

# Research into Practice: Methodology Insight

## Measurement Basics: A Must for TR Professionals Today

**Ramon B. Zabriskie**

It is recognized that there is an increased demand in healthcare today for the use of valid and reliable measurement. Current resources available to therapeutic recreation professionals including published materials and efforts among educational and professional organizations are presented. Basic measurement concepts related to scale development and evaluation including scope, theory, format, validity, reliability, and usability are discussed. Recommendations for therapeutic recreation professionals are made.

**KEY WORDS:** *Measurement, Assessment, Outcomes, Validity, Reliability*

In today's age of accountability, the demand for effective measurement in therapeutic recreation programs and services has never been greater. It is not only a prerequisite to providing appropriate interventions to the clients that we serve, but is an integral part of dictating the efficacy of services, the reimburs-

ability of services, and ultimately whether we remain in business or employed in order to provide services at all. The ability to perform valid and reliable measurement is essential to client assessment, which is the first step in the therapeutic recreation process and provides the foundation for everything else that the thera-

---

*Ramon Zabriskie, Ph.D., CTRS, is an Assistant Professor at Brigham Young University. His primary research is focused on family leisure and therapeutic recreation.*

---

\* Address correspondence to Ramon B. Zabriskie, Department of Recreation Management and Youth Leadership, Brigham Young University, 273 RB, Provo, UT 84602-2033. Phone: (801) 422-1667, email: zabriskie@byu.edu

peutic recreation professional does in terms of client care (Austin & Crawford, 2001). The assessment process not only provides vital information regarding appropriate program placement and specific interventions provided, but must be precise enough to provide necessary baseline information that can be compared to related outcome measurement after program involvement (Stumbo, 2002).

The valid and reliable measurement of client outcomes is also a key tool for determining effectiveness of services for individual clients, and has become essential for the demonstration of quality services, accountability, and continual performance improvement in health and human services (JCAHO, 2003; McCormick, 2003). Mordock (2000) argues that the current health-care environment demands the use of assessment tools that produce valid, reliable, and meaningful client information as well as clear outcome measures that are sensitive enough to provide accurate data that can be used to demonstrate change. With such a demand for effective measurement, therapeutic recreation professionals, more than ever before, are sensing the urgent need to obtain or develop psychometrically sound instruments that can be utilized in their assessment and outcome measurement procedures.

This article will discuss some of the resources that are available to therapeutic recreation professionals as they consider their current measurement procedures. It will also provide an overview of some of the measurement basics related to scale development. It is hoped that this discussion will provide usable information and resources that can assist therapeutic recreation professionals whether considering new scale development, evaluating established scales for use in their programs, or refining the current measurement approaches.

## Measurement Resources

The use of tested and standardized instruments is much more likely to provide the kind of valid and reliable client information necessary in today's environment. Fortunately, the therapeutic recreation profession seems to be responding to the current need to refine mea-

surement skills and procedures. New instruments with therapeutic recreation applications continue to be developed, tested, and presented through scholarly journals (Ackerman, 2002; Klosock, Crilly, Ellis, & Lammers, 1996; Peebles, McWilliams, Norris, Park, 1999; Petrick, 2002; Shapiro & Yun, in press). Compilations of assessment tools for therapeutic recreation professionals have been published and are available (Burlingame & Blaschko, 2002) which include examples of instruments, the theoretical foundation and purpose of instruments, the contact information for the acquisition of instruments, and the related reliability and validity information. New usable text books are hot off the press that provide specific direction for assessment (Burlingame & Blaschko; Stumbo, 2002), scale development (DeVellis, 2003) and outcome measurement (Stumbo, 2003). Journal articles and book chapters also provide specific direction related to the evaluation and selection of standardized measurement instruments (Anastasi, 1988; Dunn, 1989; Gronlund & Linn, 1990; Stumbo, 2002; Stumbo & Rickards, 1986; Ward & Murray-Ward, 1999), the use of outcome measurement for program evaluation (Stumbo & Peterson, 2004; Lee & Sampson, 1990; Widmer, Zabriskie, & Wells, 2003), and the use of outcome measurement to help meet external accreditation standards through the performance improvement process (McCormick, 2003; McCormick & Funderburk, 2000). Professional organizations (i.e. NTRS, ATRA, SPRE) promote the dissemination and utilization of new assessment and outcome measurement scales and procedures through professional conferences and symposia (Sibthorp, 2001; Zabriskie, 2001). Both NTRS and ATRA support a research committee or team comprised of educators and practitioners, who serve as resources and provide networking services to therapeutic recreation professionals. They are eager to facilitate the connection between professionals in the field who are interested in refining or testing the psychometric properties of their assessment instruments, with scholars and measurement

