hostility, frustration tolerance, attitude toward group decisions, Appearance
Cognitive Skills & Decision Making	Decision making ability, judgment ability, attention span, memory for group activities

Observer Training

In order to ensure that all observers were in consensus with OPART behavioral descriptions, team members were trained by watching videos of children with autism and completed the assessments until they reached 80% agreement. The primary researchers also provided more operational behavioral definitions at each level to promote improved differentiation between the seven functional levels.

Throughout data collection, all participants remained anonymous to research observers. Each participant was assigned a number and the agency director gave a physical description of each participant to observers as the 15-minute observation period began (i.e., "participant #1 is wearing a green shirt"). Two observational assessments of the CERT and OPART were then implemented to provide a comparison of results and subsequent validation.

Participants

Participants (N=15) in this validation study were selected by the agency program director based on the following criteria: a) diagnosis on the autism spectrum, b) participant at the site during data collection days, c) obtained consent from parents. Participants were selected throughout a range of functional levels, varying from low (n=5), mid (n=7) to high (n=3). Some participants (n=9) had other diagnosis in addition to ASD such as anxiety (n=2), ADHD (n=5), intellectual disability (n=4) and epilepsy (n=2). Participants were primarily male (n=13), but females were included (n=2). Various races were incorporated amongst participants including Caucasian (n=9), African American (n=5) and Latino (n=1).

Validation Procedures

Convergent Validity

Convergent validity is a type of validity that assumes that different measures related to the same construct should yield similar results if the measures are valid[9]. Since the CERT is a standardized and psychometrically sound assessment used with children with developmental disabilities[8] comparing the results of this instrument to the OPART provides some evidence of convergent validity if results were similar. A recreational therapist and a recreational therapy student observed scored one child with the OPART during the same fifteen-minute observation period. Simultaneously, a different recreational therapist and a RT student observed and scored the same child with the CERT. During the assessments, the children participated in an activity commonly used during their program such as arts and crafts, free play activities, organized recreational activities, and homework. Each observation period was fifteen minutes with one minute in between observation periods to finalize notes and documentation. After data collection, total scores and individual section scores were compared to determine if similar results were found between the CERT and OPART.

Interrater Reliability

Interrater reliability is defined as the measure of agreement amongst multiple raters observing the same construct[9]. To measure interrater reliability the observers' results were compared based on their observations of each child using percent agreement to determine consistency. In order for the assessment to have a high interrater reliability, the therapists must have .80 percent agreement[10.] Level of difference was used for each item on the assessment to determine the consistency among raters. This calculation was done for each item on each

child's assessment to determine the disagreement amongst therapists. In addition, a comparison of scores between the CTRS vs. non-CTRS (e.g., RT student) were compared to determine if experience level of observer impacted the over or under scoring of observed behaviors.

Predictive Validity

Predictive validity is the ability of an instrument to predict a score on the construct that it is measuring[9]. Prior to data collection, the agency director categorized the fifteen participants into "high functioning,' "middle functioning," and "low functioning." Scores of both the OPART and CERT were used to determine if rank order of the cumulative scores placed participants in an appropriate rank order. Because this is the first attempt to validate the OPART, criterion scores (i.e., score X or higher = high function, etc.) were not established. Therefore, looking at the rank order of cumulative scores may provide some evidence of the ability to predict the functional level of each participant based their rank order of cumulative scores. A comparison of both the OPART and CERT were made to determine which had better sensitivity to provide the ability to rank the functional ability of participants.

Validation Results

This study concluded that the OPART has moderate evidence of reliability and validity. Convergent validity scores for total validity (r_S=.611; p=.016) and adjusted validity (r_s = .738; p=.002) demonstrate that the OPART is a comparative instrument to an already established assessment. OPART determined adequate interrater reliability in the categories related to physical functional outcomes. However, the other four psychosocial categories had interrater reliability scores lower than the .80 acceptable criterion. Finally, the OPART showed moderate

predictive validity in that it grouped some of the lower functioning participants with lower scores and higher functioning participants with higher scores. The Spearman rank-order coefficient was chosen because it measures two variables on a scale[10]. Since OPART and CERT both use a Likert scale style, this coefficient was the most practical to use for data analysis. Spearman's coefficient can also determine the strength and direction of a relationship between two variables, and therefore can determine the correlation between the two instruments[10]. The following is a detailed explanation of validation results.