experts who have similar interests and are willing to advise and conduct collaborative works.

Therapeutic recreation professionals are strongly encouraged to take advantage of and utilize the many resources that are available as they reexamine the measurement instruments used in their assessment and outcome processes. Whether developing a new scale or evaluating an established instrument for possible use, it may be helpful to re-introduce a few measurement basics. The following is a brief review of some of the basic concepts related to the scale development and evaluation process. An example of many of these approaches can be seen in Shapiro and Yun's (in press) development and testing of the Sport Motivation Questionnaire (SMQ) presented in this issue. It must be recognized, however, that this is simply an overview, for a deeper understanding and methodological approaches please consult the previously discussed and/or other related resources.

## Measurement Basics in Scale Development

*Scope/theory.* One of the first steps, whether developing a new instrument or evaluating an established one, is to thoughtfully consider exactly what you would like to measure (DeVellis, 2003). Although it appears quite obvious, being clear on what you want to measure is often overlooked until much later when it is difficult to make changes. Therapeutic recreation professionals must take into consideration the population that they work with, the types of programs and interventions that are provided, and the specific behaviors or constructs that they hope to influence. It serves little purpose to effectively measure and collect baseline information on behaviors or constructs that are not intended to be affected or changed as a result of interventions provided. The level of desired specificity in measurement must also be considered and clearly defined. "Sometimes a scale is intended to relate to very specific behaviors or constructs, whereas at other times a more general and

global measure is sought" (DeVellis, p. 61). Finally, a clear understanding of related theory is perhaps one of the most significant aspects that must be considered when deciding what to measure. Therapeutic recreation professionals and scholars have consistently made recommendations to increase and improve theory based practice in an effort to provide sound, predictable, and generalizable interventions, services, and research that result in measurable outcomes and benefits for clients (Caldwell, 2003; Coyle, Kinney, & Shank, 1993; Devine & Wilhite, 1999; Dunn, 1993). Similarly, theories that guide intervention approaches and predict related outcomes must also guide the measurement of those outcomes. Achievement goal theory and its' related constructs, for example, provided the theoretical foundation necessary to guide the development of the SMQ (Shapiro & Yun, in press).

*Format/item development.* The next two strategies, deciding on the format of the scale and developing items for the scale, tend to occur at the same time. There are a variety of scale formats that may be considered including Likert-type scales, semantic differential scales, visual analog scales, numerical response formats, ranking scales, binary option scales (ie. agree/disagree; yes/no), and scales formatted with or without time frame specification (DeVellis, 2003). When deciding on format, including response options and instructions, it is again essential to consider the population that the scale will be used with, theories related to the constructs being measured, and the intended purpose of the scale. Items can then be developed with the format, underlying theory, and overall purpose in mind. The SMQ was developed specifically for adults with mental retardation which was the key factor in the argument for using a semantic differentiation format. It is also helpful to design items to maximize the variability in responses while operating within the most appropriate format (Widmer, Zabriskie, & Wells, 2003). Other aspects of item development to be considered include redundancy, number of items, positive and negative framing, consistency, clear word-

ing, and creating items that are relevant to and representative of the domain of interest.

*Content related evidence of validity.* Evidence of validity and reliability should be one of the most important concerns whether the therapeutic recreation professional is developing a new scale, refining or revising a scale currently utilized, or considering a new standardized scale for possible use. Validity refers to the degree to which test results reflect their intended purpose, or in other words, does the scale measure what it is supposed to measure (Stumbo, 2002)? There are three basic types of evidence related to validity which include evidence of content, criterion, and construct validity. Content related evidence of validity is concerned with determining how well the items in a scale represent the overall domain that it intends to measure. For example, if an aspect of a scale was intended to measure leisure constraints, content validity would help determine how representative the results were of all possible leisure constraints. A clear definition of the overall content or domain that is to be measured is helpful in determining content validity (DeVillis, 2003).

The primary means for gathering evidence of content validity is the use of expert opinion and evaluation. A panel of experts who have worked extensively with the particular population and the specific construct being measured are typically identified and invited to help evaluate the scale. After being informed of the theoretical foundation and working definitions related to the scale, experts are asked to utilize that framework to rate how relevant each item is to the construct that the scale is intended to measure (DeVillis, 2003). They are often asked to provide further information by evaluating the clarity and conciseness of the items and by pointing out other aspects of the construct or other ways to address the construct that have not been included in the scale. In the development of the SMQ input from the panel of experts resulted in the elimination of 2 items, the addition of 5 items and the restating of 21 items to improve word appropriateness. Such information becomes very valuable and essential in maximizing content validity and in gathering content-related evidence of validity.

*Criterion-related evidence of validity.* Criterion-related evidence of validity is concerned with demonstrating an empirical relationship between scale results and some other standard or criterion. DeVillis (2003) argued that with evidence of criterion-related validity it is irrelevant whether the theoretical basis for the relationship between the test scores and the criterion are understood. He stated "If one could show, for example, that dowsing is empirically associated with locating underground water sources, then dowsing would have validity with respect to the criterion of successful well digging" (p. 50). Criterion-related validity can be referred to as predictive validity (when scores are correlated to a future criterion), concurrent validity (scores correlated to a current criterion), or may even be postdictive (scores correlated to a previous criterion). Criterion-related validity, however, is preferred because it is a temporally neutral term and includes all of the approaches that provide evidence of validity based on a relationship with a separate criterion or standard. The primary means for gathering evidence of criterion-related validity is by examining statistical correlation between test scores and a measurement of the related criterion. Stumbo (2002) suggests that acceptable ranges of correlation coefficients are between  $r = .40$  to  $.70$  or  $r = -.40$  to  $-.70$  if negative correlations are sought.

*Construct-related evidence of validity.* Construct evidence of validity is concerned with determining how well the scores from the scale correlate with measures of some other theoretically related construct. For example, if an individual scores high on a scale which measures leisure involvement, they should also score high on a life satisfaction scale due to the theoretical relationship between leisure involvement and life satisfaction (Metzelaars, 1994; Ragheb & Griffith, 1982; Riddick, 1986). Therefore, the correlation coefficient between scores from the scale of interest and theoretically related constructs demonstrate the degree to which the construct being measured theoretically behaves like the variable it is supposed to measure, and thus, provides

evidence of construct validity (DeVillis, 2003). The primary means for collecting construct-related evidence of validity is by examining the direction and strength of the statistical correlation between test scores and measurements of theoretically related constructs gathered by previously established instruments. Other approaches include utilizing more advanced statistical procedures to determine if the factors measured are theoretically similar to the underlying constructs that the scale is intended to measure. Shapiro and Yun (in press) provided preliminary evidence of construct validity by determining that the factor structure of the SMQ scores were congruent with the theoretical explanation of achievement goal theory. Overall, evidence of content, criterion, and construct validity provide a degree of confidence that the use of a scale in a particular situation or with a certain population produces valid results and indeed measures what it is intended to measure.

*Internal consistency.* Estimates of reliability provide evidence of how accurately and consistently the scale measures what it is supposed to measure (Stumbo, 2002). There are three basic ways to estimate reliability of scores which include internal consistency, stability, and equivalence (Stumbo). Internal consistency is perhaps one of the most commonly used estimates of reliability and refers to how well the items within a scale correlate with each other. It is based on the assumption that if items on a scale have a strong relationship with the construct that they are measuring, then they should also have a strong relationship with each other. It is important to note, however, that all scales are not necessarily designed to measure the same construct in the same format; therefore estimates of internal consistency may not always be an appropriate estimate of reliability. The primary means for estimating internal consistency is by utilizing Cronbach's (1951) alpha coefficient ( $\alpha$ ). Cronbach's alpha is a statistical process that considers the interrelationship of each item and the total score to compute the alpha coefficient. The alpha coefficient therefore, is an

estimate of how consistently each item measures the same information or construct. Shapiro and Yun (in press) utilized Cronbach's alpha to provide an estimate of internal consistency for each sub-factor measured by the SMQ.