Convergent Validity

The OPART is ranked on a 1-7 scale within 7 functional outcomes. A higher score indicates higher functioning, and the highest score possible is 49. Because scores for the CERT were inversely scored to the OPART, CERT scores were reversed in order to make a comparison to the OPART. The reversed scale is 0-4, with a 4 indicating the highest functioning. There were 24 categories with a total score of 96. Relative percentage scores were calculated for both the OPART and CERT by taking the observed score and dividing it by the total possible relative scores provided.

Table 1.2 Convergent Validity

Modality	**Correlation Coefficient**	**Sig (2-tailed)**
Communication & Social Skills	.535	.049
Motivation, Attitude & Self Confidence	.600	.208
Cognitive Skills & Decision Making	.480	.070
Gross Motor Skills & Mobility	.369	.175
Adjusted OPART & Adjusted CERT	.738	.002
Total OPART & Total CERT	.611	.016

Table 1.3: Comparing OPART vs. CERT Total Scores

Participant	OPART Average/ (relative score)	CERT Average/ (relative score)
A1	31.7 (.65)	66.4 (.69)
A2	42.5 (.87)	84.3 (.88)
A3	45.5 (.93)	87.9 (.92)
A4	42 (.86)	72 (.75)
A5	38.5 (.79)	73.6 (.77)
A6	35.5 (.72)	78.3 (.82)
A7	38.5 (.79)	78.8 (.82)
A8	45.4 (.93)	92.1 (.96)
A9	38.5 (.79)	70.8 (.74)
B1	40.5 (.83)	76.6 (.80)
B2	42.6 (.87)	76.1 (.79)
B3	38 (.78)	74.4 (.78)
B4	45 (.92)	81.3 (.85)
B5	46.7 (.95)	79.3 (.83)
B6	49 (1.0)	86.9 (.91)

Inter-observation Reliability

The inter-observer reliability was strong (i.e., greater than .80 agreement) for the categories of gross motor skills/mobility, fine motor skills/grasp/release, and endurance/weakness. A "strong" or acceptable reliability is defined as greater than .80 agreement[9]. The other four categories (i.e., visual acuity, hearing, & perception; communication & social skills; motivation, attitude, & self-confidence; cognitive skills & decision making) did not have strong or acceptable inter-observer reliability coefficients (i.e., agreement lower than .80) and require further behavioral description revisions to improve reliability. For two participants, percent agreement was perfect and there was no difference in ratings among observers. Percent agreement was above .80 for each participant six times, with an

overall agreement of .40. The OPART was administered by one CTRS and one non-CTRS RT student. The average standard deviation between a CTRS vs. non-CTRS is 2.67. Scores were adjusted to account for certain behaviors that were not observed during the fifteen- minute session, primarily strength and endurance.

Table 1.4: OPART Percent Agreement

OPART Outcome	% Agreement
Gross Motor Skills and Mobility	0.80
Fine Motor Skills & Grasp/Release	0.87
Endurance & Weakness	0.87
Visual Acuity, Hearing & Perception	0.67
Communication & Social Skills	0.67
Motivation, Attitude, Self Confidence	0.67
Cognitive Skills & Decision Making	0.60

Table 1.5 OPART Participant % Agreement & Level of Difference

Participant	% Agreement	Level of Difference
A1	.14	1.5
A2	.86	.14
A3	.71	.71
A4	.71	.17
A5	.71	.43
A6	.86	.29
A7	.57	1.1
A8	.86	.29
A9	.57	.57
B1	.86	.14
B2	.71	.43
B3	.71	.57
B4	1	0
B5	.71	.57
B6	1	0