*Stability/test-retest reliability.* Stability makes estimates of score reliability based on how consistent scores are over time. It is based on the assumption that if a scale produces a particular score at one point in time, then it should produce the same or similar score at a later point in time, given the construct being measured did not change. In other words, if an individual is measured to be six foot eight inches tall on Monday, he should also be measured to be six foot eight inches tall on Friday, given he did not grow. The primary means for estimating stability is by test-retest procedures which are often referred to as test-retest estimates of reliability. In this case the scale is given to a group of individuals, then, the same scale is given to the same group of individuals on a later date. Therefore, the correlation between the two sets of scores on the scale provides an estimate of how consistent the scale measures the same thing over time. The time lapse between scale administrations is an important consideration with this method (Stumbo, 2002). If the time is too short, the first administration is likely to affect the second, and if the time is too long, it is more likely for the scores to be different due to actual change in the respondent and construct being measured. Although the time lapse must be considered in each case based on the scale, sample etc, it is commonly a matter of weeks versus just a couple of days or several months. Shapiro and Yun (in press) utilized an interval of about 3 weeks between test administrations to gather evidence of stability through the test-retest process.

*Equivalency.* Equivalency makes estimates of score reliability based on the consistency of scores collected from the same individual with two different forms of the scale, or the consistency of scores between two different individuals using the same scale to measure the same construct. The first is known as parallel-form

reliability, in which evidence is gathered by examining the correlation between scores collected with two different forms (parallel or equivalent) of the same scale that were administered to the same individuals in the same basic time frame. This approach, however, does depend on having two forms of the same instrument that are slightly different, yet measure the same construct and have similar psychometric properties, which is not always possible. Another approach to equivalency, known as inter-rater reliability, is used when assessment information is collected through observation (Stumbo, 2002). In this case, two different individuals observe the same behaviors and complete the behavioral assessment. A correlation between scores would provide evidence of how consistent the observations and scores are between the two different raters. This approach is also limited, however, to scales that measure somewhat objective or observable behaviors that different individuals can rate, as opposed to those that measure an individuals' feelings or subjective opinion. Both parallel forms and inter-rater reliability offer equivalency procedures to determine consistency. Overall, estimates of internal consistency, stability, and equivalency provide different approaches to gather evidence of reliability in order to establish a degree of confidence that the scale is performing in an accurate and reliable fashion.

*Usability.* Another concept that must be considered when establishing a new instrument or evaluating one for possible use is the idea of usability. Therapeutic recreation professionals must determine how usable the scale will be with their specific population and in their particular situation. This consists of asking practical, common sense types of questions such as: How long is the scale? How difficult is it? How long does it take to complete? What does it look like? The answers to such questions must be compared with the abilities and characteristics of the population with whom the scale will be used. Their attention span, physical capacities, and ability to understand the items, will help determine if the

scale is usable with the population. This particular area posed significant challenges for Shapiro and Yun (in press) in their efforts to design the SMQ in a manner that was developmentally appropriate for adult athletes with mental retardation. Several decisions, such as reducing the number of items from 37 to 14, were the result of usability considerations. Other questions relate to the practicality of the scale such as: How long does it take to score? Does one need specific training or qualifications to administer or interpret it? How in depth or detailed are the results? Are there established norms that can be utilized? How available or affordable is the scale? These types of questions must also be considered in comparison to resources available, client length of stay, staff qualifications, availability, time etc. If the questions of practicality and usability are not considered prior to the development or adoption of a new measurement instrument, the possibility for related difficulties and problems increase significantly.