Table 1.6 CTRS vs. Non CTRS OPART Participant Scores			
Participant	CTRS	Non-CTRS	Standard Deviation
A1	36.4	27	9
A2	42	43	1
A3	46	45	1
A4	42	42	1
A5	37	40	3
A6	36	35	2
A7	37.3	39.7	8
A8	44	46.7	2
A9	37.3	39.7	4
B1	41	40	1
B2	44.3	40.8	3
B3	36	40	4
B4	45	45	0
B5	44.3	49	4
B6	49	49	0
Average Score	**41.2**	**41.5**	
*Note: OPART average scores were modified to account for response			

Predictive Validity

The predictive validity of the OPART allowed the observers to primarily categorize participants that were high functioning and low functioning. Prior to observation, the agency director categorized the function of the fifteen participants as “high”, “mid”, or “low”. Upon OPART scoring, total scores were placed in rank order and assigned function level based on their order. The upper third was considered “high”, middle third as “mid”, and lower third as “low”. One of the participant scores was tied with two others at 38 in the “low” so it made practical sense to assign the participant as the other two participants in the “low” functional category. Results (see table 1.7) indicated that 8/15 or 53.3% of OPART scores were accurately assigned

with the functional category assigned by the agency. Considering this was the first attempt to validate, the low predictive validity for the OPART was not unexpected. The CERT accurately assigned 9/15 participants to their functional category, or 60%. Establishing a criterion score would help with predictive validity in future tests. Another consideration is that some of the scores may have varied based on the setting that the participants were observed. Many participants enjoyed free play activities more than structured activities, which could have caused negative or positive behaviors to become magnified and scores to increase or decrease.

Table 1.7 Average OPART Scores & Participant Functioning

Participant	OPART Average Score	Functioning Level
B6	49	Mid
B5	46.7	Mid
A3	45.5	Mid
A8	45.4	High
B4	45	High
B2	42.6	Mid
A2	42.5	High
A4	42	Mid
B1	40.5	Low
A5	38.5	Low
A7	38.5	Low
A9	38.5	Low
B3	38	Mid
A6	35.5	Mid
A1	31.7	Low

*NOTE: Assigned to “low” category since score equaled 2 others in low range.

Table 1.8: Average CERT Scores & Participant Functioning

Participant	CERT Average Score	Functioning Level
A8	92.1	High
A3	87.9	Mid
B6	86.9	Mid
A2	84.3	High
B4	81.3	High
B5	79.3	Mid
A7	78.8	Low
A6	78.3	Mid
B1	76.6	Low
B2	76.1	Mid
B3	74.4	Mid
A5	73.6	Low
A4	72	Mid
A9	70.8	Low
A1	66.4	Low

DISCUSSION

Summary of Findings

Overall, the OPART has the potential to be a widely used assessment for agencies that work with children with autism. There was also some evidence that the OPART had similar results to a proven assessment used in recreational therapy (CERT). The interrater reliability proved to be adequate but much higher amongst observers with the physical indicators than the psychosocial items. This is typical for observational assessment instruments. The *Functional Independence Measure* (FIM) is another observational assessment that typically has higher interrater reliability among physical measures (e.g., ambulation, transfers) than psychosocial measures (e.g., social interaction, problem solving). Additional revisions and testing of the OPART is indicated.

Limitations

Although it is a standardized and validated assessment, the CERT cognitive assessment was not the ideal choice to validate the OPART. The CERT is designed to be administered for participants age 10 and older who have been observed previously prior to the administration of the instrument. Given the time frame of this study and the population, it was not possible to administer the CERT under ideal conditions. The observers who administered the CERT and OPART were also neutral to the participants. These assessments would likely be easier to administer and essentially more accurate if they were administered by an employee or recreational therapist who interacted with the participants on a regular basis. However, the decision was made in this project to use unbiased observers so that the assessment only included those behaviors demonstrated during the 15-minute observation period without knowledge of previous information on participants.