*Example.* The following is a brief example of how some of the concepts described above were utilized in the development and testing of the Family Leisure Activity Profile (FLAP) (Zabriskie, 2001). The FLAP is a 32-item activity inventory that measures family leisure involvement. The scale was developed based on the theoretical foundation of the Core & Balance Model of Family Leisure Functioning (Zabriskie, 2000; Zabriskie & McCormick, 2001). This specific theoretical perspective provided the framework to guide and organize the development of items and the format of the scale. An initial pilot study was conducted (Zabriskie & McCormick) which examined the theoretical relationship between different family leisure patterns and different aspects of family functioning. Findings indicated correlations in the directions expected and provided preliminary construct-related evidence of validity. The scale was then sent to a variety of experts in family leisure who were asked to evaluate how representative the items were of the overall domain of all possible family leisure activities. They were then given theoret-

ical definitions and asked to rate items based on the definitions as well as to provide suggestions of areas that may not have been addressed sufficiently. This process provided important content-related evidence of validity and resulted in further refinement of the scale including the addition of two more items (Zabriskie, 2001).

Although the format of the scale was not appropriate for the use of internal consistency, stability and equivalency methods were used to address reliability. A test-retest study was conducted by administering the scale to a sample of respondents and then administering it again after a 5 week interval (Zabriskie, 2000). The correlations between administrations for both the individual items and the overall scores provided evidence of test-retest reliability.

After establishing a degree of reliability it was necessary to continue to examine the validity of scores measured by the scale. Therefore, other established scales were included in the research questionnaires that measured related criteria and constructs. Significant positive correlations between the criterion of family satisfaction and family leisure involvement provided criterion-related evidence of validity. Findings again identified significant correlations between different theoretically related aspects of family functioning and different family leisure patterns. The correlations between these constructs were again in the expected direction, were found in parent, child, and family level data sets, and provided further construct-related evidence of validity.

Subsequent known group studies have also been conducted to gather additional construct-related evidence of validity. For example, intact families that have adopted a child with special-needs have been reported to have higher levels of family cohesion and family adaptability than non-adoptive families (Groze, 1996; Hoopes, Alexander, Silver, Ober, & Kirby, 1997; Rosenthal & Groze, 1990). According to the theoretical model that the FLAP is based on, if special-needs adoptive families have higher levels of cohesion and adaptability than non-adoptive families,

then they should also have higher levels of family leisure involvement. Results of a study comparing these two samples of families (Freeman & Zabriskie, 2003) indicated that the adoptive family sample did report significantly higher levels of family leisure involvement than the non-adoptive family sample, and therefore, provided further construct-related evidence of validity.

By utilizing a variety of different approaches, the above procedures have provided considerable evidence of validity and reliability related to the use of the FLAP. Evidence such as this allows therapeutic recreation professionals to use a scale with a certain degree of confidence that the scale will consistently and accurately measure what it is intended to measure. Such evidence also adds considerable credibility when assessment and outcome measurement processes are evaluated by a variety of different stakeholders.

## Summary

In order to demonstrate accountability and provide viable therapeutic recreation services in the current healthcare environment, it has become increasingly critical for therapeutic recreation professionals to have a sound understanding of measurement. The development of the Sport Motivation Questionnaire, as described by Shapiro and Yun (in press), provides a model for understanding issues related to assessment for individuals with disabilities.

While there are a variety of resources specifically designed for therapeutic recreation professionals that can assist in the refinement of assessment and outcome measurement processes, assessment instruments from outside the field of therapeutic recreation can also be useful. It is recommended that therapeutic recreation professionals utilize a wide variety of resources to evaluate and improve measurement within their own practice.