The OPART was initially designed for a multidisciplinary agency; however, the setting was changed to the Autism Society, an agency specifically focused on children with autism. Had this been known previously, the OPART would have incorporated more sensory items such as repetitive behaviors, repetitive speech, decision-making, and environmental stimuli. The items on the OPART when it was administered were an even mix of physical and cognitive indicators, however, the cognitive items were much more subjective than the physical ones, as such behaviors are more difficult to observe.

Future Research and Implications for Practice

In the future, the OPART would need to be modified to include more outcomes specific to Autism Spectrum Disorder. Additional sensory outcomes would be incorporated to yield more

accurate results and predictive validity for populations that include sensory deficits. Adjustments in the behavioral indicators, particularly those psychosocial outcomes, may improve reliability and validity of the instrument. The descriptors would also need to be modified to include "OR" rather than "AND", since some participants may have one deficit and not the other (e.g., hearing problems but no issues with vision). Agency staff would also need to be incorporated more to see if the OPART would be a functional tool to use in their setting. Using the OPART with therapists and staff familiar with clients can provide even more specific diagnostic information that would likely improve the validity and reliability of the instrument. Additional is testing is needed to further validate predicative validity of the instrument so it can indicate accurate functional levels. Finally, there is also a need to further test the OPART to determine criterion scores (i.e., scores for low, moderate, and high function). Overall, there is encouraging preliminary evidence that the OPART may have the potential to be useful to recreational therapy practitioners working with participants with autism when creating treatment plans and tracking improvements throughout the program.

Recreational therapists can use the methods and procedures demonstrated in this article to validate agency created instruments in practice. While there are many methods to provide psychometric evidence, recreational therapists should consider: a) comparing results to an already established instrument (convergent validity); b) compare results between observations between two therapists (interobservation reliability); c) determine if instruments have the sensitivity to "predict" functional levels accurately (predictive validity).

REFERENCES:

1. Carter MJ, Van Andel GE, Robb GM. *Therapeutic Recreation: A Practical Approach*. 2nd ed. Prospect Heights, Ill: Waveland Press; 1995.

2. Gillies RM. The Role of Assessment in Informing Interventions for Students with Special Education Needs. *International Journal of Disability, Development and Education.* 2014;61:1-5.

3. Ben Itzchak, E, Lahat E, Burgin R, Zachor AD. Cognitive, behavior and intervention outcome in young children with autism. *Research in Developmental Disabilities*. 2008:29:447-458.

4. Pellecchia M, Connell JE, Kerns CM, Xie M, Marcus SC, Mandell DS. Child characteristics associated with outcome for children with autism in a school-based behavioral intervention. *Autism*. 2016;20:321-329.

5. Khowaja M., Robins D. (2013) Checklist for Autism in Toddlers (CHAT). In: Volkmar F.R. (eds) *Encyclopedia of Autism Spectrum Disorders*. Springer, New York, NY.

6. . Storch EA, Wood JJ, Ehrenreich-May J, et al. Convergent and discriminant validity and reliability of the Pediatric Anxiety Rating Scale in youth with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2012;42:2374-2382.

7. Shebl NA, Franklin BD, Barber N. Failure mode and effects analysis outputs: are they valid? *BMC Health Services Research*. 2012;12:150-155.

8. Burlingame J, Blaschko TM. *Assessment tools for recreational therapy and related fields*. 4th ed. Enumclaw, WA: Idyll Arbor, 2010.

9.. Colman AM. *A Dictionary of Psychology*. 4th ed. *Oxford University Press*; 2014.

10. Spearman's Rank-Order Correlation using SPSS Statistics. Spearman's Rank Order Correlation using SPSS Statistics-A How-To Statistical Guide by Laerd Statistics. https://statistics.laerd.com/spss-tutorials/spearmans-rank-order-correlation-using-spss-statistics.php. Accessed January 25, 2018.