The development of the Sport Motivation Questionnaire by Shapiro and Yun (in press) was based on sound principles of instrument design and provides a model for instrument design in therapeutic recreation practice. The instrument

was conceptualized from a thorough understanding of theory related to achievement motivation and issues related to assessment for individuals with cognitive impairments. In particular, the wording of the items and the response format reflected a strong awareness of concerns pertaining to assessment with individuals with mental retardation. Content validity was further enhanced through the use of a panel of experts. The items were carefully scrutinized by a number of individuals with expertise in the areas of sport, Special Olympics, and individuals with disabilities. The items were then revised based on the expert's feedback. After administering the instrument to a number of Special Olympic athletes, indications of reliability were evaluated. Internal consistency of the instrument was evaluated through using a statistical procedure (Cronbach's alpha coefficient). Some athletes completed the SMQ twice in order to assess test-retest reliability or stability over time. Finally, issues related to the usability of the instrument guided all aspects of the development and testing process. The number, wording, and response choices of the items were reviewed to determine the ease of use with the target population of Special Olympic athletes.

Utilizing instruments that provide valid and reliable measurement of client behaviors increases the likelihood of providing appropriate need-based interventions that will lead to desired client outcomes. The documentation of client outcomes is a vital component of quality care today. Such practices improve the quality of client care and provide outcome-based evidence of that care which is essential to the accountability process.

However, therapeutic recreation specialists often struggle to find appropriate existing assessment instruments for their agency or practice, and often are in the position of having to develop their own assessment instrument. In these situations, it is imperative that therapeutic professionals utilize knowledge and resources related to scale development to begin the development of sound new instruments when necessary and to gather evidence of validity and reliability related to those instruments. Such efforts will make

considerable contributions to their own agency and to the profession as a whole as it responds to the demand for valid and reliable measurement in order to demonstrate quality therapeutic recreation outcomes.

## References

- Ackerman, C. N. (2002). Development and validation of a leisure and anxiety scale. *Dissertation Abstracts International: Section B: The Sciences & Engineering*, 64(1-B), 559.
- Anastasi, A. (1988). *Psychological testing*. New York, NY: Macmillan.
- Austin, D. R., & Crawford, M. E. (2001). *Therapeutic recreation: An introduction* (3<sup>rd</sup> ed.). Needham Heights, MA: Allyn and Bacon.
- Burlingame, J., & Blaschko, T. M. (2002). *Assessment tools for recreational therapy and related fields* (3<sup>rd</sup> ed.). Ravensdale, WA: Idyll Arbor, Inc.
- Caldwell, L. L. (2003). Basing outcomes on theory: theories of intervention and explanation. In N. J. Stumble (Ed.), *Client outcomes in therapeutic recreation services* (pp. 67–86). State College, PA: Venture Publishing, Inc.
- Coyle, C. P., Kinney, W. B., & Shank, J. (1993). Purpose of positivistic research design. In M. J. Malkin & C. Z. Howe (Eds.), *Research in therapeutic recreation: Concepts and methods* (pp. 207–231). State College, PA: Venture.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- DeVellis, R. F. (2003). *Scale development: Theory and applications* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Devine, M. A., & Wilhite, B. (1999). Theory application in therapeutic recreation practice and research. *Therapeutic Recreation Journal*, 33(1), 29–45.
- Dunn, J. K. (1989). Guidelines for using published assessment procedures. *Therapeutic Recreation Journal*, 23(2), 59–69.
- Dunn, J. K. (1993). Interpretation of positivistic research: Issues of validity and reliability. In M. J. Malkin & C. Z. Howe (Eds.), *Research in therapeutic recreation: Concepts and methods* (pp. 161–179). State College, PA: Venture.
- Freeman, P., & Zabrickie, R. B. (2003). Leisure and family functioning on adoptive families: Implications for therapeutic recreation. *Therapeutic Recreation Journal*, 37(1), 73–93.

- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6<sup>th</sup> ed.). New York, NY: Macmillan Publishing.
- Groze, V. (1996). A 1 and 2 year follow-up study of adoptive families and special needs children. *Children and Youth Service Review, 18*(1/2), 57–82.
- Hoopes, J. L., Alexander, L. B., Silver, P., Ober, G., & Kirby, N. (1997). Formal adoption of the developmentally vulnerable African-American child: Ten-year outcomes. *Marriage and Family Review, 25*(3/4), 131–144.
- Joint Commission on Accreditation of Health-care Organizations (2003). *Performance measurement in health care*. Retrieved October 11, 2003, from <http://www.jcaho.org/pms/index.htm>
- Kloseck, M., Crilly, R. G., Ellis, G. D., & Lambers, E. (1996). Leisure Competence Measure: Development and reliability testing of a scale to measure functional outcomes in therapeutic recreation. *Therapeutic Recreation Journal, 30*(1), 13–26.
- Lee, L. J., & Sampson, J. F. (1990). A practical approach to program evaluation. *Evaluation and Program Planning, 13*, 157–164.
- McCormick, B. P. (2003). Outcomes measurement as a tool for performance improvement. In N. J. Stumbo (Ed.), *Client outcomes in therapeutic recreation services* (pp. 221–231). State College, PA: Venture Publishing, Inc.
- McCormick, B. P., & Funderburk, J. (2000). Therapeutic recreation outcomes in mental health practice. *Annual in Therapeutic Recreation, 9*, 9–20.
- Metzelaars, G. J. (1994). *The relationships among college satisfaction, leisure satisfaction, and life satisfaction in adult college students*. Unpublished doctoral dissertation, University of Maryland.
- Mordock, J. B. (2000). Outcome assessment: Suggestions for agency practice. *Child Welfare, 79*(6), 689–710.
- Peebles, J., McWilliams, L., Norris, L. H., & Park, K. (1999). Population-specific norms and reliability of the Leisure Diagnostic Battery in a sample of patients with chronic pain. *Therapeutic Recreation Journal, 33*(2), 135–141.
- Petrick, J. F. (2002). Development of a multi-dimensional scale for measuring the perceived value of a service. *Journal of Leisure Research, 34*(2), 119–134.
- Ragheb, M. G., & Griffith, C. A. (1982). The contribution of leisure participation and leisure satisfaction to life satisfaction of older persons. *Journal of Leisure Research, 18*(4), 248–258.
- Riddick, C. C. (1986). Leisure satisfaction precursors. *Journal of Leisure Research, 18*(4), 259–265.
- Rosenthal, J., & Groze, V. (1990). Special needs adoption: A study of intact families. *Social Service Review, 64*, 475–505.
- Shapiro, D. R., & Yun, J. (in press). Sport Motivation Questionnaire for persons with mental retardation. *Therapeutic Recreation Journal*.
- Sibthorp, J. (2001). Development and validation of the characteristics of experience scale for use in adventure based recreational programming. *Abstracts from the 2001 Symposium on Leisure Research, 30*. Ashburn, VA: National Recreation and Park Association.
- Stumbo, N. J. (2002). *Client assessment in therapeutic recreation services*. State College, PA: Venture Publishing, Inc.
- Stumbo, N. J. (2003). *Client outcomes in therapeutic recreation services*. State College, PA: Venture Publishing, Inc.
- Stumbo, N. J., & Peterson, C. A. (2004). *Therapeutic recreation program design: Principles and procedures* (4<sup>th</sup> ed.). San Francisco, CA: Pearson Benjamin Cummings.
- Stumbo, N. J., & Rickards, W. H. (1986). Selecting assessment instruments: Theory into practice. *Journal of Expanding Horizons in Therapeutic Recreation, 1*(1), 1–16.
- Ward, A. W., & Murray-Ward, M. (1999). *Assessment in the classroom*. Belmont, CA: Wadsworth Publishing.
- Widmer, M. A., Zabriskie, R. B., & Wells, M. S. (2003). Program evaluation: Collecting data to measure outcomes. In N. J. Stumble (Ed.), *Client outcomes in therapeutic recreation services* (pp. 201–219). State College, PA: Venture Publishing, Inc.
- Zabriskie, R. (2000). An examination of family and leisure behavior among families with middle school aged children. Unpublished dissertation. Indiana University, Bloomington, IN.
- Zabriskie, R. B. (2001). The validity and reliability of the Family Leisure Activity Profile (FLAP). *Abstracts from the 2001 Symposium on Leisure Research, 66*. Ashburn, VA: National Recreation and Park Association.
- Zabriskie, R., & McCormick, B. (2001). The influences of family leisure patterns on perceptions of family functioning. *Family Relations: Interdisciplinary Journal of Applied Family Studies, 50*(3), 66–74.